

Explanation by Idealized Theories

Ilkka Niiniluoto
University of Helsinki
ilkka.niiniluoto@helsinki.fi

Abstract The use of idealized scientific theories in explanations of empirical facts and regularities is problematic in two ways: they don't satisfy the condition that the explanans is true, and they may fail to entail the explanandum. An attempt to deal with the latter problem was proposed by Hempel and Popper with their notion of approximate explanation. A more systematic perspective on idealized explanations was developed with the method of idealization and concretization by the Poznan school (Nowak, Krajewski) in the 1970s. If idealizational laws are treated as counterfactual conditionals, they can be true or truthlike, and the concretizations of such laws may increase their degree of truthlikeness. By replacing Hempel's truth requirement with the condition that an explanatory theory is truthlike one can distinguish several important types of approximate, corrective, and contrastive explanations by idealized theories. The conclusions have important consequences for the debates about scientific realism and anti-realism.

Keywords approximation, concretization, explanation, idealization, scientific realism, truthlikeness.

DOI 10.2478/kjps-2018-0003

1. Aspects of Scientific Explanation

According to Carl G. Hempel's (1965) classical analysis of nomological or covering law explanation, a D-N-explanation is a relation between the explanandum E (i.e. a sentence describing a known empirical event, fact, or uniformity) and the explanans (i.e. empirically testable universal laws L and initial conditions C). The "logical condition" requires that

(R1) The explanandum E must be deducible from the explanans L&C.

The argument from L&C to E answers the explanation-seeking question “Why E?” by showing that E was to be expected.¹ To the requirement (R1) Hempel added an “empirical condition of adequacy”:

(R2) The sentences constituting the explanans must be true.

An argument which satisfies the inference requirement (R1) but not necessarily the truth requirement (R2) is only a *potential* explanation. Hempel argued that it is not appropriate to stipulate that the explanans is well-confirmed by available evidence, since this would have the awkward consequence that an argument could lose its status as an explanation when new disconfirming empirical evidence is found (*ibid.*, 248). The point of (R2) is to guarantee that being an explanation is a permanent property of an argument. It also shows that the Hempelian subsumption theory of explanation is not only “epistemic” but has an important “ontic” aspect as well (cf. Salmon, 1984, 16–17). The inference (R1) and truth (R2) conditions of Hempel’s model are highlighted in Fig. 1.

Conditions, which are similar to (R1) and (R2), can be given also for models of explanation which complement or reformulate Hempel’s covering law account. For example, if the laws L are causal, then the conditions C should specify real causes of the effect E (cf. R2). In theoretical explanations, the laws L contain theoretical terms which refer to real entities and mechanisms beyond the edge of observability, and these mechanisms should be able to generate the explanandum event E (cf. R1). (See e.g. Craver, 2006.) For this reason, in this paper I will analyze explanations by referring to Hempel’s model, but the arguments with appropriate modifications can be applied to its main alternatives as well.

While scientific realism was not a well-defined position in the 1940s, Hempel was bold enough to require in 1948 that scientific laws and

¹ Hempel (1965), 385–412, treats also inductive-probabilistic explanations, where the premises include probabilistic laws and the explanandum is inducible from the explanans. He mentions also how possible –explanations (*ibid.*, 428). Probabilistic models may be idealized and approximate, and idealized simulations may have an interesting role in showing how observable patterns can be generated. But I restrict my attention to non-probabilistic theoretical why-explanations in this paper.

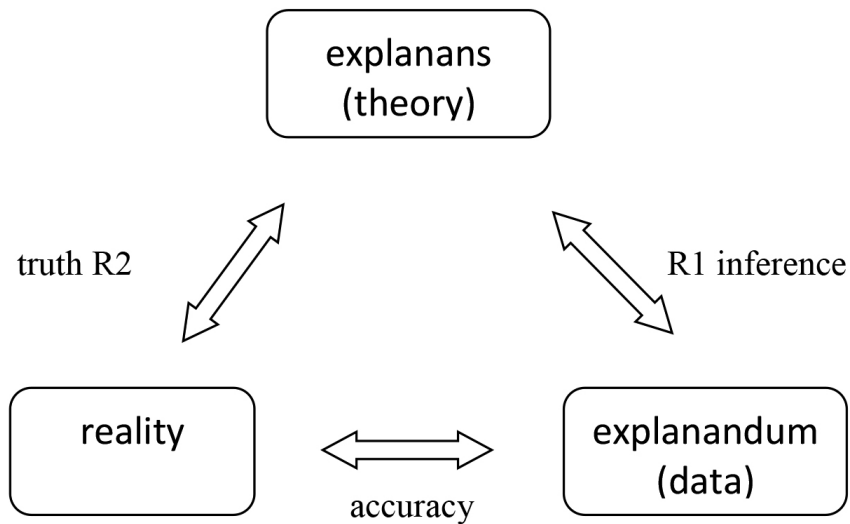


Fig. 1. Hempel's conditions for explanation.

