Epistemological Issues in the Manifest and the Scientific Images: An Introduction

Mario Alai and Francesco Orilia***

* *University of Urbino Carlo Bo*

** *University of Macerata*

1. The Manifest Image and the Scientific Image

The national PRIN 2017 project "The Manifest Image and the Scientific Image" of the Universities of Macerata, Florence, Rome 3 and Urbino (prot. 2017ZNWW7F_004) was launched in December 2019 to investigate a serious problem in our understanding of the discoveries of contemporary science, extremely deep and rich in momentous practical consequences, but often also very surprising and even paradoxical *vis a vis* our everyday experience of the world. This puzzle, already addressed by Eddington (1928), Husserl (1936) and Sellars (1962), was described by the latter as a clash between the "scientific image" and the "manifest image" emerging from common sense.

The goals of the project were accordingly fixed as follows: (a) Achieving a better understanding of the manifest image, also by recourse to experimental philosophy. (b) Getting a clearer grasp of the scientific image, especially in three areas: the sustainability of scientific realism concerning properties, relations and unobservables; the nature of time as emerging from current physics; the systems of formal logic introduced to achieve higher consistency than that provided by informal logic. (c) Investigating how the two images must be related from the logical, epistemological and metaphysical point of view if they are to be understood as compatible, in spite of their *prima facie* incompatibility.

Over the last three years these goals have been pursued by the investigators and by a number of collaborators to the project. This resulted, *inter alia*, in a number of papers presented and discussed at the two general conferences of Florence (November 29-30, 2021) and Urbino (June 20-21, 2022). This issue of *Argumenta* collects the investigations conducted from a broadly epistemological point of view, while a previous issue (*Logical and Ontological Issues in the Manifest and the Scientific Images*) collects the articles largely dealing with logical and ontological matters.

Of the papers collected here, five (those by Buonocore and colleagues, Galli, Zorzato, Cevolani-Tambolo, and Savojardo) are devoted to goal (b), viz., better understanding the scientific image and how it can be supported (i.e., with the issue of scientific realism). In addition, Cevolani-Tambolo and Savojardo also deal

Argumenta 9, 1 (2023): 9-25
ISSN 2465-2334

DOI 10.14275/2465-2334/202317.int

with the relations between the two images (goal c). Finally, the article by Angelucci and colleagues contributes both to understanding and supporting the manifest image (goal a) and to clarifying the relations between the two images (goal c).

Three of these papers discuss issues in general philosophy of science (Cevolani-Tambolo, Buonocore and collaborators, and Zorzato), while those by Angelucci and collaborators, Galli, and Savojardo concern the special philosophy of three sciences (respectively, evolutionary game theory, artificial intelligence and the neurosciences). Moreover, both Galli and Zorzato discuss the nature and function of models in science.

2. Buonocore, Margoni, Pero: "Conceiving the Inconceivable: An Assessment of Stanford's New Induction"

In recent years Kyle Stanford (2006) has introduced a new powerful argument against scientific realism, the so-called argument from "*unconceived alternatives*" (UA). He points out that

> we have, throughout the history of scientific inquiry and in virtually every scientific field, repeatedly occupied an epistemic position in which we could conceive of only one or a few theories that were well confirmed by the available evidence, while subsequent inquiry would routinely (if not invariably) reveal further, radically distinct alternatives as well confirmed by the previously available evidence as those we were inclined to accept on the strength of that evidence (2006: 19).

This we know because some of those alternatives were subsequently conceived of and found to be better (more probably or approximately true) than the previously accepted theories. It follows then by a natural induction that even today we fail to conceive theories which are better and more probably true than our own, and therefore that our theories are probably false.

Buonocore, Margoni and Pero explain that Stanford's UA argument draws on two classical antirealist arguments, those from the empirical underdetermination of theories and from the pessimistic meta-induction, although allegedly improving on both. It relies on the idea that different theories can be proposed to account for the same evidence. Stanford's argument, however, is less demanding than the classical underdetermination argument, because it does not require alternatives to be empirically equivalent, but simply empirically equally well confirmed. Therefore, the UA argument does not construct its alternatives "parasitically so as to perfectly mimic the predictive and explanatory achievements of our own theories" (Stanford 2006: 18-19) like the traditional underdetermination argument, but it points at genuine theoretical alternatives that, while unconceived up to a certain time, subsequently were actually adopted. Therefore, the UA argument is based on historical evidence, rather than on philosophical speculation.

One might object, we suppose, that this actually weakens the argument, because the very historical evidence showing that certain better alternatives were ignored at earlier times, also shows that they were recognized a later time, so perhaps we might conclude that in the long run the best alternatives (the more approximately true ones) will be found. If not, at least our theories are becoming better and better confirmed, hence, arguably, more and more approximately true and/or verisimilar.

A reply may come from the other strain in Stanford's argument: it is inductive and pessimistic, but unlike the traditional pessimistic meta-induction, his "new induction" focuses on theorists rather than on theories. In fact, it starts by noticing that at all past times scientists had cognitive (intellectual, psychological or sociological) limits, which prevented them from conceiving some better alternatives; then, it inductively argues that scientists will always have similar limits, so concluding that they will always miss many better alternatives. Again, however, a natural rejoinder is that history shows that those limits can be overcome, and they do not prevent us from continuously progressing toward the truth.

Buonocore, Margoni and Pero argue that Stanford's thesis has various problems, which surface once we ask the question:

(Q) Stanford says that at any time *t* many alternatives were empirically well supported but remained unconceived: but were they conceivable o unconceivable at that time?

By 'conceivable' they mean a theory which one could have conceived given the accepted evidential, theoretical, methodological, or metaphysical presuppositions of the time, but remained unconceived because of the subjective intellectual, psychological, or sociological limits of those scientists.

