Reconsidering the Miracle Argument on the Supposition of Transient Underdetermination

Paul Hoyningen-Huene

Center for Philosophy and Ethics of Science, Leibniz University of Hannover

Abstract. In this paper, I will show that the Miracle Argument is unsound if one assumes a certain form of transient underdetermination. For this aim, I will first discuss and formalize several variants of underdetermination, especially that of transient underdetermination, by means of measure theory. I will then formalize a popular and persuasive form of the Miracle Argument that is based on “use novelty”. I will then proceed to the proof that the miracle argument is unsound by means of a mathematical example. Finally, I will expose two hidden presuppositions of the Miracle Argument that make it so immensely though deceptively persuasive.

1. Introduction. The Miracle Argument is an important argument for the defense of realism in science. Its current name was introduced by Hilary Putnam (1975, 73); van Fraassen called it the “Ultimate Argument” (1980, 39). It is a “positive argument for realism” claiming that realism “is the only philosophy that doesn’t make the success of science a miracle” (Putnam (1975, 73)). In its clearest and strongest form, the “realism” referred to above is spelled out as “scientific realism” and the “success of science” referred to above as “(novel) predictive success of science” (see Musgrave (1999 [1988], 60)). The argument is then conceived of as an inference to the best explanation consisting of two steps. First, it is claimed that realism explains (novel) predictive success of science satisfactorily and better than any non-realist philosophy of science. Second, it is concluded that it is therefore justified to accept scientific realism.

Transient underdetermination claims, roughly speaking, that with respect to currently available evidence, any theory T relevant for and consistent with this evidence has radically false rival theories equally relevant for and consistent with the same evidence. As I will show, transient underdetermination undermines a crucial premise of the Miracle Argument. Thus, transient underdetermination shows that the Miracle Argument is unsound.

In what follows, I will first present the version of the Miracle Argument that I will be discussing (section 2). I will then introduce various forms of underdetermination, especially three variants of transient underdetermination in sections 3 and 4. I will assume one particular variant of transient underdetermination as a premise for the rest of the paper. In sections 5 and 6, I will formally analyze the consequences of transient underdetermination for the Miracle Argument. The result is that, under the assumption of transient underdetermination the Miracle Argument fails. In section 7, I will analyze why the Miracle Argument appears so immensely plausible, in spite of being fallacious given transient underdetermination. The paper closes with a short summary (section 8). Note that the paper only addresses the Miracle Argument as it relates to scientific realism, and is completely silent about structural realism.

2. The Miracle Argument. In somewhat more formal terms, the presumed situation in which the Miracle Argument is to be applied is as follows.

Let us assume that appropriate notions of truth and of approximate truth of theories have been defined, and that true and approximately true theories exist (otherwise the whole discussion about scientific realism would not make sense).

Let D_1 be a finite set of data.[1]

Let T_1 be the set of theories such that $T_1 := \{T, T \text{ is relevant for and consistent with } D_1\}$; we assume that $T_1 \neq Ø$

Let us now introduce a partition of T_1 into two subsets: those theories which are true or approximately true, and those which are radically false, by which I mean that they are not even approximately true. Thus

$T_1^{AT} := \{T \in T_1, \text{T is true or approximately true}\}$

$T_1^{RF} := \{T \in T_1, \text{T is radically false}\}$

with $T_1 = T_1^{AT} \cup T_1^{RF}$.

As stated above, we assume that $T_1^{AT} \neq Ø$. We further assume the idealization that there is a sharp boundary between true and approximately true theories on the one hand and radically false theories on the other, i.e., $T_1^{AT} \cap T_1^{RF} = Ø$. Remember that even the radically false theories contained in T_1^{RF} are relevant for and consistent with the data D_1; their radical falsity does not derive from their relationship to the data D_1 but concerns their status when judged from the presupposed notions of truth and approximate truth of theories.

Now let N be novel data relative to the data D_1 and the set of theories T_1.[2] The sense of novelty relevant in the context of the Miracle Argument is not temporal novelty but what has been dubbed "use novelty" (or "strong predictive success"; see, e.g., Worrall (1985); Worrall (1989, 148-9); Carrier (1991, 26-28); Earman (1992, 114-5); Leplin (1997); Psillos (1999, 106)). The data N are not only different from the original data D_1 but they have also not been used in the construction of the theories contained in T_1 (these theories have only been adapted to the data D_1). This concept of use-novelty is sometimes vague and ambiguous in its application (see, e.g., Earman (1992, 115)) but this fact does not have to concern us here because the existing clear-cut cases will suffice to get the Miracle Argument off the ground.

Let us now assume that there is a theory $T^* \in T_1$, i.e., a theory that is relevant for and consistent with the original data D_1, that is capable of predicting the novel data N although it has not been adapted to this task. How can this fact be explained? A possible explanation is that $T^* \in T_1^{AT}$, i.e., that T* is true or approximately true. In fact, *if* T* gets something of the world approximately or even completely right, it is not surprising that it can predict certain data that have not been used in its construction; thus the proposed explanation is *satisfactory*. Moreover, it seems impossible that a radically false theory could generate novel successful predictions—it simply lacks the resources to do so. Thus, the proposed explanation is *better than any possible rival*—it may even be the best and only explanation. This concludes the first step of the Miracle Argument, which shows that realism explains novel predictive success of science satisfactorily and better than any non-realist philosophy of science. The second step of the argument, which is not immediately important at this point, then infers the superiority of realism over anti-realism by an inference to the best explanation.