theories are true (see Hempel, 1965, 272–273). Indeed, he had given up the coherence theory of truth after the publication of Alfred Tarski's semantic theory of truth in 1936, so that for him truth is a relation between theory and reality, while the data *E* obtained by observation and measurement give a more or less accurate representation of reality (see Fig. 1 again). This idea of *correct* or *true* explanations contradicts the ancient instrumentalist tradition which demanded only that hypotheses are sufficient to “save the phenomena”. In Pierre Duhem's 1907 version of instrumentalism, scientific theories lack truth-values and their task is not to give explanations at all (see Duhem, 1954). While some other anti-realists have analyzed explanations without requiring that answers to explanatory why-questions are correct (see e.g. van Fraassen, 1980), this kind of treatment concerns only potential explanations.

The notion of potential explanation is needed if one contends that successful explanations and predictions give empirical support or confirmation to a hypothetical theory:

- (R3) If theory T potentially explains empirical phenomena E, then E gives reason to think that T is true.

The conclusion of (R3) can be understood as stating that T is highly probable given E. But if, instead, confirmation is explicated by the relation of positive relevance, i.e. $P(T/E) > P(T)$, so that evidence E increases the probability that theory T is true, then (R3) is an instance of what is usually called “the converse entailment” condition of confirmation (see Hempel, 1966, 8). In any case, for a scientific realist, principle (R3) gives empirical reasons for thinking that theory T is true. It thus expresses the basic idea of arguments that Charles S. Peirce called *abductions* (see Peirce, *CP* 5.189).

The strong conditions (R1), (R2), and (R3), typically endorsed by “hard” realists, are problematic in the case of idealized scientific theories which are known to include counterfactual assumptions and approximations. Such theories are known to be false, and they may fail to entail the explanandum. In this paper, we shall see how these conditions have been liberalized in later discussions about scientific explanation.

2. Approximate and Corrective Explanations.

In 1907 Pierre Duhem pointed out that Kepler’s observational laws formally contradict the laws of Newton’s mechanics (see Duhem, 1954, 193). In a lecture in 1948, Karl Popper admitted that “Kepler’s laws cannot be deduced from Newton’s”, but can be obtained only *approximately* by making “the assumption that masses of the various planets are negligible compared with the mass of the sun” (see Popper, 1972, 357). In a paper in 1957, Popper stated that Newton’s theory unifies Galileo’s and Kepler’s laws, but “it corrects them while explaining them” (*ibid.*, 198–202).

According to Kepler’s Third Law, $t^2/a^3 = \text{constant}$ for all planets, where t is the period of revolution around the sun and a is the semi-major axis of the ellipse. Newton’s theory entails that $t^2/a^3 = 4\pi^2/G(m+M)$, where m is the mass of the planet and M is the mass of the sun. Thus, Kepler’s Harmonic law is strictly speaking false, since the planets have different masses, but still it is approximately correct for planets with approximately equal masses or negligible masses ($m \approx 0$). With reference to

Duhem and Popper, Hempel concluded that Newton's theory might be said to provide *approximative D-N explanations* of Kepler's and Galileo's laws (see Hempel, 1965, 344). Such explanations are corrective, as the theory shows that the empirical generalizations "hold true in fairly good approximation" (see Hempel, 1966, 76).

More generally, let us say that two statements *approximate* each other if their distance or degree of similarity is sufficiently small, where the choice of the threshold of approximation may depend on the context and the judgment of the scientific community. Some distance functions measure analogy or partial identity by counting the matches and mismatches between two statements or structures,² while more generally some geometrical metrics allow that similar statements differ from each other at all points of comparison.³ In particular, an empirical law is *approximately true* if it approximates the true law. A theory is *truthlike* if it approximates the complete truth about some domain or fragment of reality.

Now we can say that theory T *approximately explains* E if one can derive from T a statement E' which approximates E. The attempted explanation fails if the distance between E and E' is too great. This notion is thus based on a liberalized version of Hempel's condition (R1):

(R1*) The explanans is approximately derivable from the explanandum.

But one can still try to maintain the truth requirement (R2) by claiming that theories correct observational regularities: an approximate explanation of E by T is *corrective* if the derived statement E' is more accurate than E (i.e. E' has a higher degree of approximate truth than E). This is typically the view of scientific realists, like Wilfrid Sellars (1963), who think that successful true theories are able to penetrate deeper into reality behind observable phenomena. The opposite view is defended

² See e.g. the applications of Amos Tversky's contrast measures by Tuomela (1979), Cevolani, Crupi, and Festa (2011) and Weisberg (2013).

