teorema

Revista Internacional de Filosofía

Vol. XXXVIII/3 • 2019

KRK EDICIONES

teorema

Revista internacional de filosofía

teorema *Revista internacional de filosofía* es una publicación cuatrimestral que aparece en febrero, mayo y octubre. Anualmente edita el suplemento **limbo** *Boletín internacional de estudios sobre Santayana.* Aunque se tomarán en consideración artículos pertenecientes a cualquier disciplina filosófica, **teorema** presta una atención especial a aquéllos que, preferentemente en español e inglés, discutan temas de lógica, filosofía del lenguaje, filosofía de la lógica, filosofía de la mente, ciencia cognitiva, filosofía e historia de la ciencia, teoría del conocimiento, metafísica, y otras áreas relacionadas. Es propósito de **teorema** dedicar especial consideración al pensamiento español en todas sus épocas y facetas. **teorema** publica también, principalmente por invitación, notas críticas y recensiones; sin embargo, las propuestas de publicación en este ámbito son muy favorablemente acogidas. La revista **teorema** sigue el procedimiento de revisión externa y anónima por pares. Los contenidos de la revista están recogidos, entre otras, en las siguientes fuentes bibliográficas: Arts and Humanities Citation Index®, Current Contents©/Arts and Humanities, Carhus Plus+, Dialnet, Dice, Elsevier Bibliographic Databases (SCOPUS), Fuente académica, ISOC-CSIC, Jstor, Latindex, Periodicals Index Online, Répertoire bibliografique de la philosophie, RESH, Sumaris CBUC, The Philosopher's Index, y Ulrich's Periodicals Directory. **teorema** ha sido declarada "revista de excelencia" por FECYT (Fundación Española para la Ciencia y la Tecnología), organismo dependiente del Ministerio de Economía y Competitividad del Gobierno de España.

teorema *Revista internacional de filosofía* is a four-monthly journal (issues in February, May and October). **limbo** *Boletín internacional de estudios sobre Santayana* is included as an annual supplement. Although papers in any philosophical discipline will be considered, the main aim of the journal is to publish original articles either in Spanish or in English in Logic, Philosophy of Language, Philosophy of Mind, Cognitive Science, Philosophy and History of Science, Epistemology, Metaphysics and related areas. The study of Spanish Thought from any period or discipline will be given special consideration. Although critical notices and book reviews are usually invited, suggestions are welcome. **teorema** is a blind- and peer-reviewed journal. The contents of **teorema** are indexed and collected in the following bibliographic sources: Arts and Humanities Citation Index®, Current Contents©/ Arts and Humanities, Carhus Plus+, Dialnet, Dice, Elsevier Bibliographic Databases (SCOPUS), Fuente académica, ISOC-CSIC, Jstor, Latindex, Periodicals Index Online, Répertoire bibliografique de la philosophie, RESH, Sumaris CBUC, The Philosopher's Index, and Ulrich's Periodicals Directory. **teorema** has been listed as a "journal of excellence" by the Spanish Government Agency FECYT (Foundation for Science and Technology).

Los contenidos de **teorema** desde 1971 están accesibles libremente en <www.dialnet.unirioja.es>.

For free access to back issues of **teorema** from 1971 up to last year go to <www.dialnet.unirioja.es>.

Redacción/Editorial Office: **teorema**. Universidad de Oviedo, Edificio de Servicios Múltiples, Campus de Humanidades, E-33071, Oviedo, Spain. **teorema**, apartado 702, E-33080, Oviedo, Spain.
Phone: (34) 98 5104378, fax: (34) 98 5104385, teorema@uniovi.es, www.unioviedo.es/Teorema, www.revistateorema.com
Suscripciones/Subscriptions: Ediciones Krk, Álvarez Lorenzana 27, E-33006 Oviedo, Spain;
phone & fax: (34) 98 5276501, correo@krkediciones.com, www.krkediciones.com. dl.:as-1736-2015