Unfortunately, Stanford does not offer an explicit answer to (Q), but according to the authors textual evidence suggests that he probably thinks of *conceivable* alternatives, for he writes that scientists

> repeatedly *failed* to conceive of scientifically serious and well-confirmed alternatives to their own proposals. [… Such alternatives] were *scientifically serious* even by the standards of the day despite being unconceived and therefore unconsidered by theorists at the time (Stanford 2006: 60; italics added).

Moreover, this failure persisted even "after we came to embrace substantive evidential, metaphysical, and methodological constraints essentially continuous with those of the present day" (2006: 60).

It might be objected that 'scientifically serious' is different from 'conceivable', and in the above quotation from p. 19, Stanford says that at *t* scientists "could" conceive only *one* alternative, so implying that the others could *not* be conceived. Yet, the authors might reply that for Stanford scientists could not conceive those alternatives because of their own personal or sociological limitations, not because those theories were unconceivable. Besides, the mention of "substantive evidential, metaphysical, and methodological constraints" suggests that Stanford has in mind just what the authors mean by 'conceivability'. Furthermore, they notice that one could not talk of a "failure" in conceiving something if it was, in fact, inconceivable.

Actually, it is not clear that by 'fail' Stanford means unsuccess, rather than simple omission or neglect.[1] Even aside from textual evidence, however, it seems that Stanford *should* be concerned precisely with conceivable theories, if his argument must be distinguished from the old pessimistic induction. In fact, if the reason why a theory remained unconceived were that it was unconceivable, then Stanford's argument would be again an induction over theories, like the old pessimistic induction ("at each past time there were many better but inconceivable

[1] In fact, unlike the Italian verb '*fallire*', the English 'to fail' has both these meanings, the evaluative one and the neutral one.

theories, so this is also happening now"), rather than a "new" induction on the limits of scientists, as it is supposed to be ("at each time *t* scientists proved unable to conceive theories which were conceivable at *t*").

However, Buonocore and friends argue that if actually Stanford's argument applies to conceivable UA, as they suggest, then it hinges on conceivability as an *a-temporal* property of *theories*, because it depends only on the relation of theories with evidence, rather than on the *temporal* limits of the *scientists*; in this sense, it is still an induction on theories, after all, rather than on theoreticians, as Stanford claims.

An even more serious flaw is that, if so, his argument cannot be applied to various remarkable cases of theory change, where clearly the theory that in time would supersede the accepted one was not even conceivable. Thus, the scope of the argument would be seriously limited, it would no longer support the generalized antirealist conclusion that *at any time t* scientists fail to conceive better theories that were serious (i.e., conceivable) alternatives even at *t*. For instance, the authors argue that, contrary to Stanford's claim, in Newton's time the Special Theory of Relativity was unconceivable, for lack of those theoretical, empirical and methodological constraints which turned out to be essential to Einstein's theory.

Therefore, since if Stanford refers to *conceivable* UA, he encounters such problems. However, since his answer to question (Q) is not explicit, Buonocore and collaborators also explore what would follow if Stanford instead referred to *unconceivable* UA (or to both conceivable and unconceivable UA). In that case, they argue, his argument would just be that at any time there are countless possible alternatives (conceivable or unconceivable), among which only one is true, and we will never be able to consider all of them in order to choose the true one. Hence, it would no longer concern a *transient but recurrent* underdetermination, as he says, but it would boil down to the traditional argument from (permanent) underdetermination. Besides, realists could argue that even if a better theory was not conceived at time *t* because the lack of the necessary evidential, methodological, and metaphysical presuppositions made it unconceivable, later on, when those presuppositions become available, it will become conceivable, and so it will probably be conceived.

One might worry that nonetheless the one true theory will escape forever, and certain moderate realists are ready to grant this possibility; however, as noticed above, history shows science is progressive, as the successively conceived alternatives are better and better approximations to the truth, and this will satisfy most current realists.

Summing up, the authors question the novelty of Stanford's argument, for no matter whether his UA are conceivable or unconceivable, his induction actually concerns theories, like the old pessimistic induction, rather than theorists, as he suggests.[2] Moreover, if his UA are conceivable the argument does not apply to some of the most important instances of theory change, where the superseding theory was not conceivable until it was actually conceived. If instead Stanford's UA are unconceivable, then his argument does not really differ from the classical underdetermination argument, and it is effective only against certain implausibly strong versions of realism. Of course, a third answer to question (Q) is possible,

[2] For a further reason why Stanford's "new induction" is not any stronger than the old pessimistic induction, see Alai 2019: §3.

viz., that Stanford doesn't care, as he intends to apply his argument to *both* conceivable *and* unconceivable UA. If so, however, the argument has the drawbacks relating to conceivable UA when it applies to them, and those relating to unconceivable UA when it applies to them.

3. Galli: "Structure Representation of Deep-Learning Models: the case of AlphaFold"

The scientific image of the world is largely drawn by using models. Models are a standard tool of scientists, perhaps their main tool, when theoretical science is concerned. Thus, to understand the scientific image and its relations to the manifest image and to how the world actually is, it is mandatory to understand what models are and how they work, i.e., how they represent. Yet, there are various kinds of models and various ways of understanding the very concept of model. No wonder, then, that this topic is so widely discussed in the philosophy of science, and that two of our papers in this issue are concerned with it.

While Zorzato focuses on a particular kind of theoretical models (the so called "fictional" ones), Galli analyses the models produced by a non-human scientist, the deep-learning neural network system AlphaFold, which has proven so successful in predicting the structure of proteins and in other tasks. Even for him, however, the basic question is still the nature of the representational relation between these models and what they represent, and whether it supports scientific realism.

Preliminarily, Galli presents an interesting taxonomy of kinds of models and of different conceptions of the models' function in science: he distinguishes a *similarity conception*, according to which models represent their target systems by being similar to them; an *inferential conception*, according to which the value of models is mainly pragmatic, consisting in the inferences they can license; and a *structuralist conception*, according to which models represent in virtue of an isomorphism they bear to their targets.