3. Underdetermination. Let us now bring underdetermination into the picture. One form of underdetermination is called “radical” or “strong” or “Quinean” underdetermination. It states that for any theory T, there are always empirically totally equivalent theories that are not compatible with T. In other words, on the basis of any conceivable empirical data, we are unable to decide which of these rival theories we should accept. Given that these theories may come with radically different ontologies, strong underdetermination is a very serious threat to scientific realism. Given empirically equivalent theories with different ontologies, we cannot rationally prefer one or the other set of theoretical entities that these theories imply, based on empirical data (see, e.g., Quine (1951, sect. VI); however, see Severo (2008) on the difficulties to exactly pin down Quine’s position and its reception). Thus, any choice between different sets of theoretical entities lacks an empirical basis, making scientific realism extremely problematic.

As strong underdetermination is a very strong concept indeed, its existence has been a matter of controversy (see, e.g., Stanford (2006, 11-16). However, rather than trying to resolve this controversy, another strategy has been to weaken the concept of underdetermination so as to produce a concept whose applicability appears less controversial. The result is typically called “weak” underdetermination or, as introduced by Lawrence Sklar, “transient” underdetermination (of theories by data) (Sklar (1975, 380-381)). Transient underdetermination holds if and only if, with respect to *currently available evidence*, any theory T relevant for and consistent with this evidence has rival theories equally relevant for and consistent with the same evidence. This sort of underdetermination is transient because any data available in the future may destroy the equivalence of T and its rivals with respect to the then available evidence. Some authors claim that transient underdetermination is a fairly uncontroversial and empirically confirmed fact of science. Sklar writes that "[e]ven those skeptical of the very possibility of radical underdetermination are likely to admit that transient underdetermination is a fact of epistemic life" (1975, 381). Kyle Stanford even suggests "that the historical record of scientific inquiry provides compelling evidence that recurrent, transient underdetermination is our actual epistemic predicament in theoretical science rather than a speculative possibility" (2006, 18). Or: "the historical record of scientific inquiry itself provides us with abundant empirical evidence that there are probably scientifically plausible alternatives to even the best contemporary fundamental scientific theories that are equally well-confirmed by the evidence available to us" (2009, 2). The reason is that quite often scientifically acceptable competitors for theories that were accepted in their time but are obsolete now, were invented only later. These competitors were conceptually out of reach at the earlier time – they were "unconceived alternatives", as Stanford calls them. The important point here is that these unconceived alternatives to well-confirmed theories are by no means bizarre philosophical constructions that no scientist would ever consider. Just the opposite: they are the theories that were later accepted as direct or indirect successors of the theories in question.

Be that as it may, transient underdetermination as just introduced appears to be a pretty clear concept. However, it turns out that there are different variants of this concept available which differ in strength and, consequently, in their possible roles in arguments relevant to the debate about scientific realism. Before entering this debate, it is thus advisable to distinguish some of these variants.

4. Variants of transient underdetermination. Before introducing these variants, let us fix our notation. As in section 2, let D_1 be a finite set of data and let T_1 be the set of theories relevant for and consistent with D_1. In its weakest form, transient underdetermination (TU) states that for every theory T from T_1 there exists another theory T′ from T_1 that is not compatible with T. In more formal terms:

Definition 1 of TU:

$$\text{TU holds iff } \forall T\, [(T \in T_1) \rightarrow \exists T'\, (T' \in T_1 \wedge \neg(T' \wedge T))].$$

Note that "$\neg(T' \wedge T)$" means that T′ and T are not compatible but the source of this incompatibility is not specified. An obvious candidate is, of course, logical incompatibility but another candidate can be incommensurability if it is specified such that it implies incompatibility (as intended by Kuhn and Feyerabend, for instance). Definition 1 of TU is very weak indeed because it is already fulfilled by two minimally differing theories consistent with the given data, for instance the true theory and one minimally differing from it at one point, i.e., an approximately true theory. Definition 1 is, in fact, only a necessary condition for an adequate definition of TU because TU posits the existence of radically false alternatives that are nevertheless relevant for and consistent with the given data, to (approximately) true theories. The basic idea is that in a given historical situation, there are also theories operating with radically false basic assumptions in spite of their agreement with the available data. For instance, at some historical time, phlogiston theory may have been in good agreement with the available data in spite of its radical falsity, as seen from today's point of view. The idea of the existence of radically false alternatives, relevant for and consistent with the given data, to a true theory can easily be articulated by means of the partition of T_1 into the two subsets: those theories which are true or approximately true, and those which are radically false. We will denote these subsets as T_1^{AT} and T_1^{RF}, respectively, and assume, as in section 2, $T_1^{AT} \neq$

Ø. The second attempt at a definition of TU, stronger than the first attempt, reads:

Definition 2 of TU:

TU holds iff $T_1^{RF} \neq Ø$.

For the purposes of my argument, definition 2 of TU is still too weak: it does not capture the idea that transient underdetermination means that there must be “quite a few” radically false theories in T_1. In the literature, there are a variety of arguments to the effect that for a given set of data, there are many more radically false theories than (approximately) true theories fitting these data. I am not going to review, let alone to evaluate, these arguments in this paper. I am rather assuming this form of transient underdetermination, and shall try to give it a precise form and evaluate its consequences for the Miracle Argument.