ÍNDICE/TABLE OF CONTENTS

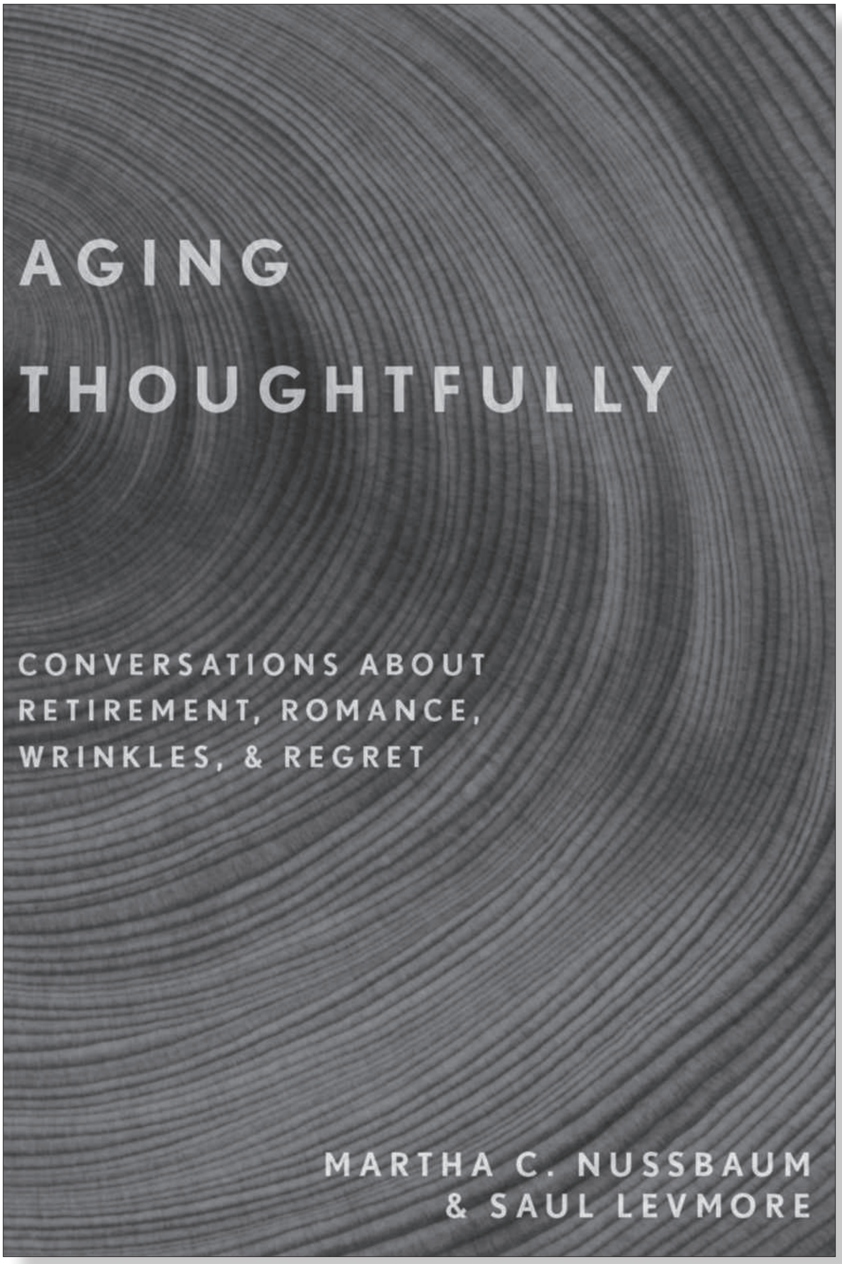
SECCIÓN MONOGRÁFICA/SPECIAL SECTION

LA EXPLICACIÓN EN CIENCIA/EXPLANATION IN SCIENCE

GUEST EDITOR: VALERIANO IRANZO

NOTA CRÍTICA/CRITICAL NOTICE

OBITUARIO/OBITUARY


AGING
THOUGHTFULLY
CONVERSATIONS ABOUT
RETIREMENT, ROMANCE,
WRINKLES, & REGRET
MARTHA C. NUSSBAUM
& SAUL LEVMORE

teorema
Vol. XXXVIII/3, 2019, pp. 5-23
ISSN: 0210-1602
[BIBLID 0210-1602 (2019) 38:3; pp. 5-23]

Introduction: Explanation in Science

Valeriano Iranzo

I. A LONG (AND WINDING) PHILOSOPHICAL DEBATE

There was a time when explaining was not considered a legitimate aim for science. Pierre Duhem and Ernst Mach, to name but two of the most representative authors, justified their scruples about explanation by invoking the autonomy of physics with respect to metaphysics and the economy of thought, respectively. The prevailing philosophical view on science at the turn of the nineteenth century was that science has to do primarily with "representing" (Duhem), "anticipating experiences" (Mach), … rather than to explaining. This may sound, indeed, a bit strange to us. After all, most scientists and philosophers of science nowadays admit that explanation is not only a legitimate aim for science, but also a valuable one. A Nobel Prize recipient in physics, Steven Weinberg, claimed that: "...the aim of physics at its most fundamental level is not just to describe the world but to explain why it is the way it is" [Weinberg (1994), p. 169]. Philosophers of science as different as Philip Kitcher and Bas van Fraassen, to mention just two examples, acknowledge that: "A crucial part of a scientist's practice consists in her commitment to ways of explaining the phenomena" [Kitcher (1993, p. 82]; "…the search for explanation is valued in science because it consists for the most part in the search for theories which are simpler, more unified, and more likely to be empirically adequate" [van Fraassen (1980), pp. 93-4]. However, it was not until almost the middle of the 20th century that explanation gained respectability thanks to Carl Gustav Hempel. His "covering law model", which can be found prefigured in other authors of the time (like Popper), became the background philosophical lore about explanation for several decades.

Explanation was understood by Hempel –in line with Logical Positivism's core assumptions– as a relationship between statements. Thus, the statement that describes the event to be explained is the explanandum; the set of further statements required to explain it is the explanans. A fundamental constraint here is that, in addition to statements referring to initial conditions, the explanans must include at least one law, so that this particular sort of general statements –lawlike statements– are essential for doing the explanatory work. On the other hand, Hempel initially insisted that the inferential link between explanans and explanandum should be deductive –hence the so-called deductive-nomological (DN) model [Hempel and Oppenheim (1948); Hempel (1965b)]. Only logical and semantic properties of the statements are taken into account in the analysis of scientific explanation. Ontological concerns, those that could offend the empiricists' feelings of the time, were carefully avoided.

Nevertheless, and despite the subsequent modifications introduced by Hempel –allowing cases in which the explanandum is not deductively followed from the explanans–[1] his proposal soon came under devastating criticism. It can be said that in the late 1960s there was a widespread consensus that the Hempelian covering law model is untenable. Alternative approaches were developed. A standard classification distinguishes four subsets: probabilistic, unificationist, pragmatist, and causal-mechanical accounts. Until 1995 approximately this multiplicity of options coexisted, but from then on there was a noticeable change in that the causal approaches to the explanation in its different variants (interventionist, mechanistic, ...) were those clearly favoured by the academic community. Thus, even though few authors would claim that "asking for explanations" simply equates to "asking for causes", many of them would subscribe that any acceptable philosophical account of scientific explanation is forced to deal with *causal* explanations. That means that reflection on explanation involves also reflection on the notion of causal relation, if not also on the notion of cause itself –an item virtually absent in the Hempelian approach.[2] Wesley Salmon summarizes this change of mentality in the philosophical community as follows:

> There is a fundamental intuition –…– according to which causality is intimately involved in explanation. Those who are familiar with Hume's critique of causality may deny the validity of that intuition by constructing non-causal theories of scientific explanation. Others may skirt the issue by claiming that the concept of causality is clear enough already, and that further analysis is unnecessary. My own view is (i) that the intuition is valid –scientific explanations does involve causality in an extremely fundamental fashion– and (2) that causal concepts do stand in serious need of further analysis.[3]