He then discusses Knuuttila's (2021) artifactual view of models, a variant of the inferentialist conception, motivated by the fact that one can model not only real existing systems, but also systems which are merely potential or not yet existing. In the latter case, it would seem, models are better seen as artifacts, i.e., as tools for investigating specific phenomena and answering scientific questions. According to Knuuttila, their function is that of exploring the spaces of possibilities and their success needs not be explained by a representational relation holding between them and a target system (which in fact does not exist in this case). Still, the question of what makes one such model successful or unsuccessful is left open.

Subsequently, Galli explains what AlphaFold is, how it works, and which kinds of models it produces. What a protein can do does not depend only on the sequence of amino acids by which it is composed, but, very importantly, also on the way its string of amino acids folds in space. Therefore, AlphaFold can produce 3-dimensional models of proteins, starting from the mere sequences of amino acids. These can be models of actually existing proteins, but also of merely possible proteins. While in the former case the representation relation is given by an isomorphism obtaining between the model and its real target, in the latter case it is given by the fact that the merely possible protein represented by the model exhibits a certain number of modal properties which in fact characterize actual proteins. Therefore, Galli holds that, contra Knuuttila's, even in this case models bear a (structural) representation relation to real systems. A robust form of scientific realism is thus implicit in his account.

4. Zorzato: "Fiction and Reality: An Uncanny Relationship"

While Galli analyses the models generated by the AlphaFold neural network, Zorzato focuses on a particular kind of theoretical models (the so called "fictional" ones). Antirealists suggest we discard the scientific *image* on the ground that it does not represent actual non-observable reality, or not correctly anyhow. One reason has always been the use of abstraction and models in science, because models are not exact replicas of their intended targets (the real systems they are meant to model): they are not complete replicas, since they are abstract, leave something out, and they are not correct replicas, since some of their features do not correspond to features of their targets. To this, however, realists reply that, as pointed out by Mary Hesse, all models include positive analogues (features we know to reproduce features of the targets), negative analogues (features we know not to reproduce features of the targets) and neutral analogues (features about which we ignore whether they reproduce features of the targets or not). Thus, we use a model only to the extent that it allows us to offer an accurate picture of the target: in describing, explaining, and predicting, we use the positive analogues, discard the negative analogues, and in advancing research we probe the neutral analogues in order to find out whether they are actually positive or negative, so discovering new features of the target.

The so-called "fictional" models, however, defy this defense of realism. They are models in which not only positive analogues (and, tentatively neutral analogues) are exploited to describe, explain, or predict, but also negative analogues. It would seem to follow, therefore, that the resulting descriptions, explanations, or predictions are false. As an example, Alisa Bokulich (2008) discusses the models produced by scientists for *Rydberg atoms*. These are certain light atoms excited to the point that their outermost electrons are at the threshold of ionization. As a result, their size becomes enormous, approaching the dimensions of minute macroscopic particles. Thus, they can be considered as sitting on the threshold between quantum and classical objects. In fact, the spectral lines emitted by these atoms in strong magnetic fields cannot be explained by current quantum theory; instead, they turn out to be nicely explained by assuming that electrons travel on classical orbits (Main et al. 1986). Furthermore, starting from the experimentally observed spectrum, it proves possible to reconstruct the corresponding orbits as described by the classical theory.

Of course, we know that electrons *do not* travel classical orbits, so this is clearly a negative analogue in the model built to account for Rydberg atoms. According to Bokulich, therefore, we cannot be realist about a model of this kind. Yet, it is successfully used to explain and it even allows some sort of prediction; besides, since the atom approaches the dimension of classical particles, something seems to suggest that there might be some truth to it. Therefore, says Bokulich, neither should the model be interpreted instrumentalistically, as a mere calculation device: what is called for is a "moderate" version of realism.

This compromise, however, has been criticized, first of all because it is not clear what exactly "moderate" realism should be, and how it differs from standard scientific realism; besides, it cannot explain how the model, being fictional, can represent reality and explain: whatever the model may achieve in this respect, if anything, must be parasitic on the theory.

In response, in her contribution to this issue Lisa Zorzato argues that Bokulich's account of fictional models is largely correct, but it doesn't call for any weakening of realism: the use of such models can be explained by "mainstream" scientific realism just like that of ordinary models. By 'mainstream realism' she understands the position of Psillos (1999), in short, the claim that at least some components of scientific theories can be justifiably believed to be true in the correspondence sense of truth.

In order to appreciate her argument, it should be remembered that authors as diverse as Poincaré (1905), Carnap (1927) and Schlick (1938) insisted that knowledge, in particular scientific knowledge, is structural, and Wittgenstein's *Tractatus* (1921) shows that linguistic representation itself is essentially structural (i.e., we can know only the relations among things or parts of things, hence we can know the intrinsic nature of complex things only to the extent that it is given by the relations of their parts, while we ignore the intrinsic nature of simple things). Therefore, the correspondence which for realists exists between representations and reality is a structural correspondence.

Zorzato does not say whether she agrees with contemporary structural scientific realism that scientific theories can represent only structures or not; but in any case, nobody would question that at least *some* scientific knowledge in the realist sense is structural. Now, she points out that real natural systems can be represented by models at a number of hierarchically ordered levels of abstraction (what she calls "the ladder of abstraction"). More precisely, there can be positive analogues in a model at different abstraction levels: this is to say, there are various more or less abstract structures in a model, which in the successful cases structurally correspond to respectively more or less abstract structures of the target system. Schematizing, the model can have a feature at Level 2 which is false of the target's structure at Level 1, but true of its structure at Level 2. Now, this is enough for "mainstream" realism.

For instance, in the case of Rydberg atoms, Zorzato claims that, while the classical orbits of the model are fictional (since there are no such orbits in reality), they play an explanatory role with respect to the behavior of the electrons, because their structure at a certain level of abstraction corresponds to certain structures of the real atomic spectra. In other words, at the level at which orbits are understood just like those of the planets, the model is false. But at a more abstract level, where only certain selected structures of the orbital behavior are considered, those structures can precisely match certain patterns of the emission spectrum.