In order to precisely articulate the different magnitudes of the theory sets in question, I need the mathematical concept of a measure on a space of theories. A measure is a generalization of the familiar concept of volume which is defined for the 3-dimensional Euclidean space. In other words, a measure states how big a subset of a space is, for more general spaces than just 3-dimensional Euclidean space. By means of a measure on the theory space T_1, we can express the idea about the differing relative size of the theory sets T_1^{AT} and T_1^{RF}. According to our supposition, the measure μ of T_1^{RF}, i.e. $\mu(T_1^{RF})$, will be much larger than $\mu(T_1^{AT})$. So underdetermination in this form tells us that $\mu(T_1^{AT}) << \mu(T_1^{RF})$. Thus, we may formulate this variant of transient underdetermination in the following way:

Definition 3 of TU:

TU holds iff $\mu(T_1^{AT}) << \mu(T_1^{RF})$.

In what follows, I will *presuppose* transient underdetermination in this form.

However, before we can proceed further, I need to say something about the general connection between the (prior) probability of finding an element of a given set in one of its subsets, and the measure of this subset. This connection is straightforward: the larger the subset is, the larger is the probability to find an element of the given set in it. In other words: The probability of finding an element of a given set S in its subset A is proportional to the measure of A, i.e., $p(s\in A|s\in S) \sim \mu(A)$

5. *Formal arguments with TU and MA.* Armed with the presupposed definition 3 of TU, we can now approach the arguments of both the anti-realists and the realists. Using the definitions of T_1, T_1^{AT}, and T_1^{RF} of section 2, the anti-realist can formulate the following argument:

Argument 1 (Transient Underdetermination)

$T_1 = T_1^{AT} \cup T_1^{RF}$ and $T_1^{AT} \cap T_1^{RF} = Ø$

$\mu(T_1^{AT}) << \mu(T_1^{RF})$

Therefore for any $T \in T_1$, it is very probable that $T \in T_1^{RF}$.

In other words, the argument states that the probability of any theory from T_1, i.e., of any theory that is relevant for and consistent with the data D_1, being radically false is much higher than its being (approximately) true. As stated above, this is because the probability of being (approximately) true or radically false is proportional to the measures $\mu(T_1^{AT})$ and $\mu(T_1^{RF})$, respectively. Thus, transient underdetermination tells us that for any $T \in T_1$, it is very probable that $T \in T_1^{RF}$.

Here we can see how, due to transient underdetermination, the empirical adequacy of a theory T with respect to some finite data set D_1 (i.e., being an element of T_1) does not translate into (approximate) truth of that theory (i.e., into T being an element of T_1^{AT}). This is how scientific realism is undermined by transient underdetermination.

However, the sophisticated form of the Miracle Argument specifically answers this move of the anti-realist by explicating the "success of science" not by mere empirical adequacy of a theory T* with respect to some finite data set but by the use-novel predictive success of science. Given T*'s successful prediction of the use-novel data N, the former probabilities for T*'s membership in T_1^{AT} or T_1^{RF}, which were solely based upon the consistency relation of T* to D_1, are obsolete. On the basis of the additional information (i.e., the use-novel data N), T*'s formerly *highly probable* membership in T_1^{RF} becomes *highly improbable* (or even impossible) because it would make T*'s use-novel predictive success a miracle. This sort of novel predictive success has happened again and again in the history of science; it can't be just a miracle but must have an intelligible cause, most likely that $T^* \in T_1^{AT}$. Thus, the Miracle Argument tells us that for any $T^* \in T_1$ that makes a use-novel successful prediction N, it is very probable (or even certain) that $T^* \in T_1^{AT}$.

Formally, the argument reads like this:

Argument 2 (Miracle Argument)

$T_1 = T_1^{AT} \cup T_1^{RF}$ and $T_1^{AT} \cap T_1^{RF} = \emptyset$

$\exists\, T^* \in T_1$ such that T* makes the novel prediction N

For any $T \in T_1^{RF}$, it is very improbable (or even impossible) to make prediction N.

Therefore, it is very probable (or even certain) that $T^* \in T_1^{AT}$.

In a sense, the Miracle Argument even seems to be strengthened by transient underdetermination. Transient underdetermination brings to the fore how immense the ocean of radically false theories is and how unlikely it is therefore to find a theory that will have novel predictive success beyond the empirical data for which the theory was designed. If, however, we hit upon a theory T* that does indeed have use-novel predictive success, we can be pretty sure (or even certain) that this theory is at least approximately true.

Let us have a closer look at the tension between the conclusions of Arguments 1 and 2:

Conclusion of Argument 1:

Therefore for any $T \in T_1$, it is very probable that $T \in T_1^{RF}$

Conclusion of Argument 2:

Therefore, it is very probable (or even certain) that $T^* \in T_1^{AT}$.

It is obvious that Argument 2 overrules Argument 1 because the former's conclusion about T* states a *posterior* probability based on *additional information*.[3]

6. *Transient Underdetermination, again.* Let us now investigate the potential effects that transient underdetermination has in a situation where a theory T* that has been adapted to the data set D_1 is capable of predicting use-novel data N. The following line of argument is very tempting. After the data N have been produced, scientists try to invent theories that are adapted to the new data set $D_2 := D_1 \cup N$. When the situation is described in this way, it is completely analogous to the situation in section 2 where we discussed the relationship between data D_1 and theories T_1.

We thus have in complete analogy with section 2:

$D_2 = D_1 \cup N$ is a finite set of data.

Let T_2 be the set of theories such that $T_2 := \{T$, T is relevant for and consistent with $D_2\}$.

