TESE DE DOUTORAMENTO

MATHEMATICAL EXPLANATION AND ONTOLOGY: AN ANALYSIS OF APPLIED MATHEMATICS AND MATHEMATICAL PROOFS

(EXPLICACIÓN E ONTOLOXÍA MATEMATICA: UNHA ANÁLISE DA MATEMÁTICA APLICADA E PROBAS MATEMÁTICAS)

Navia Rivas de Castro

ESCOLA DE DOUTORAMENTO INTERNACIONAL EN ARTES E HUMANIDADES, CIENCIAS SOCIAIS E XURÍDICAS DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO EN LÓXICA E FILOSOFÍA DA CIENCIA

SANTIAGO DE COMPOSTELA

2020

DECLARACIÓN DO AUTOR/A DA TESE

[MATHEMATICAL EXPLANATION AND ONTOLOGY: AN ANALYSIS OF APPLIED MATHEMATICS AND MATHEMATICAL PROOFS]

D./Dna. Navia Rivas de Castro

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) De selo caso, na tese faise referencia ás colaboracións que tivo este traballo.
3) A tese é a versión definitiva presentada para a súa defensa e coincide coa versión enviada en formato electrónico.
4) Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.

En Santiago de Compostela, 14 de setembro de 2020.

Asdo.. Navia Rivas

de Castro

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DAS CODIRECTORAS DA TESE

[MATHEMATICAL EXPLANATION AND ONTOLOGY: AN ANALYSIS OF APPLIED MATHEMATICS AND MATHEMATICAL PROOFS]

Dna. Concepción Martínez Vidal
Dna. Mary Leng

INFORMAN / REPORT:

Que a presente tese, correspóndese co traballo realizado por D/Dna. ***Navia Rivas de Castro****, baixo a miña dirección, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director desta non incorre nas causas de abstención establecidas na Lei 40/2015. / That the present thesis corresponds to the work done by Miss* ***Navia Rivas de Castro****, under our direction, and We authorize its presentation, considering that it meets the requirements required in the Regulations of Doctoral Studies of the USC, and that as its director does not incur the causes of abstention established in Law 40/2015.*

En Santiago de Compostela, 14 de setembro de 2020,

En York (Reino Unido), 20 de setembro de 2020.

Asdo. Concepción
Martínez Vidal

Asdo. Mary Leng

Summaries

Resumo

A presente tese ten como obxectivo dar conta da explicación matemática en dúas áreas diferenciadas: a explicación científica (matemática aplicada) e a explicación matemática en sentido interno. A investigación abrangue dúas perspectivas diferentes: a que xorde da preocupación ontolóxica sobre as entidades matemáticas, e a que xorde do interese polo papel explicativo da matemática, centrado na práctica matemática e incluíndo un estudo de casos para o caso da matemática interna. Neste contexto, analízase a posibilidade de que haxa unha teoría unificada da explicación para dar conta de tódolos casos de explicación matemática.

A investigación ten lugar desde dúas perspectivas diferentes: unha que xorde da preocupación ontolóxica sobre as entidades matemáticas, e outra que xorde dunha escolla metodolóxica: estudar as cuestións sinaladas (a explicación matemática en ciencia e na propia matemática) na práctica matemática, é dicir, examinando a forma en que os matemáticos entenden o seu traballo nestas áreas, que inclúe un estudo de casos para o contexto da explicación intramatemática.

O obxectivo central é a análise do papel que ten a explicación matemática en ciencia e a súa relevancia para o éxito ou fracaso das teorías científicas. Isto trae consigo unha análise da cuestión ontolóxica de se o papel explicativo dos obxectos abstractos e, en concreto, dos obxectos matemáticos, é suficiente para postular a súa existencia. Ademais, considerarase a posibilidade de que haxa unha teoría unificada para dar conta da explicación matemática tanto interna como externa.