A possible concern, here, is that the emission spectrum is an empirical structure; hence, it might be objected, the model simply has an instrumental role, saving the phenomena. However, especially in view of the fact that Rydberg atoms resemble classical particles also in other respects (e.g., their size), it seems quite possible that further analysis identify structural correspondences also at a theoretical level. At any rate, further progress of research in this respect appears to be both desirable and possible.

An important take-home lesson, here, is that when the scientific image is given by models, we must distinguish between the literal picture offered by a model, and its *intended* picture, i.e., the structural one. Clearly, the literal picture is closer to the manifest image (because the model is often drawn from ordinary empirical knowledge, or at least from consolidated scientific results already incorporated in common sense); however, it is false. Thus, the scientific image can be considered as true only if identified with the intended structural picture offered by the model. Moreover, the latter picture bears a structural resemblance to the empirical data patterns, which are one of the facets of the manifest image.

5. Cevolani and Tambolo: "Empirical Success, Closeness to Evidence, and Approximation to the Truth"

Empirical success is the success of a scientific theory or hypothesis in describing, organizing, explaining, and predicting experience, i.e., in accounting for empirical data, i.e., in entailing true empirical propositions. Henceforth it will be called simply "success". Given its empirical nature, anyway, it can be appreciated in a non-theoretical way, i.e., from the vantage point of the manifest image. Thus, it provides an interface between the scientific and the manifest image: when the latter is used as a benchmark for assessing the validity of the former, as in the debates on scientific realism, success figures as a necessary and most important requirement that hypotheses or theories are called to satisfy.

In fact, Cevolani and Tambolo explain that realists are committed to Laudan's (1981: 32-36) "downward path" (DP) and "upward path" (UP), i.e., respectively, the claim that true or approximately true hypotheses or theories are probably very successful, and that very successful hypotheses or theories are probably at least approximately true.

Scientific antirealists are also interested in success; since they deny that science provides theoretical knowledge, they understand the progress of science simply as the idea that science is growingly successful. Moreover, both realists and antirealists can account for the perduring value and utility of falsified hypotheses by pointing out at their success.

Popper held that we cannot ever know whether a hypothesis is true, but only, sometimes, recognize when it is false. Moreover, it's likely that hypotheses we hold now will be falsified in the future. However, even falsified hypotheses may be more or less "similar" to the truth: hypothesis H1 is more *verisimilar* than hypothesis H2 iff H1 has a true content larger than H2, or a smaller false content, or both. Furthermore, if we find that the subsequent and superseding hypotheses are more verisimilar than the earlier and superseded ones, then we know that there is progress in science (Popper 1963). Popper's idea was subsequently developed by a research tradition in which Oddie (1986), Kuipers (1987), Niiniluoto (1987, 1998), Festa (1982), and lately Cevolani himself have been prominent.

This tradition must face two main problems: first, how do you measure the content of a hypothesis? Intuitively, it can be spelled out as its logical strength, or the number of its consequences. Popper's original definition of verisimilitude had a fatal technical flaw which was exposed by Tichy (1974) and Miller (1974), essentially due to the fact that all propositions have infinite consequences. Niiniluoto, Kuipers, Oddie fixed this (roughly) by considering exclusively the number of the atomic propositions "relevant" to the hypothesis H which are entailed by it, and by relativizing the definition to a language. Cevolani and Tambolo's own definition of verisimilitude is

$$vs(H) = \frac{t}{n} - \frac{f}{n}$$

That is, H's verisimilitude (vs) is measured by the difference between the number of true atomic propositions *t* and of the false atomic propositions *f* entailed by H, both weighted by the total number *n* of atomic propositions of the language. In practice, the larger is the *proportion* of true propositions of the language entailed by H and the smaller is that of false propositions, the more verisimilar H is (see also Cevolani et al. 2011, 2013).

The second problem confronting the verisimilitude tradition is how to estimate how much of H's content is true and how much is false, i.e., the numbers of its true and false consequences, respectively. In fact, H's content (i.e., the *t* true propositions and the *f* false propositions entailed by H) includes: (1) H's empirical consequences which we observed to be (1a) true or (1b) false; (2) H's empirical consequences which we have not been able to observe to be true or false; (3) H's theoretical consequences. Thus, the truth-value of propositions in (1) is known, and propositions in (1a) constitute H's success. Instead, the truth-value of propositions in (2) and in (3) (which are many more than those in (1)) must be estimated, and this can be done first and foremost on the basis of H's success and failures, i.e., of the truth-value of propositions in (1).

In this estimation, therefore, success plays the key role; yet, it is a very difficult and risky extrapolation, since the propositions in (1) are so few in comparison with both the propositions in (2) and in (3), and so different in subject from the propositions in (3). Many realists hold that this task can be aided by considering also the "theoretical (or *nonempirical*) virtues" of H (see Alai 2019: §3.2), but antirealists contend that we will never have enough reasons to justify the claim that any consequence of H is true (or false), except for those in (1) (e.g., van Fraassen: 1980). This is therefore the main focus of contemporary discussions on realism.

Moreover, Cevolani and Tambolo explain that the idea of success is a vague one, and though there are different ways to explicate it precisely, none is completely satisfactory. Hempel (1948) characterized the success of a hypothesis or theory H as its "systematic power", viz. a measure of the proportion of the content of the available evidence E entailed by H. In other words, E is the set of all the atomic propositions of the language currently known to be true, and systematic power is a function of how many of those propositions H entails. Thus, even falsified hypotheses can be more or less successful: for instance, under this characterization a falsified hypothesis H1 turns out to be more successful of a non-falsified hypothesis H2 if H1 includes a wider proportion of E than H2,[3] which can happen when H1 is more informative than H2. On the other hand, this has the undesirable consequence that if H1 entails H2, then H1 is always at least as successful as H2: this is unacceptable, because, for instance, if H1 is built simply by adding to H2 some false or irrelevant claims, H1 is by definition as successful as H2.