Again, we introduce a partition of T_2 into two subsets: those theories that are true or approximately true, and those that are radically false. Thus

$T_2^{AT} := \{T \in T_2$, T is true or approximately true$\}$

$T_2^{RF} := \{T \in T_2$, T is radically false$\}$

with $T_2 = T_2^{AT} \cup T_2^{RF}$.

Remember, again, that the radically false theories contained in T_2^{RF} are also relevant for and consistent with the data D_2.

At this point, we can bring in transient underdetermination again. As in the case of the sets T_1^{AT} and T_1^{RF}, transient underdetermination tells us about the relative sizes of the sets T_2^{AT} and T_2^{RF}. It tells us that

$$\mu(T_2^{AT}) << \mu(T_2^{RF}).$$

Thus, most of the theories that manage to be relevant for and consistent with the old data D_1 *and* the novel data N are *not* even approximately true.

Clearly, the theory T* that produces the novel prediction N (beyond being consistent with and relevant for D_1) is a member of T_2. Now the crucial question is to which subset of T_2 does T* belong, to T_2^{AT} or to T_2^{RF}? Judged *solely* on the basis that T* is a member of T_2, the answer is clear: it is very probable that T* is a member of T_2^{RF} (because $T_2 = T_2^{AT} \cup T_2^{RF}$, $T_2^{AT} \cap T_2^{RF} = \emptyset$ and $\mu(T_2^{AT}) << \mu(T_2^{RF})$). Yet, we know more about T*: T* is not only consistent with and relevant for $D_2 := D_1 \cup N$, but having been adapted *only* to D_1 it was still able to predict the novel data N. Does this fact change the picture and qualify T* to be with very high probability a member of T_2^{AT} as the Miracle Argument has it? To answer this question, we must again compare the sizes of the relevant sets. Let us first define these sets.

To begin, we need the set of all theories from T_1 that are able to produce the use-novel prediction N which I call $T_{1\rightarrow 2}$. This is the set of all theories that have been adapted to D_1, thus they are members of T_1, but nevertheless predict the new data N, i.e., they are relevant for and consistent with $D_2 := D_1 \cup N$, which means that they are members of T_2. This yields

$$T_{1\rightarrow 2} := \{T, T\in T_1, \text{T is predictively successful with respect to N}\}.$$

We are interested in the relation between the sizes of the subsets $T_{1\rightarrow 2}^{AT}$ and $T_{1\rightarrow 2}^{RF}$ of $T_{1\rightarrow 2}$: $T_{1\rightarrow 2}^{AT}$ contains all (approximately) true theories of $T_{1\rightarrow 2}$,

$$T_{1\rightarrow 2}^{AT} := \{T, T\in T_{1\rightarrow 2} \wedge \text{T is (approximately) true}\}$$

whereas $T_{1\rightarrow 2}^{RF}$ contains all radically false theories of $T_{1\rightarrow 2}$,

$T_{1\to2}^{RF} := \{T, T\in T_{1\to2} \wedge T \text{ is radically false}\}$.

As earlier, we assume $T_{1\to2}^{AT} \cup T_{1\to2}^{RF} = T_{1\to2}$ and $T_{1\to2}^{AT} \cap T_{1\to2}^{RF} = \varnothing$.

Thus the question is, what is the relation between $\mu(T_{1\to2}^{AT})$ and $\mu(T_{1\to2}^{RF})$? According to the Miracle Argument, $\mu(T_{1\to2}^{AT}) >> \mu(T_{1\to2}^{RF})$ which justifies the claim that any T* that produces use-novel predictions is with high probability (approximately) true.

However, it is not too difficult to construct model examples that show that the Miracle Argument is unsound given transient underdetermination. In the mathematical example that I discuss in the appendix, the sets $T_{1\to2}^{AT}$ and $T_{1\to2}^{RF}$ and the corresponding measures $\mu(T_{1\to2}^{AT})$ and $\mu(T_{1\to2}^{RF})$ can be explicitly analyzed. First, it turns out that $T_{1\to2} \equiv T_2$. In other words: The set of theories that are constructed in order to fit the data D_1 and are – unexpectedly – also able to predict the use-novel data N, is the same as the set of theories that have been constructed to fit the data $D_2 := D_1 \cup N$ in the first place. Second, transient underdetermination does not only hold for T_1, but also for T_2. In other words, $\mu(T_2^{AT}) << \mu(T_2^{RF})$. Third, because of $T_{1\to2} \equiv T_2$, it also holds that $T_{1\to2}^{AT} \equiv T_2^{AT}$ and $T_{1\to2}^{RF} \equiv T_2^{RF}$. Therefore, forth, $\mu(T_{1\to2}^{AT}) << \mu(T_{1\to2}^{RF})$. In other words, a theory T* that was constructed to fit the data D_1 and is – unexpectedly – also able to predict the use-novel data N, is still very likely radically false, contrary to what the Miracle Argument claims. Fifth, in the mathematical example discussed in the appendix the picture does not even change for multiple use-novel predictions. In the model case, a theory that has repeatedly produced independent use-novel predictions is still very likely radically false.