This paragraph was firstly published in 1984, but Salmon's statement conveys a generalized attitude among philosophers of science at the early nineties. Here is a brief sketch of the story that led to Salmon's predicament.[4]

a) *Explanation as unification*

Michael Friedman and Philip Kitcher endorsed two different unificationist accounts of explanation. Friedman defended that explanation is tantamount to unification and the latter is understood as "reducing the total number of independent phenomena that we have to accept as ultimate or given" [Friedman (1974), p. 15)]. The law of ideal gases, for instance, is explained by the kinetic theory of gases insofar as a number of independently acceptable phenomena –unexplained phenomena, actually- are reduced to one. In the vein of the Hempelian account, the explanatory task is attached to laws, especially to the more comprehensive theoretical ones. Kitcher, in turn, underwrites that we "derive descriptions of many phenomena, using the same pattern of derivation again and again, and in demonstrating this, it teaches how to reduce the number of types of fact that we accept as ultimate" [Kitcher (1989), p. 432]. Theories unify to the extent that they provide one pattern (or a few number of patterns) to derive the greatest number of sentences accepted by the scientific community. An *argument pattern* is an ordered triple composed by a schematic argument (a sequence of *schematic sentences*; i.e.: sentences in which some of the non-logical vocabulary has been replaced by dummy letters), *filling instructions* for completing the dummy letters in the schematic sentences, and *classifications* (they describe which sentences in schematic arguments are premises and conclusions). Here is an example:

QUESTION: Why do the members of *G*, *G'* share *P*?

ANSWER:

(1) *G*, *G'* are descended from a common ancestor G_0

(2) G_0 members had *P*.

(3) *P* is heritable.

(4) No factors intervened to modify *P* along the G_0-*G*, G_0-*G'* sequences.

Therefore, (5) Members of *G* and *G'* have *P*.

In this example there are five schematic sentences. Filling instructions require that *G*, *G'*, G_0 be replaced by names of groups of organisms, and that *P* be replaced by the name of a trait of organisms. Finally, the classification would state that (1)-(4) are the premises and that (5) is the conclusion deduced from them [Kitcher (1993), p. 83].

Generally speaking, the fewer the argument patterns employed and the larger the number of sentences derived, the better systematization we have. Particularly, the "explanatory store" over a corpus of statements *K* –all those currently accepted by the scientific community– is the best systematization of *K*, that is, the minimal set of explanatory patterns which allow the derivation of *K*.

The goal is unification, yes, but the explanatory import is attached to particular argument patterns –explanatory schemata–, and not to the most basic regularities found in nature (pace Friedman). We look, rather, for the minimal explanatory store for *K*. However, both authors agree on the idea that the explanatory relationship is a deductive relationship. Kitcher does not explicitly demand the necessity of laws for putative explanations, but he stills endorses the idea that explaining equates to giving an argument whose (deductive) conclusion is the explanandum.

How does Kitcher's unificationist approach tackle the problems previously raised against Hempel? The flagpole example is one of the most famous counterexamples against the Hempelian D-N model. A flagpole shadow is entailed by the height of the pole plus the angle of the sun above the horizon plus laws about the rectilinear propagation of light. Consequently, the flagpole shadow –the explanandum– is "D-N explained". But it is also true that we could change the argument so that the height of the pole is entailed by the flagpole shadow plus the remaining items. However, we would not say that the height of the pole is explained by its shadow (plus the other items). Unfortunately, the D-N model does not discriminate between cases where the explanatory relation is asymmetrical, even though the deductive constraint is fulfilled.[5]

Now, what is the answer provided by Kitcher to the flagpole counterexample? When confronted to those asymmetries, he resorts to our entrenched argument patterns. He argues that here we have two explanatory schemata: the "origin and development pattern" and the "shadow-pattern". The former appeals to the conditions under which the object originated and the subsequent changes it has suffered; the latter invokes the shadow of objects to derive their dimensions. The "origin and development pattern" should be favoured, according to him, because the "shadow pattern" does not allow us to derive the dimensions of those objects which do not have shadows. Given that the number of sentences derived by the "shadow pattern" is less than those that can be derived from the alternative pattern, the latter is more unifying and should be preferred because of its higher explanatory value.

On account of this example, someone could think that the most explanatory patterns according to Kitcher are precisely those that fit with the causal order of the phenomena explained. But he insists that there is no objective causal structure in the world to ground the asymmetries of explanatory relationships: "one event is causally dependent on another just in case there is an explanation of the former that includes a description of the latter" [Kitcher (1989), p. 420]. Putting the matter in other words, our judgments/beliefs about causality just mirror our judgments/beliefs about explanatory relationships.

b) *Explanation as Statistical Relevance*

A further difficulty for Hempel's approach has to do with explanatorily irrelevant information. "Mr. Jones fails to get pregnant" –the alleged explanandum– is a deductive consequence from "All males who take birth control pills regularly fail to get pregnant" plus "Mr. Jones is a male" plus "Mr. Jones has been taking birth control pills regularly". Again, "Mr. Jones fails to get pregnant" is "explained" according to the D-N model, but we do not consider this is a putative explanation [Salmon (1971), p. 34]. Of course, taking birth control pills have no effect concerning pregnancy in males, so why should we consider it has any explanatory import for this particular explanandum?

The moral of the story is that only *relevant* information should be counted when explaining an event. Salmon's "Statistical Relevance" (S-R) model appeals to a probabilistic criterion. The idea is that a bit of information is explanatorily relevant if and only if it is statistically relevant, that is, if it affects the probability of what has to be explained. Since taking birth pills does not increase/decrease the probability of Mr. Jones getting pregnant, it has no epistemic import at all for it. Putting the matter in formal terms, if M=male, T=taking birth pills, and P=pregnancy, $p(P|M \& T) = p(P|M) = 0$. However, being F=female, and taking for granted that the percentage of females who get pregnant after taking the pills is less than that of those females who do not take the pills, $p(P|F \& T) \neq p(P|F)$. Therefore, T is explanatory relevant for F (regarding P), but completely irrelevant for M.