This problem is avoided by Kuipers (2000), according to whom H1 is more successful than H2 iff (a) the confirming instances (the true empirical consequences) of H1 are at least as many as those of H2, (b) the disconfirming instances (the false empirical consequences) of H2 are at least as many as those of H1, and (c) H1 has at least one more confirming instance or one less disconfirming instance than H2. In this way, even if H1 entails H2 it may be less successful than H2, for the false or irrelevant surplus content of H1 with respect to H2 may (and typically will) have some disconfirming instances. Unfortunately, however, when success is so defined it becomes impossible for a falsified hypothesis H1 to be

[3] Since a false hypothesis may have some true consequences.

more successful than a non-falsified one H2, because H1 will have at least one more disconfirming instance than H2.

Cevolani and Tambolo take a clue from Zamora Bonilla's (1992, 1996) notion of "estimated truthlikeness" (i.e., roughly, similarity to the evidence), which is defined by him as directly proportional to the portion of the available evidence E entailed by H and inversely proportional to the "rigor", i.e., informativity, or improbability, of E (where E is the set of all the *m* empirical propositions relevant to H currently known to be true). That notion, however, has the drawback that all falsified hypotheses measure 0. Thus, in the present article, Cevolani and Tambolo define success as "similarity to the evidence *es*", where

$$es(H,E) = \frac{t_E}{m} - \frac{f_E}{m}$$

Here t_E is the number of propositions in E entailed by H (hence, its confirming instances), f_E is the number of propositions in E contradicting H (hence, its disconfirming instances), and m is the number of propositions in E. Therefore, the success of H is given by the difference between the ratio of the confirming instances t_E to the m elements of E, and the ratio of the disconfirming instances f_E to the m elements of E. In a nutshell, a successful hypothesis is one that entails a large proportion of the observations (the elements of E) and contradicts a small proportion of them.

This notion has all the advantages of those of Hempel, Kuipers and Zamora Bonilla, but none of their disadvantages: falsified hypotheses may still be successful, success is still directly proportional to the confirming instances and inversely proportional to the disconfirming instances, but a logically stronger hypothesis is not necessarily as successful as a weaker one. Besides, this notion seems to be precisely what scientific antirealists need to account for scientific progress merely in terms of increasing empirical success, without any realist presuppositions, i.e., without assuming that the theoretical content of hypotheses or theories is even approximately or partly true.

Scientific realists, instead, need a clear notion of success in order to argue that if H is approximately true, then it is very successful (the "downward path", DP) and, more importantly, if H is very successful, then it is probably approximately true (the "upward path", UP). UP, of course, is our best bet to estimate verisimilitude.

However, Cevolani and Tambolo show that, if success is constructed as *similarity to evidence* (*es* above), and verisimilarity as *vs* above, neither DP nor UP can be expected to hold in general. For instance, suppose that

(C1a) E is very poor, consisting of just the two propositions p_1, p_2, and suppose H is highly verisimilar.

Yet, quite possibly,

(C1b) H entails p_1 but contradicts p_2. In this case the above definition entails that the similarity of H to evidence equals ½ - ½ = 0; hence, H is highly verisimilar, but without success: DP fails.

Conversely, suppose that

(C2a) As before, E is very uninformative, e.g., consisting of just one proposition p_1, and

(C2b) H entails p_1 (and hence it is maximally successful) but it is extremely poor, to the point of coinciding with p_1 itself.

In this case, although H is maximally successful, its verisimilitude is very low. For instance, if there are 1,000 atomic propositions in the language, by the above definition of *vs*,

$vs\,(\mathrm{H}) = 1/1{,}000 - 0/1{,}000 = 0.001$

Thus, UP is violated in (C2).

Actually, a case like (C2a) is irrelevant to the current debates on scientific realism, for they concern only the possibility of justifying belief in the truth of theoretical hypotheses, while H here is merely empirical. Moreover, few if any realists believe we can show that any hypothesis is (more or less) *verisimilar* (i.e., that it entails most of the propositions of a language, i.e., that it tells a large part of what there is to know, or of "the whole truth"). They are quite content to argue that a hypothesis is (more or less) *approximately true*, i.e., that it is *largely*, or at least *partly*, true (i.e., that most or at least some of its consequences, both empirical and (especially) theoretical, are true—see Musgrave 2006-2007, Alai 2014b: 279-80), *irrespectively of how informative H is*, i.e., of how many propositions it entails: small is beautiful, if it is true. From this point of view, if H entails just itself, and it is true (or, say, if it entails just a theoretical proposition, itself, and an empirical one, and both are true), H is completely true (it has the maximal approximation to the truth), hence, UP works perfectly for it.

Nonetheless, that *es* and *vs* do not support DP and UP can be shown by different examples. For instance, suppose that

(C3a) As above, E consists of just one proposition p_1, correctly entailed by H, so that H's success is maximal, i.e., $\frac{1}{1} - \frac{0}{1} = 1$; still,

(C3b) The theoretical content of H is completely false, and H accounts for p_1 *by pure luck*, or simply because it has been purposefully imagined, or modified *ad hoc*, in order to accommodate p_1.

In this case, then, H is maximally successful, but neither verisimilar nor even slightly approximately true. Therefore, UP fails.

However, Cevolani and Tambolo's formalization shows what is missing from the notion of similarity to evidence *es* to support DP and UP: the definition of *es* is "completely silent on what E is [while] the precise relationship between H and E […] is obviously crucial to assess the success of H on E".

A few comments may be made on this conclusion. First, it might seem that it simply provides a formal confirmation of the intuitive and even commonplace idea that scarce evidence, even if favorable to a hypothesis, cannot confirm it (as required by UP), and that even highly verisimilar and approximately (i.e., largely) true hypotheses might not be successful (as required by DP) at the very first moment, but only in the long run.