Defenders of the Miracle Argument may object that the mathematical example I give in the appendix is so artificial and contrived that it has no real meaning for actual science whatsoever. I strongly disagree. I think that the situation of classical physics, for instance in the 18th and 19th century, bears structural similarity to my idealized mathematical example. In these two centuries, classical physics produced one use-novel prediction after the other. As is well-known, some physicists were so impressed that they believed that the fundamental principles of physics were secured once and for all, and only some mopping up work was left. However, as we now know, all these repeated predictive successes were no indicators for success (or even truth), regarding areas outside of contemporary technical reach (there may even be a more general lesson to be learned about (heuristic) predictivism, see (Harker 2008)). On three fronts, classical physics collapsed: at very high velocities, at very strong gravitational fields, and at very small energies. This situation is completely general. If some physical theory is extremely successful regarding the prediction of use-novel phenomena within a certain range of accuracy, nothing can be inferred about this theory's potential, let alone its truth, when the experimental accuracy is increased by one, by two, or by twenty-six orders of magnitude.

7. *Why does the Miracle Argument appear to be so plausible?* The obvious question now is, how can we account for the impression that despite its existing fundamental weakness, the Miracle Argument appears to be so plausible?[4] I think that its plausibility derives from two (hidden) presuppositions that are built in to the Miracle Argument. Let us look at Hilary Putnam's classic statement of the Miracle Argument: "The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle" (Putnam 1975, 73). Firstly, note that the question, Why is science successful? (in the sense of use-novel predictive success) is a *general* question concerning *all* cases of successful use-novel predictions. Note further that in his statement, Putnam presupposes that this question has a *uniform* answer, i.e., an answer on the level of a philosophy. In other words, Putnam immediately looks for an answer in terms of some very general conception of science, and not in terms of a detailed case-to-case analysis that may yield different answers for different cases. Clearly, this is a dubitable presupposition. Secondly, by naming his candidate answer, realism, Putnam implies what pair of philosophies he has in mind, namely realism and anti-realism. Putnam suppresses the fact that the contrast to realism is not just one philosophy but

an extremely heterogeneous class of philosophies some of which might have the potential to explain the success of science while others might not. In Putnam's statement, these two presuppositions are taken for granted such that the possibility or even necessity of critical discussion of their appropriateness does not come to mind. Let us now discuss these presuppositions in turn.

The existence of a uniform answer to the question why some theories are successful in producing novel predictions, i.e., an answer that identifies a feature possessed by all those theories and only by them, is highly dubitable. There are several substantially different possibilities why a particular theory may appear to be capable of producing successful novel predictions. Firstly, in rare cases the source of successful novel predictions may indeed be a pure coincidence: a lucky combination of numbers or of mathematical features, a cancelling of falsities, an unjustified but nevertheless successful extrapolation, etc. Secondly, the supposed use novelty of the prediction may be the result of ignorance and not really a case of novelty. Looked upon with a better understanding of the very same theory or from the vantage point of a successor theory, what appeared use-novel at some point in time may later turn out to have been implicitly inbuilt into the theory. Thirdly, a theory may have relevant similarities to a later, empirically more successful theory that translate into successful use-novel predictions, in spite of the two theories' possibly fundamental ontological discrepancies (examples are presented and discussed, e.g., in (Carrier 1991) and (Lyons 2006)). Notice that from the point of view of the later theory, the earlier theory may be in some areas just a good numerical approximation to the later theory, in spite of being ontologically utterly divergent from it. Notice further that the later theory *does not have to be (approximately) true* for this situation to hold. Finally, in principle a theory may indeed be (approximately) true and therefore be able to make novel predictions. Thus, there are quite a few different possibilities why some theory may be capable of making successful use-novel predictions. The (approximate) truth of a theory is but one of several possibilities.

The second presupposition of the Miracle Argument is that even among the uniform answers, only the alternatives of realism and anti-realism are considered. However, as has been discussed in the literature, among the anti-realist positions there are variants available that are also able to explain the success of science – realism is just not the only game in town. For instance, Timothy Lyons has discussed possible alternatives competing with realism in their explanatory role for the novel success of theories (Lyons 2003). His view is that serious competitors to realism include empirical adequacy, strong surrealism, and modest surrealism; Lyons himself favors the latter account (Lyons (2002, 78-79); Lyons (2003, 900-901)). Be that as it may, my point here is that the Miracle Arguments gains much of its persuasive force by making invisible possible competitors to realism that are located within the set of anti-realist positions.

8. *Conclusion*. Given our result, under the (plausible) supposition of transient underdetermination the Miracle Argument as an argument for scientific realism fails. First, our idealized example suggested that due to transient underdetermination, theories capable of making successful use-novel predictions are nevertheless most likely radically false. Second, in the previous section I have more concretely sketched several ways in which radically false theories may nevertheless be capable of producing successful novel predictions. Thus, (approximate) truth is not a candidate for uniformly explaining novel predictive success. Third, if scientific realism does not explain novel predictive success, we cannot proceed to the second step in the Miracle Argument, namely, that it is justified to accept scientific realism (see section 1). Thus, on the supposition of transient underdetermination, the Miracle Argument for scientific realism collapses.

Acknowledgements. Fruitful comments on earlier drafts of this paper especially by Katie Plaisance and Stathis Psillos, but also by Martin Carrier, Mehmet Elgin, Eric Oberheim, Thomas Reydon, Howard Sankey, Kyle Stanford, Marcel Weber, and a meticulous anonymous referee are gratefully acknowledged.