Para profundar nestas cuestións, o presente traballo inclúe (1) unha análise de como a explicación matemática foi tratada na literatura, unha análise do papel da proba matemática e das razóns polas que ten sentido buscar probas máis explicativas de resultados xa aceptados, (2) un estudo sobre a relación entre o uso das matemáticas na explicación científica e o compromiso ontolóxico das teorías como ferramentas explicativas na ciencia.

Parte do traballo da tese consistirá nun exame do papel explicativo das matemáticas mediante o estudo de casos que reflicten dito papel. O estudo de casos é unha das principais fontes de datos do presente traballo, co fin de clarificar o papel que a matemática ten, entre outros recursos metodolóxicos.

Palabras clave: explicación matemática, ontoloxía matemática, contrafácticos, virtudes explicativas, proba matemática.

Abstract

The present work aims at providing an account of mathematical explanation in two different areas: scientific explanation and within mathematics. The research is addressed from two different perspectives: the one arising from an ontological concern about mathematical entities, and the other originating from a methodological choice: to study our chosen problems (mathematical explanation in science and in mathematics itself) in mathematical practice, that is to say, looking at the way mathematicians understand and perform their work in these diverse areas, including a case study for the context of intra-mathematical explanation.

The central target is the analysis of the role that mathematical explanation plays in science and its relevance to the success or failure of scientific theories. The ontological question of whether the explanatory role of abstract objects, mathematical objects in particular, is enough to postulate their existence will be one of the issues to be addressed. Moreover, the possibility of a unified theory of explanation which can accommodate both external and internal mathematical explanation will also be considered.

In order to go deeper into these issues, the research includes: (1) an analysis how the question of what is involved in internal mathematical explanation has been addressed in the literature, an analysis of the role of mathematical proof and the reasons why it makes sense to search for more explanatory proofs of already known results, and (2) an analysis of the relation between the use of mathematics in scientific explanation and the ontological commitment that arises from these explanatory tools in science.

Part of the present work consists of an analysis of the explanatory role of mathematics through the study of cases reflecting this role. Case studies is one of the main sources of data in order to clarify the role mathematical entities play, among other methodological resources.

Key words: mathematical explanation, mathematical ontology, counterfactuals, explanatory virtues, mathematical proof.

Resumen

La presente tesis tiene como objetivo tratar la explicación matemática en dos áreas diferentes: la explicación científica y la propia matemática. La investigación abarca dos perspectivas diferentes: una que surge de la preocupación ontológica sobre las entidades matemáticas, y otra que surge de una elección metodológica: estudiar las cuestiones señaladas (la explicación matemática en ciencia y en la propia matemática) en la práctica matemática, es decir, examinando la forma en que los profesionales matemáticos entienden su trabajo en estas áreas, que incluye un estudio de casos en el contexto de la explicación intramatemática.

El objetivo central es el análisis del papel que tiene la explicación matemática en ciencia y su relevancia para el éxito o fracaso de las teorías científicas. Esto conlleva un análisis de la cuestión ontológica de si el papel explicativo de los objetos abstractos y, en particular, los objetos matemáticos, es suficiente para postular su existencia. Además, se considerará la posibilidad de que exista una teoría unificada que dé cuenta de la explicación matemática tanto interna como externa.

Con el fin de profundizar en estas cuestiones, el proyecto incluye: (1) un análisis de como la explicación matemática ha sido tratada en la literatura, un análisis del papel de la prueba matemática y de los motivos por los que tiene sentido buscar pruebas más explicativas de resultados ya aceptados, y (2) aclarar la relación entre el uso de las matemáticas en la explicación científica y el compromiso ontológico de las teorías como herramientas explicativas en ciencia.

Parte del trabajo de la tesis consistirá en el examen del papel explicativo de las matemáticas mediante el estudio de casos que reflejan dicho papel. El estudio de casos es una de las principales fuentes de datos del presente trabajo, con el fin de aclarar el papel que las matemáticas juegan, entre otros recursos metodológicos.

Palabras clave: explicación matemática, ontología matemática, contrafácticos, virtudes explicativas, prueba matemática.