This may be right, but there is much more. In fact, even if the body of the *available* evidence E were very large, it would be typically very small (hence of little statistical relevance) with respect to the infinite body UE of the *unavailable* evidence, which escapes us because it is remote in space or time, or beyond the reach of our senses, instruments, or experiments, etc. (UE may be understood as the set of all the true empirical propositions in the language relevant to H but not included in E, i.e. not confirmed by observation). Therefore, it might happen that

(C4a) H is a highly verisimilar and largely true hypothesis, which would get most of UE right, but

(C4b) H it is not successful, because its relatively few empirical failures happen to concern precisely E, hence DP fails.

Conversely, it might be the case that

(C5a) H is quite successful, getting all of E right. Yet,

(C5b) H has a very large and completely false theoretical content,[4] so that most of its empirical consequences are false: in fact, (unbeknownst to us) it contradicts all of UE. Therefore, H is successful but not even partly true,[5] nor verisimilar, and UP is violated.

All this indicates that the main trouble is the gap between *success*, which is empirical, and *truth*, which in the realism-antirealism debate is pre-eminently theoretical, i.e., unobservable, or at least unobserved. Now, the most promising strategy to bridge this gap seems to be one suggested (again) by Popper. In fact, while he initiated the research on *static* properties of hypotheses, like his *verisimilitude*, Hempel's *systematic power*, Kuiper's *empirical success*, and *similarity to evidence*, he also stressed the need to investigate their *dynamics*. An almost trivial example of how dynamic considerations may help in this respect is this: the counterexamples to DP and UP based on the extreme weakness of E (as in cases C1, C2 and C3) can be ruled out because in the absence of a consistent body of evidence to be accounted for, H would not have been proposed in the first place, since there wouldn't have been any need for it, nor enough empirical guidance to conceive it.

This is not all, however, since counterexamples to DP and UP can be envisaged even for large bodies of evidence, like in cases (C4) and (C5). Now, for instance, in a case like (C5a), how can we understand whether (C5b) also holds, i.e., UP is violated, or not? Well, if E was fully known and H was framed precisely to accommodate E, it is very likely that (C5b) holds (i.e., H is neither approximately true nor verisimilar), so that UP fails: in fact, by the principle of empirical underdetermination, there are countless possible *false* hypotheses and only a true one accounting for E. On the other hand, if E was completely unknown beforehand and genuinely predicted by H, by the "no miracles" argument it is overwhelmingly improbable that H is mostly or completely false (Alai 2014a). Hence, it is utterly unlike that (C5b) holds and UP is violated: on the contrary, UP supports the claims that H is at least partly true and to some extent verisimilar.

Theory dynamics also rescues DP, by ruling out cases like (C4). (C4) is impossible because it would be impossible to conceive an almost completely true hypothesis H which is contradicted by all the available evidence E: scientists work out their hypotheses starting from the available evidence. Besides, even if one were so crazy to imagine a hypothesis H which contradicted all the propositions in E, she would have no clue on how to construct H in such a way that all of its theoretical and unobserved empirical content were true. Therefore, it would be cosmically improbable that, among the countless hypotheses contradicting E, she picked just one that happens to be significantly verisimilar or partly true.

[4] This is quite possible if H was shaped *ad hoc* to accommodate E, just like in (C3b) H was shaped to accommodate p_1. More on this below.

[5] Mind, H has a true content, viz., E itself, but this is not in question in the realism-antirealism debate, which, as explained above, is concerned only with the theoretical and the unobserved empirical content of hypotheses (respectively, the propositions in (3) and (2) above).

Therefore, while Cevolani and Tambolo are right that, in order to vindicate DP and UP, we need to take into account "what E is", we need to consider not only the *quantity* of E (i.e., whether it is small, like in (C1), (C2) and (C3), or large, like in (C4) and (C5)), but also its *quality* (e.g., whether it was predicted or just accommodated), as well as the quality of H (e.g., whether it was just constructed *ad hoc* to accommodate E, or it made (also) some daring novel predictions). In other words, as the "predictivist" research tradition has shown (for an overview see Alai 2014a), what matters is more the quality of success than its quantity: even just one novel prediction can confirm more than many pure retrodictions.

6. Savojardo: "The Representation of Reality in the Intelligent Use of Tools"

Savojardo points out how a conflict between the manifest and the scientific image might emerge from the neurosciences. According to the Embodied Cognition account, cognitive activity does not depend only on the brain, but very importantly also on the action of the body on the mind. In particular, in order to avoid an opposition between motor and cognitive aspects, the abilities related to the use of tools are reduced to the sensory-motor level. This opposition won't go away, however: in fact, while the use of familiar tools requires just the retrieval of manipulative sensorimotor information or skills, when we create and use new tools, or use familiar tools in a new way, we need certain specific conceptual skills and certain purely cognitive inferential functions.

Thus, we are threatened by an irreconcilable separation between a prevailingly practical and sensorimotor knowledge, predominant in common everyday use of familiar tools, and a more abstract and theoretical knowledge, especially in science, where instruments themselves become objects of pure reasoning when they are devised, designed, produced and used in order to investigate the world: the manifest image would then become the reign of embodied and sensorimotor cognition, and scientific image the reign of abstract, theoretical knowledge.

According to Savojardo, however, this cleavage may be avoided by two arguments. The former, mainly relying on Buzzoni 2008, begins by maintaining that an intelligent use of tools is essential both in our everyday activities and in science. Whether we use a stick to move in the dark or a probe to explore space, we always do so "guided by an underlying intention to know the environment in order to intervene on it". In any case, "the use of an instrument […] mediates between our body and reality […] and this presupposes an important link between thought and action, and between cognitive and motor elements of knowledge".