Appendix

Let us assume that the unknown "true theory" is a real-valued function $y = f(x)$, defined for $0 \leq x \leq 10$. Let us further assume that we have a set of functions $f_{a1, \ldots, a100}(x)$ with 100 free parameters a_i and that $f(x)$ can also be represented in this form (think, for example, of polynomials of degree 100, i.e., functions of the form $f_{a1, \ldots, a100}(x) = \sum_{i=0,\ldots,100} a_i\, x^i$). We want to get the true function $f(x)$ by fitting the parameters a_i to given data. For simplicity, let us assume that the parameters to be considered have a lower bound L and an upper bound U, i.e., $\forall_{i = 1, \ldots, 100}\ L \leq a_i \leq U$. We are thus considering a set of functions $T_0 := \{f_{a1, \ldots, a100}(x), L \leq a_i \leq U, i = 1, \ldots, 100\}$. Any such function $f_{a1, \ldots, a100}(x)$ is thus described by the values of the 100 parameters, and any set of such functions is described by a set of values of the 100 parameters. A natural measure for such a set of functions $f_{a1, \ldots, a100}$ is the Euclidean volume in the respective 100-dimensional parameter space. More explicitly, let F be a set of functions $f_{a1, \ldots, a100}(x)$ characterized by a set P of parameter values, then the measure μ_{100} (the Euclidian volume in the 100-dimensional parameter space) of the set F is defined as $\mu_{100}(F) := \int_P da_1 \cdots da_{100}$. For instance, the measure of all functions $f_{a1, \ldots, a100}$ in a certain parameter range $\Delta a = (\Delta a_1, \ldots, \Delta a_{100})$ (a 100-dimensional cuboid) is $\int_{\Delta a} da_1 \cdots da_{100} = \prod_{i=0,\ldots,100} \Delta a_i$. Due to our restriction on the values of the parameters a_i by the lower bound L and the upper bound U, the total measure of the function space T_0 is finite: $\mu(T_0) = (U - L)^{100}$.

Let us now assume that for 10 different values of x between 0 and 1, we have correct measurements of $f(x)$, i.e., we have correct data points (x_j, y_j), $j = 1, 2, \ldots, 10$. Thus $D_1 := \{(x_j,y_j), 0 \leq x_j < 1, j = 1, 2, \ldots, 10\}$. Now we look for all functions $f_{a1, \ldots, a100}(x)$ from T_0 that fit the data set D_1, i.e., these functions must fulfill the ten equations $f_{a1, \ldots, a100}(x_j) = y_j$, $j = 1, 2, \ldots, 10$. These are 10 equations for the 100 parameters a_i. By means of these equations, we can eliminate 10 parameters, say $a_{91}, \ldots a_{100}$, by means of expressing them in terms of $a_1, \ldots, a_{90}$.

Having done so, we have a set of functions $g_{a1, \ldots, a90}(x)$ that by their construction fit the data set D_1.[5] Thus, $T_1 = \{g_{a1, \ldots, a90}(x), g_{a1, \ldots, a90}(x_j) = y_j, j = 1, 2, \ldots, 10, L \leq a_i \leq U, i = 1, \ldots, 90\}$ contains all theories relevant for and consistent with D_1. Clearly $T_1 \subset T_0$.

Do we have a case of transient underdetermination in the relevant sense? Remember that TU holds iff $\mu(T_1^{AT}) << \mu(T_1^{RF})$. Now we have to be careful about our choice of a measure. If we take our original measure $\mu_{100} := \int da_1 \cdots da_{100}$, both T_1^{AT} and T_1^{RF} have measure 0. The reason is that $\mu_{100}(T_1) = 0$ because the parameter space of T_1 is a 90-dimensional hypersurface in a 100-dimensional space, and T_1^{AT} and T_1^{RF} are subsets of T_1. Instead, we have to use the measure $\mu_{90} := \int da_1 \cdots da_{90}$ that yields a non-zero measure for T_1: $\mu_{90}(T_1) = (U - L)^{90}$. The subset T_1^{AT} of T_1 is very small in comparison with T_1, because in the case of approximately true and true theories, the parameter values are allowed to vary only very slightly around the true value, say by a small amount Δa. Therefore, $\Delta a << U - L$. We get

$$\mu_{90}(T_1^{AT}) = \Delta a^{90} << (U - L)^{90} = \mu_{90}(T_1)$$

As $\mu_{90}(T_1^{RF}) = \mu_{90}(T_1) - \mu_{90}(T_1^{AT})$, we have indeed transient underdetermination in the relevant sense: $\mu_{90}(T_1^{AT}) << \mu_{90}(T_1^{RF})$.

Let us now consider use-novel predictions. So far, we have used the data $D_1 = \{(x_j,y_j), 0 \leq x_j < 1, j = 1,\ldots, 10\}$ in order to restrict the range of all available functions from T_0. Now we bring in additional data, namely 10 correctly measured values y_j of $f(x)$ in the interval $1 \leq x < 2$. The set of data points $\{(x_j,y_j), 1 \leq x_j < 2, j = 11, 12, \ldots, 20\}$ has obviously not been used in the construction of T_1 and thus represents use-novel data (relative to D_1 and T_1). In other words, the novel data set N is defined as $N := \{(x_j,y_j), 1 \leq x_j < 2, j = 11,\ldots, 20\}$. We are now interested in functions T* from T_1 that despite their having been constructed only out of data measured in the interval $0 \leq x < 1$, are still capable of correctly predicting the data from the interval $1 \leq x < 2$. Do such functions ('theories') exist, and if so, are they approximately true?