³ See Niiniluoto (1987, 2013). Note that Weisberg (2013) defends Schelling's checkerboard model of segregation as an explanation by claiming that this model is similar to real cities by the contrast measures, while Niiniluoto (2013) argues that Schelling's ideal toy cities share no attributes with real cities in the sense of positive analogy.

by Nancy Cartwright (1983), who claims that the phenomenological laws of applied physics are highly accurate or true, while the theoretical laws “lie” so that the strict predictions from fundamental theories are less accurate or further from the truth. If a theory gives a correct description of a model, then for Cartwright such a model is only an idealized simulacrum of a real-world system.

For the fallibilists, who think with Peirce that complete certainty cannot be achieved in science, Hempel’s requirement (R2) seems too strong. As Popper (1972) argued, even our best theories are typically false, but still they can be improved by replacing them with more truthlike successors (see Niiniluoto, 1999). Popper’s unsuccessful attempt to explicate this idea can be saved by defining the *degree of truthlikeness* or *verisimilitude* of a theory T in language L in terms of the distances of the complete states of affairs (“possible worlds”) allowed by T from the target t^* which is the true state of affairs as described in L (i.e. the most informative generalization in L).⁴ Maximal truthlikeness is achieved by a theory which is equivalent to the target t^* . For true theories, truthlikeness covaries with logical strength. It is important that also false theories can be compared by this criterion: the truthlikeness of a false statement allows that its (smaller) mistakes are compensated by its (more) truthful components. Truthlikeness is thus a holistic notion which applies to a theory as whole.

From the fallibilist perspective, Hempel’s (R2) sets the bar for actual explanations too high, and there is a need to relax the requirement that the premises of explanations are true. We shall explore below the idea that explanations are required to be truthlike. Thus, Hempel’s requirement (R2) for correct explanations can be replaced with the condition that explanations are truthlike:

(R2*) The explanans must be truthlike.

An important additional motivation for this move comes from the question whether idealized theories can be used as explanations. Idealized

⁴ If theory T is formulated as a statement with modal operators (like physical or nomic necessity), the target t^* has to be the strongest nomic truth about the world. For details about truthlikeness measures, see Niiniluoto (1987), Kuipers (2000), Cevolani, Crupi, and Festa (2011).

scientific theories tell how natural and social systems behave under counterfactual conditions, so that they are not true – indeed, they are often known to be false, as they ignore or exaggerate some relevant features. The use of such theories as covering laws in correct explanations is therefore at least problematic. Moreover, due to the counterfactual assumptions, idealizations are always to some extent removed from empirical reality, so that their ability to serve as premises of explanations is questionable even by the weakened standard (R1*).

With the changed principles (R1*) and (R2*), the condition (R3) has to be reformulated as well. Instead of concluding abductively that successful explanations give reasons for the truth of an explanation, abduction should be understood as inference to the truthlikeness of a hypothetical explanation:

(R3*) If theory T potentially explains empirical phenomena E, then E gives reason to think that T is truthlike.

In particular, if the success of an explanatory theory is only approximate (cf. R1*), then (R3*) is more adequate than Peirce's original schema of abduction (see Kuipers, 2000; Niiniluoto, 2005).

3. Idealization and Concretization.

Leszek Nowak (1971) argued that Hempel's D-N model ignores the idealizational character of scientific laws and theories. Similarly, Nowak (1972) criticized Ernest Nagel's treatment of scientific laws. Nowak (1980) claimed that Karl Marx followed in economics the same method of "idealization and concretization" as Galileo in physics. It is remarkable that a young Polish Marxist had the courage to question the Received View of analytic philosophy of science. The success of his approach is reflected in the twelve volumes on idealization in the monograph series *Poznan Studies in the Philosophy of the Sciences and the Humanities* in 1990–2005.⁵

5 For the Poznan school, see also Krajewski (1977); Brzezinski (2007).

Nowak's own approach is based on the systematic development of the method of *idealization and concretization*: hidden ideal assumptions are first made explicit as antecedents of idealizational laws and then these assumptions are eliminated or relaxed by modifying the consequent by a process of concretization. This idea can be illustrated by the Boyle-Marriott law

$$(1) \quad pV = RT$$

which is at best a truthlike description of the relations of pressure p , volume V , and absolute temperature T of real gases (R is a constant). Therefore, it is called the "ideal gas law" which correctly describes only the ideal gas model. This law ignores the influence of intermolecular attractive forces (a) and the finite size of gas molecules (b). When these hidden idealizing assumptions are made explicit, the factual but false law (1) is transformed to an *idealizational law*

$$(2) \quad \text{If } a = 0 \text{ and } b = 0, \text{ then } pV = RT.$$

This law is concretized by removing the assumptions step by step:

$$(3) \quad \text{If } b = 0, \text{ then } pV = RT - a/V$$

$$(4) \quad (p + a/V^2)(V - b) = RT.$$

Here the new factual claim (4) is known as van der Waals' law, which still may be an idealization with respect to some further hidden factors.⁶ An alternative to the concretization steps (3) and (4) is to introduce an "approximation statement" with respect to (1) of the form

$$(5) \quad \text{If } a \leq \alpha \text{ and } b \leq \beta, \text{ then } pV \approx RT,$$

where α and β are small constants. Statements of the form (5) may be testable by systematic experiments.