According to the S-R model, an explanation for a particular event is all the information statistically relevant to it, that is, the set of all factors that make any difference to the probability of the event. It's worth noticing here that both Hempel and the unificationists agreed on the idea that explaining an event is making it expected. Explanation demands a set of statements –laws, descriptions of initial conditions, explanatory schemata, …– that either entail or make highly probable the explanandum. But Salmon's S-R model departs from this assumption. Strictly speaking, to give an explanation equates to providing a probability distribution rather than providing an argument whose conclusion is the explanandum. Certainly, we must be careful to get the correct probability values and also not to overlook any statistically relevant factor involved. And this, and only this, is all we need to explain an event, regardless of its probability value. In fact, a highly improbable event may be explained by citing the relevant conditional probabilities. A consequence of this is that inconsistent explananda may be appropriately explained by the same corpus of information. If the aforementioned constraints are fulfilled, the explanation is fully satisfactory for both explananda. Here is an example:

> Two patients, *x* and *y*, are infected by streptococcus. Let V = recovery, T (¬T) = 'treated (untreated) with penicillin', and R (¬R) 'the strain is resistant (non-resistant)'. According to our medical statistics, p (V|T & ¬R) = 0.9; p (V|¬T & ¬R) = 0.4; p (V|T & R) = 0.1; p (V|¬T & R) = 0.1

Now, let's suppose that *x* has been infected by a resistant strain and *y* by a non-resistant one but, after receiving the treatment, both of them recover. The relevant information for explaining both events is the same, no matter that *x*'s recovery is much more unlikely than *y*'s recovery. Furthermore, the same information should be taken into account for explaining two inconsistent explananda (i.e.: *x*'s recovery and *x*'s non-recovery).[6]

This could be considered as a counterintuitive consequence of the S-R model. Notwithstanding, the main limitations for it have to do with the prospects to grasp causal links by means of statistical dependencies. Let's see what these are.

Science students are advised at introductory courses in scientific methodology not to confuse correlations with causes. If A is the cause and B is the effect, then presumably p (B|A) > p (B). Two events causally related are statistically dependent since the cause raises the probability of the occurrence of the effect. But very often the way we proceed in science is, firstly, collecting data about a potential association/correlation between the variables (measuring frequencies, for instance), and secondly, inferring a causal relation from those data. But after detecting a statistical dependency between A and B four possibilities remain open: (i) A causes B; (ii) B causes A; (iii) A and B are effects of a common –and, perhaps, unknown– cause; (iv) A and B are associated by chance. Another famous example nicely shows how the S-R model can circumvent this difficulty. The reading of a barometer (B) and the occurrence of a storm (S) are highly correlated so that p (B|S) > p (S) –and it is also the case that p (S|B) > p (S). But we would hardly consider that B explains S (nor, alternatively, that S explains B) since both events are explained by a common cause, i.e.: the decrease of the atmospheric pressure (P).

The S-R model perfectly fits with our intuitions about this example. Since p (S|P) = p (S|P&B), B is statistically irrelevant to S given P. But P is statistically relevant to S given B, because p (S|B) ≠ p (S|P&B). Analogously, p (B|S) = p (B|P&S) –so, S is irrelevant to B given P. And p (B|S) ≠ p (B|P&S), so P is relevant to B given S. Shortly, P explains B and also S, but neither B explains S nor S explains B. In a situation like this it is said that B is *screened off* from S by P (and also that S is screened off from B by P).[7]

But winning a battle is not like winning the war. The point is that causal nets are not always statistically indistinguishable and different causal networks can accommodate the same class of probability distributions: "…, the resolving power of any possible method for inferring causal structure from statistical relationships is limited by statistical indistinguishability. If two causal structures can equally account for the same statistics, then no statistics can distinguish them" [Spirtes *et al.* (2000), p. 59].

Even imposing reasonable constraints on those probability distributions in principle intended to infer causal relationships –the Causal Markov Condition and the Minimality Requirement–, statistical indistinguishability cannot be avoided. Causal nets may be underdetermined by conditional probabilities. And we should not think that this is a problem just for very complex causal structures. An example of "strongly statistical indistinguishability" is:[8]

G_1 = A causes C; D causes B; B causes C; A causes D.

G_2 = A causes C; D causes B; B causes C; D causes A.

It's worth adding that this is a serious objection against the S-R model only insofar as it is taken for granted that explaining an event is closely related to locating it in *a causal network*, to making explicit its *causal history*, and so on. But the fact is that after several decades of debate "the great majority of philosophers is convinced that an account of explanation must provide a starring, if not exclusive role for causation" [Strevens (2014), p. 48)]. Woodward (2003), where it is defended a causal-interventionist interpretation of explanation, was probably a definitive turning point in this direction even though further versions within the causal framework –not necessarily interventionist– has been subsequently developed. The causal-mechanistic approach, which traces back to the eighties [Salmon (1984)], for instance, reemerged with strength –expurgated from its strongest physicalist commitments– at the turning of the century [Machamer, Darden & Craver (2000)]. Right now, it is surely the most discussed option in the literature.[9]

The list of papers included below reflects this state of the matter. Three of those four related to the analysis of explanation (those of J. Reiss, S. Psillos & S. Ioannidis, and S. Pérez-González) discuss problems internal to the causal tradition –two of them are particularly concerned with the explanatory import of mechanisms.

However, despite the widely predominance enjoyed nowadays by the causal tradition, that's not the full story. The consensus around the centrality of the notion of cause in order to explicate explanation does not entail assuming that there are no exceptions. Some authors have pointed at the limits of causal explanation through particular examples mainly –but not always– taken from physics [Lange (2016)]. The contribution of J. Suárez & R. Deulofeu, see below, goes along the non-causalist path but appeals to an episode of biology. Equilibrium explanations –a sort of ubiquitous explanation in biology and economics– are those favoured examples that, supposedly, cannot be reduced to the causalist-mechanical framework.[10]

This point raises some doubts about the prospects for giving an all-encompassing analysis for scientific explanation. Given the huge variety that can be found among different scientific fields, is it reasonable to look for "explanatory monism", so to say? It has been maintained that laws do not play a basic role in biology, for instance, in contrast to what happens in physics. Granted that, an account of explanation which exploits the explanatory import of laws is handicapped when dealing with biomedical sciences –conversely, physics would, in principle, be a more comfortable place for unificationist accounts. Analogously, mechanisms seem specially fitted for explanation in medicine, bio-chemistry, geology, … But, what about social sciences? Even though talking about "social mechanisms" may be perfectly sound, it is debatable to what extent the sort of mechanistic explanation for fluctuations in the financial markets are similar to that invoked concerning the DNA replication in meiosis, for instance. Comparative detailed research, focused on specific scientific episodes, is required here. Even though addressing this issue is beyond the scope of this introduction, we will have a brief look at those views which highlight the contextualist constraints –not necessarily related to the peculiarities of the scientific fields– operating on explanation.