Moreover, both in common knowledge and in science "the theoretical moment and the technical moment […] can be distinguished […] only on the level of reflection". Just like in everyday life "the mind often constructs possible alternative scenarios to real situations" to allow successful interaction with the world, at a more elaborate level scientists use counterfactual reasoning to explore aspects of reality more remote from everyday experience. Thus, "there is no human knowledge that is absolutely non-technical, just as there can be no knowledge that is merely practical-technical, unmediated by concept".

Savojardo's second argument hinges on Polanyi's (1958, 1969) distinctions between explicit and tacit knowledge on the one hand and subsidiary and focal awareness on the other. In this relation, commonsense knowledge might appear to be mainly tacit, and scientific knowledge exclusively explicit. Nonetheless, tools are essential to both, and in both they can either be used automatically, as sensorimotor prolongations of our body, or be explicitly conceived and planned as means to certain cognitive ends. In either way, however, tools are known, although tacitly in the former and explicitly in the latter.

For instance, if I use a hammer to drive a nail, I explicitly consider the hammer and the nail, i.e., I have focal awareness of them and of their operations; however, I couldn't achieve my goal unless, at the same time, I were perfectly aware, although in a merely *subsidiary* and tacit way, of the hammer's impulses on my palm and fingers (Polanyi 1958: 57). We cannot be focused *at the same time* on the instrument with its goals as a whole, and on its details: for instance, a pianist who shifts his attention to his fingers while playing risks to lose sight of the melody.

Nevertheless, whenever it is needed awareness can shift from subsidiary to focal, and *vice versa*, and there are intermediate degrees between them, and thus between purely implicit and fully explicit knowledge, and this clearly applies to the use of scientific instruments as well. For instance, we notice an analogous difference in approach when electrons are studied to investigate their properties, and when they are "sprayed", i.e., used as instruments, to reveal the existence of quarks with fractional charges (Hacking 1983: Ch. 16), or to study the trajectories of neutrons (Giere 1988: Ch. 5).

Therefore, while no doubt the use of tools in common knowledge is largely driven by implicit corporeal knowledge, whereas in science it is largely driven by explicit and highly sophisticated knowledge, this difference is gradual and reversible in perspective and approach, rather than radical.

7. Angelucci, Fano, Ferretti, Macrelli, Tarozzi: "Does Evolution Favor Accurate Perception?"

When one deals with the problem of reconciling the "two images", one usually takes the image s/he considers as more dubious or questionable and tries to understand whether its truth can be proven starting from the other image, which s/he assumes as true by default, or at least as standing on firmer grounds. For instance, in the debates on scientific realism, the manifest image is taken as basically true, and the question is whether the scientific image can stand up to the same standards. Other debates, however, proceed in the opposite direction: according to Sellars himself, it is the scientific image that must be taken as the benchmark for the manifest image,[6] and for philosophers like Paul Churchland (1981, 1988) and Steve Stich (1983) the progress of scientific psychology is showing that the manifest "folk psychology" is radically mistaken.

The paper "Does Evolution Favor Accurate Perception?" by Angelucci and colleagues is of the latter kind: there is a widespread tendency to draw on evolutionary biology to support the reliability of our sensory perception of physical reality, by claiming that in normal conditions our perceptual representations are largely accurate, since natural selection favors epistemically reliable perceptual systems.

This claim, however, has been rejected by Hoffman and colleagues (2013, 2015), who argued that the perceptual systems of animals are adapted to pursue utility (e.g., food, shelter, safety), rather than objective reality. To this end, they

[6] "In the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not" (Sellars 1963: 173).

imagine organisms (call them 'pragmatists') whose perceptual system can distinguish only what is useful to them from what is not, but ignore other objective differences in the environment; on the other hand, imagine organisms (call them 'realists') which can perceptually distinguish a wider range of properties and distinctions. For instance, certain blue things and certain green things may appear to pragmatists of one and the same color (say, grey), since they are all useful, while certain other blue things and green things may again appear to them as sharing one color but a different one (say, brown), because they are not useful (or even harmful) to them. In this way, however, pragmatists can immediately recognize what is useful and what is not, in spite of their wrong perception of colors. Thus, they are evolutionarily favored over realists, whose perceptual systems offer a more accurate picture of things, but which need time and effort to collect a more detailed chromatic information and to compute from it whether a given thing is useful or not. A model in evolutionary game-theory set up by Hoffman and collaborators showed then that pragmatists would flourish and realists would be driven to extinction.

If this were true, evolution would favor useful but false perception, and this would mean that our own perceptual representations of the world are largely wrong (this is also argued by Stich (1991: Ch. III)). This, by the way, might suggest that we should largely discard the manifest image and rely on theoretical science for a more precise picture of physical reality. Moreover, evolutionary epistemology could no longer support both commonsense realism and scientific realism by arguing that true perceptual beliefs are favored by evolution, and philosophical skepticism would gain momentum.

In their paper, however, Angelucci and colleagues argue that the above study failed to consider environmental modifications: when conditions change, differences which were previously irrelevant to utility may become relevant. For instance, it may become the case that all and only green things are useful. In this way, pragmatists would become utterly confused, still "believing" that certain blue things (appearing grey to them) are useful and that certain green things (appearing brown to them) are not. Thus, they would soon become extinct, while realists would readily adapt to the new conditions, because they can properly distinguish green from other colors. To press their point, they propose a different model, incorporating the effects of environmental change, showing that in this model organisms able to produce more realistic representations of the world are favored in the long run.

Of course, these kinds of models are necessarily quite idealized, and their scope is limited by the particular assumptions incorporated. They are rather like particular thought-experimental settings. For instance, much depends on whether environmental evolution is discontinuous, with prolonged periods of stability between one change and the next, or it is ubiquitous and continuous: in the latter case, it seems, realists would always be ahead of pragmatists.