⁶ Note that (1) and (4) are quantitative laws of coexistence which not merely express a Humean regularity about the gases which happen to exist in the actual world, but assert what combinations of quantitative variables are physically possible. They can be expressed modally as physically necessary statements, so that their truth conditions can be given in the possible world semantics. See Niiniluoto (1987), 111–112.

Nowak himself treated (2) and (3) as material conditionals with false antecedents, which are then trivially true. The problem is that this would make these statements true whatever their consequent is, so that Nowak was compelled to postulate ad hoc a new notion of truth. But if (2) and (3) are read as counterfactual conditionals, their truth (or truthlikeness) means that the consequent is true (or truthlike) in those possible worlds which satisfy the ideal condition and are otherwise like the actual world (see Niiniluoto, 1990). In such possible worlds some real factors are eliminated or neutralized by assigning them extreme values. Now we have the important result that the concretization (4) logically entails (3) and (2), so that the so called Correspondence Principle is satisfied (see Krajewski, 1977) and concretization is a method of “potential truth approximation”.⁷ Note also that if $a \rightarrow 0$ and $b \rightarrow 0$, then the law (4) approaches (1) in the limit. If (4) is true, then (2) and (3) are true as well, but as weaker truths they are less truthlike than (4). The initial ideal gas law (1) is truthlike if it is sufficiently close to the true concretization (4): the law (1) is strictly speaking false everywhere, but at least for sufficiently large values of V it is approximately equal to the law (4). If (4) is false but truthlike, its consequences (2) and (3) are also truthlike but to a degree less than (4).

Nowak’s account covers Galileo’s study of free fall with the idealization that the moving body is not affected by the resistance of air. This law can be concretized, “factualized” (cf. Krajewski, 1977) or “de-idealized” (cf. McMullin, 1985) by adding the resistance of air as a new linear force and calculating the modification by Newton’s theory. Such *Galilean idealizations* are methodological in the sense that it is easier to study first a simpler framework which can later be enriched by new factors. For Weisberg (2007), Galilean idealizations are temporary simplifications for the purpose of computability.

Nowak’s (1980) treatment captures also another idea: to study free fall without other disturbances allows us to see this phenomenon in its “pure” or “essential” form. He assumes that for a given problem one can hierarchically list the primary and secondary factors which influence the variable under investigation, and idealization starts from the introduc-

⁷ Unsuccessful concretization can be made in a mistaken way, so that it leads us further from the truth. See Kuipers (2000).

tion of the essential or privileged ones and then adds step by step the initially eliminated secondary ones. This has been called the method of isolation (see Mäki, 1994, 2009) or minimalist idealization (see Weisberg, 2007). If one chooses the target t^* as the complete truth in a language which includes only the essential core factors, then the distance from t^* expresses the “essesimilitude” of a theory (see Mäki, 1994; Niiniluoto, 2002).

Let us now return to Nowak’s (1971) critique of Hempel. He dismissed Hempel’s proposal that “theoretical idealizations” are consequences of more general factual theories, so that such theories can be used as premises of explanations (see Hempel, 1965, 164; Rudner, 1966, 62). For example, projectiles could be explained directly by Newton’s theory instead of idealized ballistic equations. For Nowak, there is no guarantee that such stronger factual background theories are always available – and Newton’s theory itself turns out to be an idealization with the hidden assumption that $1/c = 0$, where c is the velocity of light. While Hempel and Rudner were aware that lawlike theories sustain counterfactual conditionals,⁸ Nowak (1980), 66, interpreted Hempel’s “theoretical idealizations” as the consequents (cf. (1)) of genuine idealizational laws (cf. (2)). So Nowak’s achievement was to formulate statements of the form (2) and their concretizations, but he failed to interpret them as counterfactual conditionals.

It may happen that from a theory with hidden idealizations one can derive consequences which are close to observed regularities. For example, Newton’s mechanics gives an approximately correct description of middle-sized bodies moving with finite velocities. Sometimes such consequences give corrective approximate explanations of empirical laws. For example, the correction of Kepler’s Harmonic Law by Newton’s theory is a concretization in Nowak’s sense, as the relevant factor m , which is missing in Kepler’s formula, is introduced to the derived law. In this case, the conditions $(R1^*)$ and $(R2^*)$ are satisfied.

On the other hand, the consequences E' of an idealized theory T may be too distant from the explanans E , so that T fails even to approximately

⁸ This condition of lawlikeness was suggested by Nelson Goodman. See Hempel (1965), 339.

explain E. Cases of this kind are illustrated by Kristin Shrader-Frechette (1989) by means of Darcy's Law in hydrogeology. In this sense, there is no general answer to the question whether idealized theories can serve as potential explanations: whether a theory satisfies the condition (R1*) depends on the case and the chosen standard of approximation.