Resumo extenso

O ser humano fai uso de explicacións matemáticas a diario en todo tipo de actividades científicas ou cotiás, pero temos unha idea moi limitade de como esas explicacións funcionan. Non obstante, claramente preferimos certas explicacións concretas sobre outras. Isto pode deberse a varias situacións: que algunhas probas dean conta do fenómeno máis eficazmente, que sexan máis fáciles de captar por unha audiencia particular, que sexan máis simples ou precisas, que identifiquen a causa do fenómeno, que respondan máis preguntas "*que pasaría se...?*" ou que unifiquen diversos fenómenos baixo unha mesma explicación. Non hai dubida de que a matemática ten un papel central en explicacións de feitos do mundo. As explicacións científicas inclúen un prominente compoñente matemático, que está case omnipresente nas ciencias naturais (desde a mecánica cuántica ó uso de ecuacións diferenciais e estatísticas en bioloxía) e ciencias sociais, polo que o interese en como funcionan estas explicacións non é sorprendente.

A presente tese, con título *Mathematical Explanation and Ontology: an Analysis of Applied Mathematics and Mathematical Proofs* (*Explicación e ontoloxía mateática: unha análise da matemática aplicada e probas matemáticas*), centra os seus obxectivos fundamentalmente en dous: o primeiro ten a ver con dar conta da explicación matemática no ámbito da explicación científica (polo que estariamos a falar da matemática aplicada ás diversas ciencias que fan uso dela) e a explicación matemática en sentido interno. Trátase, por tanto, dunha análise da explicación matemática en dous ámbitos diferenciados. Do mesmo xeito, a investigación abrangue dúas perspectivas diferentes: a que xorde da preocupación ontolóxica sobre as entidades matemáticas, e a que xorde do interese polo papel explicativo da matemática. Este está centrado na práctica matemática e, por este motivo, inclúe un estudo de casos para o caso da matemática interna. Neste contexto, analízase a posibilidade de que haxa unha teoría unificada da explicación para dar conta de tódolos casos de explicación matemática, tomando a Teoría Contrafáctica da Explicación (TCE) como candidata para esta análise.

Algunhas das hipóteses deste traballo inclúen as seguintes: a matemática ten un rol crucial na explicación científica; a explicación matemática, como un todo, pode ser pensada en termos contrafácticos, de maneira que a teoría contrafáctica da explicación é unha boa candidata como marco de análise da explicación matemática en diversas áreas de investigación; que aceptar a matemática como xenuinamente explicativa non implica que o enfoque platonista con respecto ás entidades matemáticas sexa correcto; e que non tódalas probas matemáticas constitúen explicacións xenuínas dos resultados relevantes.

Tendo en conta o anteriormente exposto, a tese consta de cinco capítulos, que forman un todo no que se poden diferenciar claramente dous grandes bloques: un primeiro bloque centrado na matemática aplicada e o debate ontolóxico que xorde arredor dela, e un segundo bloque centrado na análise do concepto de proba e de probas matemáticas concretas para elucidar aspectos relevantes sobre a capacidade explicativa da matemática en explicacións que son puramente matemáticas. Dun xeito máis específico, a continuación segue un resumo dos contidos da disertación.

Os **Capítulos 1 e 2** constitúen unha discusión sobre o estado do debate que concirne a explicación extra-matemática, isto é, a aplicación da matemática á explicación científica. Isto inclúe unha avaliación da cuestión ontolóxica relativa ós obxectos matemáticos, proporcionando algunhas claves para entender mellor o debate e, ademais, algunhas ideas sobre como se debería abordar esta problemática. Son consideradas algunhas opcións en relación á natureza das entidades abstractas (máis particularmente, as entidades matemáticas), incluíndo enfoques platonistas, nominalistas e deflacionistas. De maneira especial, é abordada a cuestión da elucidación de se o papel explicativo das entidades abstractas é suficiente para argumentar a súa existencia, como defenden as posicións platonistas, ou se este papel explicativo pode ser tratado sen traer consigo un compromiso coa existencia das entidades relevantes.