Therefore, while a model like the present one cannot warrant too general and certain conclusions, at least it suggests that evolutionary game theory might not bring so grim news for the perceptual accuracy of the manifest image, after all. In fact, it might be observed that the various species of the genus *homo* were distinguished from other animals precisely by their flexibility and ability to exploit even minor changes. Even more importantly, they didn't wait for the environment to modify the utility functions, but they always actively changed them by "inventing" ever new ways to take advantage of the environment. From this point of view, it might be argued that perceptual realism has been one of the distinctive features of our species, and one of the keys to its evolutionary success.

References

Alai, M. 2014a, "Novel Predictions and the No Miracle Argument", *Erkenntnis*, 79, 2, 297-326.

Alai, M. 2014b, "Defending Deployment Realism against Alleged Counterexamples", in Bonino, G., Jesson, G., and J. Cumpa (eds.), *Defending Realism: Ontological and Epistemological Investigations*, Boston-Berlin-Munich: De Gruyter, 265-90.

Alai, M. 2019, "The Underdetermination of Theories and Scientific Realism", *Axiomathes-Epistemologia*, 29, 6, 621–37.

Bokulich, A. 2008, *Reexamining the Quantum–Classical Relation – Beyond Reductionism and Pluralism*, Cambridge: Cambridge University Press.

Buzzoni, M. 2008, *Thought Experiment in the Natural Sciences: A Transcendental-Operational Conception*, Würzburg: Königshausen & Neumann.

Carnap, R. 1928, *Der Logische Aufbau der Welt*, Leipzig: Meiner Verlag.

Cevolani, G., Crupi, V., and Festa, R. 2011, "Verisimilitude and Belief Change for Conjunctive Theories", *Erkenntnis*, 75, 183-202.

Cevolani, G., Festa, R., and Kuipers, T.A.F. 2013, "Verisimilitude and Belief Change for Nomic Conjunctive Theories", *Synthese*, 190, 3307-24.

Churchland, P.M. 1981, "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy*, 78, 2, 67-90.

Churchland, P.M. 1988, "Folk Psychology and the Explanation of Behaviour", *Proceedings of the Aristotelian Society*, Supp. Vol. 62, 209-21.

Eddington, A.S. 1928, *The Nature of the Physical World*, New York: The Macmillan Company.

Giere, R.N. 1988, *Explaining Science: A Cognitive Approach*, Chicago: Chicago University Press.

Hacking, I. 1983, *Representing and Intervening*, Cambridge: Cambridge University Press.

Hempel, C.G. and Oppenheim, P. 1948, "Studies in the Logic of Explanation", *Philosophy of Science*, 15, 2, 135-75; reprinted in Hempel 1965.

Hempel, C.G. 1965, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.

Hoffman, D.D., Singh, M., and Mark, J.T. 2013, "Does Evolution Favor True Perceptions?", in Rogowitz, B.E., Pappas, T.N., and de Ridder, H. (eds.), *Proceedings of the SPIE 8651, Human Vision and Electronic Imaging*, XVIII, 865104.

Hoffman, D.D., Manish, S., and Prakash, S. 2015, "The Interface Theory of Perception", *Psychonomic Bulletin & Review* 22, 6, 1480-1506.

Husserl, E. 1936, *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie: Eine Einleitung in die phänomenologische Philosophie*, The Hague: Martinus Nijhoff, 1954. Engl. Transl.: *The Crisis of European Sciences and Transcendental Phenomenology*, Evanston: Northwestern University Press, 1970.

Knuuttila, T. 2021, "Epistemic artifacts and the Modal Dimension of Modelling", *European Journal for Philosophy of Science*, 11, 65.

Kuipers, T.A.F. 2000., *From Instrumentalism to Constructive Realism*, Dordrecht: Springer.

Laudan, L. 1981, "A Confutation of Convergent Realism", *Philosophy of Science*, 48, 19-49.

Main, J., Weibusch, G., Holle, A., and Welge, K.H. 1986, "New Quasi-Landau Structure of Highly Excited Atoms: The Hydrogen Atom", *Physical Review Letters*, 57, 2789-92.

Musgrave, A. 2006-2007, "The 'Miracle Argument' for Scientific Realism", *The Rutherford Journal*, *The New Zealand Journal for the History and Philosophy of Science and Technology*, 2, http://www.rutherfordjournal.org/article020108.html.

Poincaré, H., 1905, *La Science et l'Hypothèse*, Paris: Flammarion.

Polanyi, M. 1958, *Personal Knowledge: Towards a Post-Critical Philosophy*, London: Routledge and Kegan Paul; quotes from II ed. 1962.

Polanyi, M. 1969, *Knowing and Being*, London: Routledge & Kegan Paul.

Psillos, S. 1999, *Scientific Realism: How Science Tracks Truth*, New York: Routledge.

Schlick, M. 1938, *Form and Content: An Introduction to Philosophical Thinking*, in Schlick, M. *Gesammelte Aufsätze*, 1926-1936, Wien: Gerold, 151-249.

Sellars, W. 1962, "Philosophy and the Scientific Image of Man", in Colodny, R.G. (ed.), *Frontiers of Science and Philosophy*, Pittsburgh: University of Pittsburgh Press, 35–78.

Sellars, W. 1963, *Science, Perception and Reality*, London: Routledge & Kegan Paul and New York: The Humanities Press; reissued in 1991 by Ridgeview: Atascadero.

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

Stich, S. 1983, *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press.

Stich, S. 1991, *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*, Cambridge, MA: MIT Press.

Van Fraassen, B. 1980, *The Scientific Image*, Oxford: Oxford University Press.

Wittgenstein, L. 1921, "Logisch-Philosophische Abhandlung", in *Annalen der Naturphilosophie*, 14.

Zamora Bonilla, J. 1992, "Truthlikeness without Truth: A Methodological Approach", *Synthese*, 93, 343-72.

Zamora Bonilla, J. 1996, "Verisimilitude, Structuralism, and Scientific Progress", *Erkenntnis*, 44, 25-47.