Let us first identify the subspace of T_1 in which possible T*'s are located. First, we consider the set T_2 of all theories that are relevant for and consistent with the new data set $D_2 := D_1 \cup N$. T_2 is constructed out of all functions $f_{a1, \ldots, a100}(x)$ in exactly the same way as T_1 was constructed. The only difference is that now we have to fit the functions $f_{a1, \ldots, a100}(x)$ to D_2 instead of to D_1 (see above). As we now have 20 data points, the result is that T_2 consists of functions with only 80 free parameters: $T_2 = \{h_{a1, \ldots, a80}(x), h_{a1, \ldots, a80}(x_j) = y_j, j = 1, 2, \ldots, 20, L \leq a_i \leq U, i = 1, \ldots, 80\}$. Clearly, $T_2 \subset T_1$. *If* T*'s exist, they are certainly located in T_2 (they fit all 20 data points), yet we cannot say more. Let us therefore consider the set $T_{1\to2} := \{T, T \in T_1$, T is predictively successful with respect to N$\}$ that picks out those elements of T_1 that produce the use-novel prediction N. We can construct $T_{1\to2}$ out of T_1 by imposing on *its* members the 10 additional conditions $g_{a1, \ldots, a90}(x_j) = y_j$, $j = 11, \ldots, 20$. Again, by these equations we can eliminate 10 parameters and obtain $T_{1\to2} = \{k_{a1, \ldots, a80}(x), k_{a1, \ldots, a80}(x_j) = y_j, j = 1, 2, \ldots, 20, L \leq a_i \leq U, i = 1, \ldots, 80\}$. Note that in the definition of this set the ability of its members of having produced the use-novel prediction N is not explicit. But this ability is inbuilt in its construction because $T_{1\to2}$ just filters out those members of T_1 that are indeed able to produce the use-novel prediction N.

What is the relation between T_2 and $T_{1\to2}$? The members of T_2 were fitted to accommodate the 20 data points $(x_{j,}, y_j)$, $0 \leq x_j < 2$, $j = 1, \ldots, 20$. By contrast, the members of $T_{1\to2}$ were only fitted to accommodate the 10 data points in the interval $0 \leq x_j < 1$, $j = 1, \ldots, 10$, but were nevertheless capable of correctly predicting the other 10 data points in the interval $1 \leq x_j < 2$, $j = 11, \ldots, 20$. Clearly, $T_{1\to2} \subset T_2$ because the members of the former set fit all 20 data points. But perhaps surprisingly, also $T_2 \subset T_{1\to2}$ holds. Take some theory $T' \in T_2$. Clearly T' is also a member of T_1 because $T_2 \subset T_1$. If you scan theories in T_1 and you hit upon T', you can certainly use it to correctly predict the data points $(x_{j,}, y_j)$, $j = 11, \ldots, 20$ in the interval $1 \leq x_j < 2$ because T’ $\in T_2$. Therefore, T’ $\in T_{1\rightarrow2}$. Thus, also $T_2 \subset T_{1\rightarrow2}$ holds.[6] We finally obtain $T_{1\rightarrow2} \equiv T_2$.

Now we have to discuss the proportion of (approximately) true theories and radically false theories in $T_{1\rightarrow2}$ ($\equiv T_2$). Note first that $T_{1\rightarrow2}^{AT} \equiv T_2^{AT}$ and that $T_{1\rightarrow2}^{RF} \equiv T_2^{RF}$. Thus, instead of discussing $T_{1\rightarrow2}^{AT}$ and $T_{1\rightarrow2}^{RF}$, we may discuss the identical sets T_2^{AT} and T_2^{RF}. As we have only 80 free parameters, we have to use the measure μ_{80}. By the same reasoning that we applied earlier in the discussion of the measures of T_1 and its subsets T_1^{AT} and T_1^{RF}, we obtain

$$\mu_{80}\ (T_2^{AT}) = \Delta a^{80} << (U - L)^{80} = \mu_{80}(T_2)$$

and

$$\mu_{80}(T_2^{AT}) << \mu_{80}(T_2^{RF}),$$

i.e., transient underdetermination holds in the relevant sense. Because of $T_{1\rightarrow2}^{AT} \equiv T_2^{AT}$ and $T_{1\rightarrow2}^{RF} \equiv T_2^{RF}$, we also get

$$\mu_{80}(T_{1\rightarrow2}^{AT}) << \mu_{80}(T_{1\rightarrow2}^{RF}).$$

Thus, any theory T* from T_1 that happens to produce the use-novel prediction N (i.e., T* is also a member of $T_{1\rightarrow2}$) is with high probability radically false, contrary to what the Miracle Argument claims.

Note that in our exemplary case even several repeated, independent novel predictions do not change the picture. A theory T* from T_1 (it was fitted to the data in [0, 1)) that happens to produce 10 use-novel predictions not only in the interval [1,2), but also in [2, 3), then in [3, 4), then in [4,5) and finally even in [5, 6), is nevertheless very probably radically false. The reason is also intuitively clear. We have a very large number of different functions defined in the interval [0. 10]. When we pick out those functions that agree with the true function f (x) at 10 data points in the interval [0, 1), these functions may still be *very* different from f (x) in the remaining interval [1, 10] of the total interval. Even restricting the set of those functions to those that happen to reproduce 10 data points each in [1,2), in [2, 3), in [3, 4), in [4,5), and even in [5, 6), nothing can be predicted about their behavior in the remaining interval [6, 10]. Because we have still 40 free parameters left, most of those functions will do whatever they want and are certainly not even near any true data points (x, f(x)) in [6, 10]. Thus, we may not expect that even theories that have been fitted to some data and are capable of multiple use-novel predictions, are (approximately) true. Use-novel predictions are by no means good indicators for (approximate) truth if transient underdetermination obtains.