c) *The Contextual/Pragmatic Dimension of Explanation: Is That All?*

Given the difficulties to provide a general characterization of explanation, some authors have insisted that explanation is irremediably *contextual*. These approaches, labelled as "pragmatic" accounts" of explanation, highlight the relation between the explainer and her audience.[11] They are focused on questions as the assumptions required in the act of explaining to get some understanding for the audience, the role played by the agents' beliefs and interests concerning what counts as a correct explanation, the peculiarities of explanations related to idiosyncratic domains, …

Pragmatic approaches are intended to cast doubt on the philosophical task of giving a general or "structural" definition of explanation, like all those aforementioned. However, it is debatable to what extent the issues raised by pragmatic accounts cannot be accommodated in those standard approaches. The contextual relativity of explanation could be restricted, perhaps, to accepting that an amount of information related to the local context where the explanatory demand arises may be highly relevant. But this does not mean that contextual factors turn explanation into a purely psychological or subjectivist affair.[12]

Putting at the forefront the pragmatic dimension of explanation introduces a further topic deserving of attention. At the outset of this introduction we pointed out that philosophers and scientists nowadays agree that explanation is a matter of concern in scientific research. Theory-building, in particular, is driven –albeit, non-exclusively– by this concern. And, in principle, scientists prefer *theories* that unify different phenomena or domains, …, that have diverse empirical consequences (and some of them at least, about novel phenomena), that can be embedded in our background scientific knowledge, that are simple, … It could be said, then, that generally speaking scientists prefer good explanations to theories that score badly in those factors –commonly called 'theoretical virtues'. The debatable issue here, however, is whether these explanatory advantages have any confirmational import. When confronted with two alternative explanations for the same explanandum, should we consider that the best one qua explanation is also more confirmed than the other? Alternatively, if we confer more credibility to the best explanation of both, precisely because it shows better explanatory credentials, are we also favouring the most confirmed option of both?

Thinking that explanatory goodness increases the plausibility of the *explanans* is a key idea for partisans of "inference to the best explanation" (IBE, hereafter).[13] A standard way of introducing this inferential pattern is:

(P_1) F is some fact or collection of facts.

(P_2) Hypothesis H_1, if true, would explain F.

(P_3) No competing explanations (H_2, H_3, ..., H_n) would explain F better than H_1.

(Conclusion) One is justified in believing that H_1 is true.

The peculiarity of IBE is that the conclusion –the *explanans; H_1* in this example– is inferred because of its explanatory yieldings about a particular explanandum. However, this is somewhat ambiguous. Thus, those who subscribe the importance of IBE do not entirely agree about its role. While some authors think it is primarily related to the context of discovery (IBE understood as a heuristical strategy), other authors insist that it has full epistemic import (see Iranzo (2007) for further discussion). There are still those overtly sceptics about IBE who do not consider that IBE refers to a specific inferential pattern whose reliability must be taken for granted. Bas van Fraassen, for instance, claims that the explanatory appeal of a hypothesis, however great, does not provide any *confirmational* advantage for the explanatory hypothesis. Rather, that feature is just an *informational* virtue –to use van Fraassen's words– that can be justified by pragmatic reasons alone [van Fraassen (1980), p. 87 and ff.].

It could be argued that differences between good and bad scientific explanations could hardly be qualified unless a consensus on what is explanation is reached. But the fact is that both debates –the nature of explanation and the significance and the epistemic value of IBE– have been developed separately for decades. Whatever it is, current discussion on this issue has evolved along two main paths.[14] Firstly, elaborating a precise characterization of the various virtues encompassed under the generic label of 'explanatoriness'; secondly, forging a conceptual link between IBE and Bayesianism, which is the most well-established theory of confirmation at present.[15] It is expected then, that a careful scrutiny of those properties that qualify a hypothesis as a good explanation are somehow positively connected to its probability or degree of confirmation. Admittedly, the results obtained do not always play in favour, far from it, of the explanationists, that is, in favour of those who attach an epistemic (confirmational) import to "explanatoriness". W. Roche & E. Sober argue head-on against this view, while J. Schupbach defends IBE against a popular, potential criticism (see both papers below).

II. THE PAPERS

Four of the six contributions included in this monographic section –those of REISS, PSILLOS & IOANNIDIS, PÉREZ-GONZÁLEZ and SUÁREZ & DEULOFEU– are devoted to the analysis of explanation itself: what it is and how could we understand it, if possible, in terms of a more fundamental or pristine notion (causation, mechanism, ….). It should be added that Reiss and Psillos & Ioannidis address this question from a general perspective, while Pérez-González and Suárez & Deulofeu are focused on particular scientific disciplines (economics and biology, respectively). There are two more contributions, those of ROCHE & SOBER and SCHUPBACH, that are devoted to "inference to the best explanation" (IBE). The general concern here is whether the empirical assessment of hypotheses should be constrained by their respective explanatory merits. While Roche & Sober defend a skeptical argument against this possibility, Schupbach offers an interpretation of IBE that allows it to sidestep the so-called challenge of conjunctive explanations. Let's pause on all this.

According to explanatory causalism explaining an event has to do with ascertaining the causes that provoke it so that causality is the grounding notion for explanation. A basic associated insight is that scientific explanation is objective insofar as it reveals the framework of causal relationships actually operating in a particular context. In "Causal Explanation: Is All There Is to Causation?", Julian Reiss argues that absence causation is a challenge not only for physicalist and realist theories of causation but also for counterfactual and difference-making ones. He suggests an anti-objectivist account of causation —he explicitly acknowledges its Humean flavour— in order to cope with this problem: causes are inferred from explanatorily successful stories. They are picked out by virtue of explanatory considerations since there is no objective causal structure in the world which legitimate causal explanations should reflect. His slogan is: "Explanation comes first; causation, second". Reiss defends that explanations are a kind of speech acts, i.e.: "transfers of understanding" between agents. Causal explanations, in particular, are those explanations which enable agents to make plausible causal inferences. But they are not considered "causal" to the extent that the explanans provides information about the causal history of the explanandum. Rather, what counts as causal explanation is established according to "social norms for causal inference" or "intersubjective facts about inferential practice". Among those norms Reiss mentions the evidential standards to trade-off between Type-I and Type-II errors in statistics or the injunction to discard alternative causal hypotheses before asserting a causal claim. Rules like these are, indeed, the effective constraints on causal explanation.