Nowak's proposal is to narrow the gap between an idealized theory T and empirical reality E by relaxing the hidden idealizing assumptions of T: to explain E by T one has to concretize T into a new theory T' and then derive E from T'. If \dashv is used as the sign for concretization, Nowak replaces the initial theory T with a sequence of concretizations $T, T_0 \dashv T_1 \dashv \dots \dashv T_k$ so that the Correspondence Principle is satisfied, and then explains E by combining initial conditions C with the least idealized theory T_k among the known concretization of T, or an approximation statement with respect to T_k . If the process is successful, we reach at least an approximate explanation which satisfies the condition (R1*).

As successful concretization increases the truthlikeness of an idealized theory, Nowak's insight can be captured by the modified condition (R2*) as well. But again there is no general answer to the question whether idealized theories satisfy (R2*). We have seen that idealized theories may be truthlike, so that they may as such give acceptable approximate explanations. But if they do not satisfy (R2*), we should continue their concretization up to the point that the modified theory is truthlike.

An interesting modification of these ideas is given by Frank Hindriks (2008), who suggests that false theories and models are "explanatory engines" with a lot explanatory power, as they contain many idealizing assumptions to be relaxed. Suppose that an idealized theory T implies an empirical consequence E, and the concretization T' of T implies another consequence E'. Then we may explain why the consequence E fails to obtain by answering with T' the *contrastive* explanatory question "Why E' rather than E?". Here it might be the case that E' is a successful concretization of E, so that this explanation is corrective. For example, E' could be the corrected version of Kepler's law derivable from Newton's theory, or the observations of projectiles in real conditions with resistance of air, or T' might be Einstein's special theory of relativity which corrects the descriptions of rapidly moving bodies given by Newton's mechanics T. However, Hindriks (2013) suggests that in "explanation by relaxation"

neither of the theories T and T' has to be “realistic in the sense that it approximates the way things are”, and neither of the consequences E and E' has to be approximately true. Thus, his account does not require condition (R2*).

4. Recent Debates.

The earlier sections may be summarized by the hypothesis that an idealized theory T explains data E if and only if the following conditions are satisfied:

(R1*) E is approximately derivable from T

(R2*) T is truthlike.

Both of these success conditions use notions like “approximately derivable” and “truthlike”, which are relative to some threshold, but this is similar to widely used methodological notions like “probable”, “confirmed” or “well supported”. These liberalized conditions include as special cases Hempel’s original conditions (R1) and (R2).⁹ Further, success in explanation is an indicator of truthlikeness:

(R3*) If T would explain E , then E supports the truthlikeness of T .

This requirement is a modification of Peirce’s model of abduction.

In a recent survey, Alisa Bokulich (2017) argues that “the use of idealized models to explain phenomena is pervasive across the sciences”, but admits that this theme “remains an ongoing area of research”. In the light of my presentation, one of the recurring problems in recent contributions is that many philosophers do not make a clear distinction between the inference and truth requirements of Fig. 1. Another problem is ignorance of the notions of approximate truth and truthlikeness, and thus the failure to employ them in formulating criteria for the use of idealizations in explanatory arguments. A further problem is the con-

⁹ Note, however, that (R2*) is stronger than the original (R2) in one respect: weak uninformative truths like tautologies fail to be truthlike. By (R2*) explanatory theories have to be both informative and close to the truth.

fusion between genuine explanations and merely analogical arguments. These critical points can be illustrated by showing how conditions (R1*), (R2*) and (R3*) provide us a tool of analyzing some recent debates about explanation by idealized theories.

Some philosophers, who work with causal and mechanistic models of explanation, have maintained “hard” conditions which correspond to Hempel’s original (R1) and (R2). For example, Carl Craver (2006) requires that mechanistic models, which are adequate to give how-actually explanations, should give a correct and (at least contextually) complete description of components which correspond to the real components that in fact produce the phenomenon in the target system. This excludes the possibility, allowed by (R2*), that idealized models are explanatory. Similarly, Michael Strevens (2008), 297, who has developed his own “kairetic” account of the relevant “difference-makers” in causal explanations, states that “no causal account of explanation allows nonveridical models to explain” (*ibid.*, 297).

Strevens continues with the puzzling thesis that “an idealized explanation is always better than its veridical counterpart”. But, on closer look, by “idealized explanation” Strevens means here the derivation of the approximate truth of Boyle’s law (1) from idealized assumptions (such as $a = 0$ and $b = 0$ in (4)), so that the explanandum is not real gas but gas with “Boylean behavior” (*ibid.*, 311). This is similar to Sorin Bangu’s (2015) view that fictional idealizations can be “honestly” used to explain only fictional explananda. These arguments fail to take notice of the notion of idealized approximate explanation of real phenomena allowed by (R1*).