References:

Carrier, M. (1991). What is Wrong With the Miracle Argument? Studies in the History and Philosophy of Science 22, 23-36

Earman, J. (1992). Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory. (Cambridge: MIT Press)

Harker, D. (2008). On the Predilections for Predictions. British Journal for the Philosophy of Science 59 (3), 429-453

Hempel, C.G. (1965). Aspects of Scientific Explanation. (In Aspects of Scientific Explanation and other Essays in the Philosophy of Science (pp. 331-496). New York: Free Press.)

Leplin, J. (1997). A Novel Defense of Scientific Realism. (Oxford: Oxford University Press)

Lyons, T.D. (2002). Scientific Realism and the Pessimistic Meta-Modus Tollens. (In S. P. Clarke and T. D. Lyons (Eds.), Recent Themes in the Philosophy of Science: Scientific Realism and Commonsense (pp. 63-90). Dordrecht: Kluwer.)

Lyons, T.D. (2003). Explaining the Success of a Scientific Theory. Philosophy of Science, 70, 891-901

Lyons, T.D. (2006). Scientific Realism and the Stratagema de Divide et Impera. British Journal for the Philosophy of Science, 57, 537-560

Musgrave, A. (1999 [1988]). The Ultimate Argument for Scientific Realism. (In Essays on

Realism and Rationalism (pp. 52-70). Amsterdam: Rodopi (originally in Robert Nola (Ed.), Relativism and Realism in Science. Dordrecht: Kluwer Academic, 1988))

Psillos, S. (1999). Scientific Realism: How science tracks truth. (London: Routledge)

Putnam, H. (1975). Mathematics, Matter and Method. Philosophical Papers, Vol. 1. (Cambridge: Cambridge University Press)

Quine, W.V.O. (1951). Two Dogmas of Empiricism. The Philosophical Review 60 (1), 20-43

Severo, R. (2008). "Plausible insofar as it is intelligible": Quine on underdetermination. Synthese 161 (1), 141

Sklar, L. (1975). Methodological Conservatism. Philosophical Review 84 (3), 374-400

Stanford, P. K. (2000). An Antirealist Explanation of the Success of Science. Philosophy of Science 67 (2), 266-284.

Stanford, P. K. (2006). Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives. Oxford: Oxford University Press.

Stanford, P. K. (2009). Scientific Realism, the Atomic Theory, and the Catch-All Hypothesis: Can We Test Fundamental Theories Against All Serious Alternatives? British Journal for the Philosophy of Science.

van Fraassen, B. C. (1980). The Scientific Image. Oxford: Clarendon.

Worrall, J. (1985). Scientific discovery and theory-confirmation. In J. C. Pitt (Ed.), Change and progress in modern science (pp. 301-332). Dordrecht: Reidel.)

Worrall, J. (1989). Fresnel, Poisson, and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories. (In D. Gooding, T. Pinch and S. Schaffer (Eds.), The Use of Experiment: Studies in the Natural Sciences (pp. 135-157). Cambridge: Cambridge University Press.)

[1] The condition of the finiteness of D_1 could be relaxed to include infinite sets but this is of no concern here.

[2] N may even be a set of data resulting from a series of experiments repeatedly generating novel data; I will revisit this point later.

[3] We can analyze the tension between Argument 1 and Argument 2 also as a case of what Hempel has called "the ambiguity of statistical explanation" (Hempel (1965, 394)), where two statistical arguments with true premises have contradicting conclusions. This tension is resolved because, firstly, the conclusion of Argument 1 does not strictly exclude any T from being an element of T_1^{AT}. Secondly, Argument 2 provides us with *more specific* information about T* than argument 1. According to Hempel's "requirement of maximal specificity" (Hempel (1965, 397-401)), we have to accept Argument 2 instead of Argument 1 because Argument 2 refers to a smaller reference class (to only those $T^* \in T_1$ that make the novel prediction N) than does Argument 1. Thus, due to its higher specificity, Argument 2 overrules Argument 1.

[4] The following argument is inspired by personal communication with P. Kyle Stanford in April 2006 and by the footnote 8 on p. 274 of (Stanford 2000).

[5] The relation between the functions $g_{a1, \ldots, a90}(x)$ and the functions $f_{a1, \ldots, a100}(x)$ is the following. Express the parameters $a_{91}, \ldots a_{100}$ in terms of the parameters $a_1, \ldots, a_{90}$, i.e., $a_i = a_i (a_1, \ldots, a_{90})$, $i = 91, \ldots, 100$ and in this way eliminate the parameters $a_{91}, \ldots, a_{100}$ in $f_{a1, \ldots, a100}(x)$. Thus: $g_{a1, \ldots, a90}(x) = f_{a1, \ldots, a90, a91(a1, \ldots, a90), \ldots, a100(a1, \ldots, a90)}(x)$.

[6] Pragmatically, there is of course a difference between the members of $T_{1\rightarrow 2}$ and the members of T_2, as the anonymous referee rightly remarks: the members of $T_{1\rightarrow 2}$ are fitted to D_1 and are *then* discovered to correctly predict N, whereas the members of T_2 are fitted to $D_2 = D_1 \cup N$ *from the start*. Nevertheless, the two sets $T_{1\rightarrow 2}$ and T_2, are identical. As I am using their identity only to evaluate the sizes of some of their subsets, their pragmatic difference may be

legitimately ignored.