In "Mechanistic Causation: Difference-Making is Enough", Stathis Psillos and Stavros Ioannidis assume that causal explanation is crucial in scientific practice. Although they agree with Reiss on this point, they focus on an influential way to understand causal explanations, that is, on mechanistic accounts of it. In contrast to Reiss's approach, however, they think that causation "through mechanisms" comes first and explanation, second. According to them mechanisms are: (i) what turn a relation between A and B into a causal relation and (ii) what give causes their explanatory import. Shortly, mechanisms are necessary to causation and also to scientific explanation. They criticize, however, the prevailing account about mechanisms, according to which mechanisms essentially involve activities (in addition to entities, properties and relations). Psillos and Ioannidis think, rather, that "difference making is prior to production". Mechanisms are "networks of difference-making relations" –the latter usually understood in terms of counterfactual dependence– for them. Admittedly, activities are implemented to account for the productive dimension of mechanisms: a mechanism produces a result that can be properly considered as its effect. But Psillos and Ioannidis argue that establishing causality necessarily involves contrary-to-fact commitments.

Nevertheless, even taking for granted that understanding causality in terms of production cannot avoid difference making (since A cannot be the putative cause of B, unless A makes some difference to the occurrence of B), we could still think that that is not enough. In response to this Psillos and Ioannidis insist that mechanism is a concept effectively used in scientific practice. They resort to an episode in the history of medicine –i.e.: the discovery of deficiency in vitamin C as the cause of scurvy– which is actually an example of absence causation (recall that this was the leading issue in Reiss's paper). They maintain that scientific practice demands reconstruction of "stable causal pathways", certainly, but identifying and detailing them equates to detecting the factors which make differences concerning the disease. The sort of evidence invoked here is not some sort of "mechanistic evidence" qualitatively distinct from evidence about difference-making relations. They conclude, then, that no metaphysical baggage related to activities, powers or capacities is required to understand the causal/explanatory role played by mechanisms in science.

Advocates of the mechanistic standpoint on explanation think that mechanisms play a substantial role in nearly all scientific domains. Besides, most of them think that an appropriate notion of mechanism should be suitable for all those domains. In "The Search for Generality in the Notion of Mechanism", Saúl Pérez-González discusses the prospects for such project. According to him, the development of an all-encompassing notion of mechanism is pursued through two different and alternative strategies. The "extrapolation strategy" tries to articulate a notion of mechanism taking one or a few fields of science as reference, and then applies that notion to the remaining fields. The "across-the-sciences" strategy consists of thinking about how mechanisms are understood across all the sciences and elaborates a notion of mechanism that includes just the shared features. After analysing paradigmatic examples of both strategies, Pérez-González argues that both face outstanding difficulties. The extrapolation strategy leads to notions unable to account for the varieties of mechanisms, while the across-the-sciences strategy leads to vacuous characterizations of mechanisms. He concludes that the search for generality does not look promising and suggests that it would be preferable to develop field-specific notions of mechanism.

A different approach is endorsed in "Equilibrium explanation as structural non-mechanistic explanations: The case of long-term bacterial persistence in human hosts". Javier Suárez and Roger Deulofeu depart from the widespread acceptance of the "New Mechanism" standpoint with the aim of questioning its universality. In contrast to the causal-mechanistic framework, they appeal to "structural explanations", that is, explanations that account for the phenomenon to be explained in virtue of the mathematical properties of the system where the phenomenon obtains, rather than in terms of the mechanisms that causally produce the phenomenon. Structural explanations are very diverse in kind depending on the relevant structural properties invoked (bowtie structures, topological properties of the system, equilibrium constraints). Suárez and Deulofeu focus on a particular biological model, i.e., Blaser and Kirschner's nested equilibrium model of the stability of persistent long-term human-microbe associations. After investigating the role played by the mathematical properties of this model, they consider that it has fully explanatory import since: (i) it provides a set of differential equations —a mathematical structure— that satisfies an evolutionarily stable strategy (ESS); (ii) the explanation of host-microbe persistent associations is robust to any perturbation due to the nested nature of the ESSs; and more importantly for their case, (iii) this is so because the properties of the ESS directly mirror the properties of the biological system *in a non-causal way*. They conclude that this example vindicates the claim that equilibrium explanations look more similar to structural explanations than to causal-mechanistic ones.

Two further papers cope with the alleged link between explanatory value and inference.

In "Inference to the Best Explanation and the Screening-Off Challenge" Roche & Sober argue that "explanatoriness" is evidentially irrelevant. The "screening-off" thesis (SOT) affirms that the statement 'H would explain O if H and O were true' adds nothing at all to the empirical support that O by itself gives to H. The formal rendition of this is: p (H|O&EXPL) = p (H|O), where EXPL is the proposition that if H and O were true, then H would explain O. The main example for them is an extrapolation from a frequency estimate found in a sample to a particular member of the population. Thus, if *freq* (heavy smoking before age 50 | lung cancer after age 50) = α, and Joe —a random member of the population not included in the sample— got lung cancer after fifty, the probability that Joe was a heavy smoker before age 50 given that he got lung cancer after fifty —that is, p (H|O)— equates to α. Now, if we add EXPL —i.e.: the proposition that if H and O were true, H would explain O—, then p (H|O) = p (H|O&EXPL) = α. Consequently, EXPL is evidentially irrelevant to H.

Roche & Sober qualify the scope of SOT to examples in which the background information includes frequency data. However, they claim that there are realistic cases, similar to the aforementioned example, which fulfil this condition. Furthermore, they think that these cases go against IBE. They discuss two versions of IBE according to which inferring (=believing) H is licensed when H is the best potential explanation and also when H's overall score regarding the explanatory virtues usually invoked in this debate (explanatory power, fertility, parsimony, ….) is high. Roche & Sober argue that even for these strengthened versions of IBE there are realistic counterexamples where all those explanatory considerations are screened-off by O. From this they conclude that there are corresponding versions of SOT –logically stronger than it, indeed– that undermine IBE.