Alisa Bokulich (2012) rejects Hempel’s truth requirement (R2), but is worried that one should not too easily grant explanatory power to idealized fictions. For example, Ptolemy’s epicycles are nonexplanatory fictions. She attempts to specify conditions for “model explanations” by demanding that to some extent the model correctly captures or adequately represents the target system. In another paper, Bokulich (2011) identifies “structural model explanations”, where the counterfactual dependence required by causal relations is grounded by the mathematical structure of the theory. These formulations are not quite precise, but they may give warrant to the condition (R2*) that the idealized model is truthlike.

Yasha Rohwer and Collin Rice (2013) consider “hypothetical pattern idealizations” which are “explanatory” without being “explanations”. In their special terminology, explanatory representations justify true beliefs relevant to answering why-questions, but are not sufficiently veridical about real-world systems to provide explanations. It seems that this proposed distinction can be explicated by our conditions: “explanatory” idealizations satisfy (R1*), while “explanations” should satisfy (R2*).

Richard Batterman and Collin Rice (2014) argue that “minimal model explanations” are able to explain the macrobehavior of the target system if they belong to the same universality class, even when their microdetails do not share any common features. Due to the multiple realizability of the macro-level, causal details of the micro-level become irrelevant: having the same macrofeatures of a universality class allows explanation even when the model does not represent any relevant microscale features of the system. This amounts to the principle (R1) without (R2*), so that their minimal models only save the phenomena without any requirement of their truthlikeness. Lange (2015) further objects that Batterman and Rice in fact rely on the “common features approach” which they try to reject.

Many philosophers have drawn attention to the peculiar nature of nonlinear systems. Alexander Rueger and David Sharp (1996) optimistically conclude against Cartwright (1983) that theories in a “messy world” of nonlinear dynamics can be simple, explanatory, and confirmable at the same time (cf. R3*). But it is often suggested that such nonlinear theories are *non-Galilean idealizations*, so that their idealizing assumptions are not harmless in the sense that they could be de-idealized (cf. McMullin, 1985). Chang Liu (1999) argues that causal idealizations cannot be improved by concretization. Batterman (2002) analyzes critical phenomena (such as boiling water or drop formation in liquids) which involve physical discontinuities and singular limits, where the idealized thermodynamical limit (i.e. the assumption that the number of particles is infinite) cannot be de-idealized. The reason is that without the infinity assumption no mathematical solutions of dynamic equations are possible, so that the derivability condition (R1*) would fail. Still, Andrew Wayne (2010) argues that not only “Galilean, harmless, controllable” idealizations can explain, but we should allow that also non-Galilean idealiza-

tions, which are not true or approximately true of the physical system, have explanatory capacities. However, he adds, this would require that we give up the assumption that explanation presupposes successful representation even in the liberal sense of (R2*). Similarly, Eric Winsberg (2010) notes that non-realistic models with “fictions” (e.g. artificial viscosity with an infinite value), which are not even approximately true, may be empirically successful and reliable, and concludes that the rule “success implies truth” does not hold (cf. R3).

These “non-Galilean” arguments deserve serious consideration. They question Nowak’s treatment of idealization and concretization (usually without mentioning the extensive work of the Poznan school), and challenge the modified principles (R2*) and (R3*). They are also highly relevant to the debates between scientific realists and anti-realists. But they are not quite conclusive, as their use of the key notions of “approximate truth” or “successful representation” is largely intuitive.

First, it is clear that idealized assumptions taken as separate single claims (e.g. infinite viscosity, infinite velocity of light) may fail to be approximately true. However, together with other claims, an idealized theory or model as a whole may be truthlike or sufficiently similar to the real system (see Weisberg, 2013; Niiniluoto, 2013).

Secondly, one may ask in what sense non-Galilean models cannot be de-idealized. Wayne’s (2010) discussion of the quadratic damping of air resistance is illuminating. The problem is not in adding quadratic air resistance to pendulum models, but rather in the impossibility of solving the new oscillator equation exactly or with approximation techniques involving regular limits. But this failure of (R1*) is compatible with the possibility that the addition of quadratic damping has increased the truthlikeness of the pendulum model (cf. R2*).

Thirdly, as we already noted, in some cases zero or infinity assumptions may be introduced in order to make the equations solvable. Then we deliberately sacrifice something from the truthlikeness of the theory in order to make it manageable, since otherwise no explanations or predictions could be derived from the theory. Here Nowak’s proposal to concretize the theory before giving an explanation cannot be accomplished in practice, but nevertheless such a theory which cannot be de-ideal-

ized any more may be a sufficiently truthlike representation, so that its explanations satisfy both (R1*) and (R2*). Thus, it is not decisive for the possibility of explanation by T whether T is a Galilean or non-Galilean idealization, but rather whether it satisfies (R2*).