Jonah Schupbach's paper ("Conjunctive Explanations and Inference to the Best Explanation") starts with an observation that is hardly disputable, i.e.: that sometimes there are different potential explanations for the same explanandum. This may occur both in everyday and scientific contexts. In case that accepting them all (or, at least, two of those explanations) provides us with a richer explanation, we have a "conjunctive explanation". At first sight, however, IBE urges us to infer the best option among *competing* explanatory hypotheses. But, if competition occurs just when hypotheses are incompatible (either because they are directly inconsistent by themselves or because the available evidence renders them incompatible), conjunctive explanations are straightforwardly excluded from the domain of applicability of IBE. Hence, a weaker notion of competition is required. His proposal here –jointly developed in a previous paper with D.H. Glass– is to define competition between hypotheses in terms of their (dis)confirmatory relations. He suggests a measure for the "net" degree of competition, based on the log-likelihood measure of confirmation, which contains two addends. One of them is related to the "direct competition" between H_1 and H_2 –the reciprocal disconfirmational effect without taking into account the evidence E –i.e.: the explanandum. The other addend alludes to the "indirect competition" since H_1 and H_2 could be competitors relative to some explanandum E even though they are entirely compatible (because, for instance, only one of the hypotheses is needed to explain E). Particularly, direct competition takes into account conditional probabilities between H_1 and H_2 –that is, $p(H_1|H_2)$, $p(\neg H_1|H_2)$ $p(H_1|\neg H_2)$ and $p(\neg H_1|\neg H_2)$–, while indirect competition considers the likelihoods of the conjoined hypothesis and its negations with respect to E –i.e.: $p(E|H_1\&H_2)$, $p(E|\neg H_1\&H_2)$, $p(E|H_1\&\neg H_2)$, $p(E|\neg H_1\&\neg H_2)$.

Nonetheless, even though Schupbach and Glass's probabilistic explication of competition plausibly widens the domain of applicability for IBE, there are problematic cases. Schupbach discusses an example where the conjunctive explanation is the best explanation but it includes competing hypotheses (on Schupbach and Glass's weak reading of competition). Thus, we should embrace the conjunctive explanation (H_1&H_2) since: (i) H_1 and H_2 together account for the evidence better than either does individually –that is, $p(E|H_1\&H_2) > p(E|\neg H_1\&H_2)$ and also $p(E|H_1\&H_2) > p(E|H_1\&\neg H_2)$, and (ii) the available evidence separately supports both hypotheses, *even though they disconfirm one another unconditionally and conditional on E*. According to this, the core prescription of IBE –"choose the best explanation among competing hypotheses"– is challenged. Schupbach's final considerations minimize the importance of competition as a necessary requirement to apply IBE. Accordingly, after pointing at the difference between "the *single* most explanatory hypothesis" and "the most explanatory conclusion", he recommends that IBE should be interpreted as inference to the most explanatory conclusion (regardless of that conclusion's logical form) as opposed to inference to the most explanatory single hypothesis.

Departament de Filosofia
Universitat de València
Avda. Blasco Ibáñez, 30
46010 València
E-mail: iranzov@uv.es

ACKNOWLEDGMENTS

Research leading to this monographic section has been supported by the Ministry of Economy and Competitiveness (MINECO), Spain, project FFI2016-76799-P

NOTES

[1] The "inductive-statistical" explanation (I-S model) [Hempel (1965b), pp. 381 and ff.].

[2] Hempel did not completely withdraw the notion of "causal explanation". See below, footnote 5.

[3] Salmon (1997), p. 323. See also, Cartwright (2004).

[4] For a detailed story, see Salmon (1989).

[5] Hempel distinguished between "laws of coexistence" and "laws of succession" [Hempel (1965b), p. 352]. The main difference between them is that the latter ineluctably refer to time order. Usually they describe changes in a physical, biological, …, system, through differential equations. Causal explanations are, according to Hempel, a subset of D-N explanations which include laws of succession. Then, his reply to the flagpole counterexample is that the

laws involved in it are laws of coexistence, so *they are not causal laws*. Therefore, even if we have two alternative D-N explanations when we interchange explanans and explanandum, the charge cannot be that the D-N model fails because it does not adequately discriminate the causal order of events. This reply, however, is hardly convincing (see the illuminating discussion in Psillos (2002a), sect. 8.5).

[6] Incidentally, p (¬V | T & R) = 1 – p (V | T & R) = 0.9. Hence, if the same explanans is appropriate for those inconsistent explananda, then that very same explanans is appropriate for both an expected and an unexpected event.

[7] Common causes are good examples of screening-off relations, but they are not the only ones. See below the paper from W. Roche & E. Sober for a discussion in a different context.

[8] Spirtes *et al.* (2000), p. 60. Causal structures use to be represented by means of directed acyclic graphs. An introductory discussion of Bayesian nets can be found in Illari and Russo (2014), chap. 7. For more details, see Spirtes *et al.*

[9] Actually, some authors allude to the "New mechanistic" philosophy, which expands the scope of the notion of mechanism beyond philosophy of science. For a comprehensive view of the current debate on the notion of mechanism –and mechanistic explanation–, see Glennan and Illari (2018).

[10] See Reutlinger and Saatsi (2018) for a state of the art of non-causalist approaches to explanation. By the way, there are neo-Hempelian proposals still in play. An example is Diez (2014).

[11] Van Fraassen (1980), chap. 5, and Achinstein (1983) are the most refined proposals to date.

[12] See Woodward (2014) for this suggestion. The paper of Julian Reiss included below could also be seen as a compatibilist proposal between causalism and pragmatism.

[13] Presumably, the expression "inference to the best explanation" was coined by Gilbert Harman [Harman (1965)]. A historical antecedent related to IBE is Charles Peirce's term '*abduction*', a specific mode of reasoning irreducible to deduction and induction [see Campos (2011) and Psillos (2002b)].

[14] And there may be good reasons to remain so. In Cabrera (2018) it is argued that both issues should be kept separated.