Fourthly, Roman Frigg et al. (2013) have argued that weather and climate models as nonlinear dynamical systems are unpredictable, if there is only the slightest model imperfection in the basic equation. Even though this may seem to imply that (R2*) excludes (R1*), it is not clear whether this feature of the models has any consequences for their ability to give explanations. Short-run predictions (with few iterations) are not excluded in these systems, so that the same holds for explanations. Further, what Hempel called self-evidencing explanations seem to be possible within such idealized models. For example, in a single toss of a coin heads and tails are equally probable and thus unpredictable outcomes,¹⁰ but when heads has occurred we can infer backward to the kind of initial condition which has produced this result. More generally, if we know what state the system has reached, we may be able to reason inversely to its initial state, and then use it as the initial condition of an explanatory argument (cf. Hempel, 1965, 274–276).

Finally, it is important to observe that some successful inferences from idealized models and simulations to target systems are not explanatory, since they apply analogical reasoning (see Niiniluoto, 2013). For example, animals are sufficiently analogous to humans, so that they are used as evidence in medical experiments. But in such cases the rule (R3*) cannot be applied: successful drug testing by animal experimentation does not prove that human beings are mice or zebra-fish.

A fallibilist realist would not accept the simplified thesis that “success implies truth”, as the correct principle (R3*) is much more cautious. The successful engineering applications of mechanical theories (e.g. Newton’s mechanics) do not prove that all of their idealizations are strictly true. Empirically or computationally successful applications do not prove that *literally* the atomic nucleus could be a liquid drop or that electrons could have classical trajectories in quantum dots (cf. Bokulich, 2012).

¹⁰ For frequencies in repeated tosses of coins, approximate probabilistic predictions and explanations are possible.

Here we have two options, depending on the case. One is to conclude that, in spite of their success in the sense of (R1*), which is based on analogical rather than explanatory reasoning, such idealizations lack ontological depth which is required of explanations by (R2*). The other is to accept them as parts of explanatory theories which by (R3*) maintain a fallible Peircean abductive link between explanation and epistemic warrant: if theory T is a successful potential explanation of sufficiently broad empirical data or regularity E, then E supports the claim that T as whole is truthlike.

In conclusion, while some hard realists and anti-realists may agree with Duhem that idealized theories and models have no explanatory capacities, there is also a new tendency of thinking that such theories save the appearances or give explanations without being successful representations of reality. By assessing recent debates with the liberal conditions (R1*), (R2*) and (R3*), we have argued against such an instrumentalist trend and supported the realist view that the explanatory success of idealized scientific theories goes together with their closeness to the truth.¹¹

References

- Bangu, Sorin (2015), “Why Does Water Boil? Fictions in Scientific Explanation”, in Uskali Mäki et al. (eds.), *Recent Developments in Philosophy of Science*, Springer, Cham, 319–330.
- Batterman, Robert W. (2002), *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, Oxford University Press, Oxford.
- Batterman, Robert W. and Rice, Collin C. (2014), “Minimal Model Explanations”, *Philosophy of Science* 81, 349–376.
- Bokulich, Alisa (2011), “How Scientific Models Can Explain”, *Synthese* 180, 33–45.

¹¹ The first version of this paper was presented in the Fifteenth International Congress of Logic, Methodology and Philosophy of Science in Helsinki, August 6, 2015. I am grateful to the organizers and participants of the 3rd Lisbon ICPOS, especially António Zilhão, for stimulating discussions in December 2016.

Bokulich, Alisa (2012), "Distinguishing Explanatory from Nonexplanatory Fictions", *Philosophy of Science* 79, 725–737.

Bokulich, Alisa (2017), "Models and Explanation", in Lorenzo Magnani and Tommaso Bertolotti (eds.), *Springer Handbook of Model-Based Science*, Springer, Dordrecht, 103–118.

Brzezinski, Jerzy et al. (eds) (2007), *The Courage of Doing Philosophy: Essays Presented to Leszek Nowak*, Rodopi, Amsterdam.

Cartwright, Nancy (1983), *How the Laws of Physics Lie*, Clarendon Press, Oxford.

Cevolani, Gustavo, Crupi, Vincenzo & Festa, Roberto (2011), "Verisimilitude and Belief Change for Conjunctive Theories", *Erkenntnis* 75, 183–202.

Craver, Carl E. (2006). "When Mechanistic Models Explain". *Synthese* 153, 355–376.

Duhem, Pierre (1954), *The Aim and Structure of Physical Theory*, Princeton University Press, Princeton.

Frigg, Roman, Bradley, Seamus, Machete, Reason & Smith, Leonard (2013), "Probabilistic Forecasting: Why Model Imperfection Is a Poison Pill", in Hanne Andersen et al. (eds.), *New Challenges to Philosophy of Science*, Springer, Dordrecht, 479–491.

Hempel, Carl G. (1965), *Aspects of Scientific Explanation*, The Free Press, New York.

Hempel, Carl G. (1966), *Philosophy of Natural Science*, Prentice-Hall, Englewood Cliffs.