[15] Some recent works on theoretical virtues are: Sober (2015), Keas (2018) and Schindler (2018). On the alleged connection between Bayesianism and IBE, see Lipton (2004) and Psillos (2007) for a positive and a negative assessment, respectively. Glymour (2015) is a critical perspective on probabilistic measures – not necessarily related to the Bayesian Criterion of Relevance to incremental confirmation, see Schupbach and Sprenger (2011)– for explanatory virtues.

REFERENCES

ACHINSTEIN, P. (1983), *The Nature of Explanation*; New York, Oxford University Press.

CABRERA, F. (2018), 'Does IBE Require a 'Model' of Explanation?'; *British Journal for the Philosophy of Science* *<https://doi.org/10.1093/bjps/axy010>.*

CAMPOS, D. (2011), 'On the Distinction between Peirce's Abduction and Lipton's Inference to the Best Explanation'; *Synthese*, vol. 180, pp. 419-42.

CARTWRIGHT, N. (2004), 'From Causation to Explanation and Back'; in B. Leiter (ed.), *The Future for Philosophy*, Oxford: Oxford University Press.

DÍEZ, J. (2014), 'Scientific W-Explanation as Specialised Embedding: a Neo-Hempelian Approach'; *Erkenntnis,* vol. 79, pp. 1413-43.

FRIEDMAN, M. (1974), 'Explanation and Scientific Understanding'; *Journal of Philosophy*, vol. 71, pp. 5–19.

GLENNAN, S. and ILLARI, P. M. (eds.) (2018), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*; Abingdon, Routledge.

GLYMOUR, C. (2015), 'Probability and the Explanatory Virtues'; *British Journal for the Philosophy of Science*, vol. 66, pp. 591-604.

HARMAN, G. (1965), 'The Inference to the Best Explanation'; *Philosophical Review*, vol. 74, pp. 88–95.

HEMPEL, C. (1965a), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*; New York, Free Press.

— (1965b), 'Aspects of Scientific Explanation'; in Hempel 1965a, pp. 331–496.

HEMPEL, C. and OPPENHEIM, P. (1948), 'Studies in the Logic of Explanation'; *Philosophy of Science*, 15, pp. 135–175. Reprinted in Hempel (1965a), pp. 245–290.

ILLARI, P. and RUSSO, F. (2014), *Causality – Philosophical Theory Meets Scientific Practice*; Oxford, Oxford University Press.

IRANZO, V. (2007), 'Abduction and Inference to the Best Explanation'; *Theoria,* vol. 22, pp. 339-46.

KEAS, M. N. (2018), 'Systematizing the Theoretical Virtues'; *Synthese* vol. 195, pp. 2761–93.

KITCHER, P. (1989), 'Explanatory Unification and the Causal Structure of the World'; in *Scientific Explanation*, P. Kitcher and W. Salmon (eds.), Minneapolis, University of Minnesota Press, pp. 410–505.

— (1993), *The Advancement of Science. Science Without Legend, Objectivity Without Illusions*; New York, Oxford University Press.

LANGE, M. (2016), *Because Without Cause. Non-Causal Explanations in Science and Mathematics*, New York, Oxford University Press.

LIPTON, P. (2004), *Inference to the Best Explanation* (2nd edition), London, Routledge.

MACHAMER, P., DARDEN L. and CRAVER, C. F. (2000), 'Thinking about Mechanisms'; *Philosophy of Science*, vol. 67, pp. 1-25.

PSILLOS, S. (2002a) *Causation and Explanation*; Chesham, Acumen.

— (2002b), 'Simply the Best: A Case for Abduction'; in A. C. Kakas and F. Sadri (eds.), *Computational Logic: Logic Programming and Beyond*, Berlin, Springer-Verlag, pp. 605–26.

— (2007), 'The Fine Structure of Inference to the Best Explanation'; *Philosophy and Phenomenological Research*, vol. 74, pp. 441-48.

REUTLINGER, A. and SAATSI, J. (eds.) (2018), *Explanation without Causation – Philosophical perspectives on Non-Causal Explanations*; Oxford, Oxford University Press.

SALMON, W. (1970), 'Statistical Explanation'; in R. G. Colodny (ed.), *The Nature and Function of Scientific Theories*, Pittsburgh, University of Pittsburgh Press, pp. 173-231.

— (1984), *Scientific Explanation and the Causal Structure of the World*; Princeton, Princeton University Press.

— (1989), *Four Decades of Scientific Explanation*: Minneapolis, University of Minnesota Press.

— (1998), *Causality and Explanation*; New York, Oxford University Press.

SCHINDLER, S. (2018), *Theoretical Virtues in Science: Uncovering Reality Through Theory*; Cambridge, Cambridge University Press.

SCHUPBACH, J. N. and SPRENGER, J. (2011), 'The Logic of Explanatory Power', *Philosophy of Science*; vol. 78, pp. 105–127.

SOBER, E. (2015), *Ockham's Razor: A User's Manual*; New York, Cambridge University Press.

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000), *Causation, Prediction and Search*; Cambride, Mass., The MIT Press.

STREVENS, M. (2014), 'Probabilistic Causality'; in L. Sklar (ed.), *Physical Theory – Method and Interpretation*; New York, Oxford University Press, pp. 40-62.

VAN FRAASSEN, B. (1980), *The Scientific Image*; Oxford, Oxford University Press.

WEINBERG, S. (1994), *Dreams of a Final Theory*; New York, Vintage Books.

WOODWARD, J. (2003), *Making Things Happen: A Theory of Causal Explanation*; New York, Oxford University Press.

— (2017), 'Scientific Explanation'; *The Stanford Encyclopaedia of Philosophy*, E. N. Zalta (ed.), URL =*<https://plato.stanford.edu/archives/fall2017/entries/scientific-explanation/>*.

RESUMEN

La presente introducción contiene dos partes. En la primera se ofrece una visión general de las principales posiciones defendidas en el debate filosófico sobre la explicación científica. En la segunda se resumen y comparan los seis artículos incluidos en la sección monográfica.

PALABRAS CLAVE: *explicación, explicación científica, inferencia hacia la mejor explicación.*

ABSTRACT

This introduction contains two parts. The first part offers an overview of the main positions developed in the philosophical debate about scientific explanation since Hempel's covering-law model. The second part summarizes and compares the six papers included in the monographic section.

KEYWORDS: *Explanation, Scientific Explanation, Inference to the Best Explanation.*