Hindriks, Frank (2008), "False Models as Explanatory Engines", *Philosophy of the Social Sciences* 38, 334–360.

Hindriks, Frank (2013), "Explanation, Understanding, and Unrealistic Models", *Studies in History and Philosophy of Science* 44, 523–531.

Krajewski, Wladislaw (1977), *Correspondence Principle and the Growth of Knowledge*, D. Reidel, Dordrecht.

Kuipers, Theo (2000), *From Instrumentalism to Constructive Realism*, Kluwer, Dordrecht.

Lange, Marc (2015), "On 'Minimal Model Explanations': A Reply to Batterman and Rice", *Philosophy of Science* 82, 292–305.

Liu, Chang (1999), "Approximation, Idealization, and Laws of Nature", *Synthese* 118, 229–256.

Mäki, Uskali (1994), "Isolation, Idealization and Truth in Economics", in Bert Hamminga and Neil De Marchi (eds.), *Idealization in Economics*, Rodopi, Amsterdam, 147–168.

Mäki, Uskali (2009), "MISSing the World: Models as Isolations and Credible Surrogate Systems", *Erkenntnis* 70, 29–43.

McMullin, Ernan (1985), "Galilean Idealizations", *Studies in History and Philosophy of Science* 16, 247–273.

Niiniluoto, Ilkka (1987), *Truthlikeness*, D. Reidel, Dordrecht.

Niiniluoto, Ilkka. (1989), "Theories, Approximations, Idealizations", in Jerzy Brzezinski et al. (eds.), *Idealization I: General Problems*, Rodopi, Amsterdam, 9–57.

Niiniluoto, Ilkka (1999), *Critical Scientific Realism*, Oxford University Press, Oxford.

Niiniluoto, Ilkka (2002), "Truthlikeness and Economic Theories", in Uskali Mäki (ed.), *Fact and Fiction in Economics: Models, Realism, and Social Construction*, Cambridge University Press, Cambridge, 214–228,

Niiniluoto, Ilkka (2005), "Abduction and Truthlikeness", in Roberto Festa et al. (eds.), *Confirmation, Empirical Progress, and Truth Approximation*, Rodopi, Amsterdam, 255–275.

Niiniluoto, Ilkka (2013), "Models, Simulations, and Analogical Inference", in Vassilios Karakostas & Dennis Dieks (eds.), *EPSA 11 Perspectives and Foundational Problems in Philosophy of Science*, Springer International, Cham, 19–27.

Nowak, Leszek (1971), "The Problem of Explanation in Karl Marx's *Capital*", *Quality and Quantity* 5, 311–330.

Nowak, Leszek (1972), "Laws of Science, Theories, and Measurement", *Philosophy of Science* 39, 533–548.

Nowak, Leszek (1980), *The Structure of Idealization: Towards a Systematic Interpretation of the Marxian Idea of Science*, D. Reidel, Dordrecht.

Peirce, Charles S. (1931–35), *Collected Papers* 1–6, ed. by C. Hartshorne and P. Weiss, Harvard University Press, Cambridge, MA: (CP)

Popper, Karl R. (1972), *Objective Knowledge: An Evolutionary Account*, Clarendon Press, Oxford.

Rohwer, Yasha & Rice, Collin (2013), “Hypothetical Pattern Idealization and Explanatory Models”, *Philosophy of Science* 80, 334–355.

Rudner, Richard (1966), *Philosophy of Social Science*, Prentice-Hall, Englewood Cliffs.

Rueger, Alexander & Sharp, W. David (1996), “Simple Theories of a Messy World: Truth and Explanatory Power in Nonlinear Dynamics”, *The British Journal for the Philosophy of Science* 47, 93–112.

Salmon, Wesley C. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press, Princeton.

Sellars, Wilfrid (1963), *Science, Perception, and Reality*, Routledge & Kegan Paul, London.

Shrader-Frechette, Kristin (1989), “Idealized Laws, Antirealism, and Applied Science: A Case in Hydrogeology”, *Synthese* 81, 329–351.

Strevens, Michael (2008). *Depth: An Account of Scientific Explanation*, Harvard University Press, Cambridge, MA.

Tuomela, Raimo (1979), “Scientific Change and Approximation”, in Ilkka Niiniluoto & Raimo Tuomela (eds.), *The Logic and Epistemology of Scientific Change*, North-Holland, Amsterdam, 265–297.

van Fraassen, Bas (1980), *The Scientific Image*, Clarendon Press, Oxford.

Wayne, Andrew (2011), “Expanding the Scope of Explanatory Idealization”, *Philosophy of Science* 78, 830–841.

Weisberg, Michael (2007), “Three Kinds of Scientific Idealizations”, *Journal of Philosophy* 104, 639–659.

Weisberg, Michael (2013), “Getting Serious about Similarity”, *Philosophy of Science* 79, 785–794.

Winsberg, Eric (2010), *Science in the Age of Computer Simulation*, The University of Chicago Press, Chicago.