Partindo do argumento da indispensabilidade de Quine-Putnam, hai dúas principais respostas no lado nominalista: o programa de Field e diversos enfoques *fáciles* (*easy-road*) á cuestión. Desde o bando platonista, xurdiu unha nova estratexia enfocada no papel explicativo substancial das matemáticas en explicacións extra-matemáticas particulares (tales coma o caso das cigarras), que deron lugar ó debate do *argumento da indispensabilidade mellorado* (*Enhanced Indispensability Argument* ou *EIA*). No EIA tómanse casos concretos de explicacións extra-matemáticas onde supostamente a matemática ten un poder explicativo distinguiblemente xenuíno que leva á idea de que debemos adquirir un compromiso ontolóxico coa existencia dos obxectos matemáticos (e, por extensión, dos obxectos matemáticos).

Nesta sección ponse o enfoque no debate entre Baker, no lado platonista, e Knowles e Saatsi no lado oposto, para concluír que o tipo de papel explicativo substancial mencionado non abonda para xustificar un compromiso coa existencia das entidades matemáticas. Polo tanto, este tipo de papel que a matemática xoga na explicación é compatible cun enfoque realista débil ou nominalista. É compatible defender que os obxectos matemáticos non existen nun sentido platonista forte coa idea de que teñen poder explicativo. Neste sentido, semella razoable renunciar a unha visión platonista forte (como o denominado platonismo *heavy-duty*) e decantarse por un platonismo máis débil ou mesmo un enfoque nominalista. Outra conclusión crucial a este respecto é a de que o debate *EIA* non abonda para decantar a cuestión ontolóxica sobre as entidades matemáticas.

Neste debate, son moitos os traballos analizados no debate candente que se pode observar na literatura sobre o *EIA*. Así, son tratados entre outros, os realizados por Baker (2005, 2009, 2017), Balaguer (1996, 1998), Baron (2016, 2019), Colyvan (2010, 2012), Field (1980, 1989), Knowles e Saatsi (2019), Lange (2013), Leng (2010, 2020), Maddy (1995, 1997) e Melia (2000, 2002), e algúns enfoques máis clásicos coma o de Hempel (1965) e Salmon (1989). Ademais, realízase unha presentación detallada da Teoría Contrafáctica da Explicación (TCE), xunto coa avaliación do debate ontolóxico a respecto das entidades matemáticas.

De acordo con Maddy, podemos crer unha teoría científica e permanecer nunha posición agnóstica ou instrumentalista sobre a existencia dos obxectos relevantes para esa teoría, e neste traballo arguméntase que unha perspectiva similar a respecto dos obxectos matemáticos tamén pode ser aceptable. Da análise, extráese que a decisión entre un realismo débil e un enfoque nominalista acaba sendo unha cuestión de elección ou interpretación, que probablemente dependerá en último termo de cuestións intuitivas e non argumentos fundamentados.

Outras visións analizadas, coma a de Baron, sosteñen que os feitos matemáticos conteñen información non descritiva sobre feitos do mundo. Polo tanto, debemos crer na existencia dos *feitos* matemáticos. Agora ben, isto non trae consigo un compromiso coa existencia dos *obxectos* matemáticos, así que esta visión tamén sería compatible tanto cun platonismo forte como cun platonismo débil. De ser este o caso, e tendo en conta que o *EIA* non proporciona unha distinción clara no plano ontolóxico, entón a visión de Baron tamén sería en último termo compatible co nominalismo. Isto abre espazo para que o enfoque sexa compatible con calquera marco ontolóxico (aínda que isto quede, certamente, lonxe dos propósitos do autor), polo que semella proporcionar máis evidencias de que non hai motivos decisivos para mantermos unha posición platonista forte.

No lado negativo destas conclusións atopamos que, no contexto do EIA, non temos forma de determinar cal é a visión correcta na ontoloxía matemática polo que, para quen busque un modo de defender unha visión radicalmente nominalista, este marco teórico non sería o adecuado. O mesmo sucedería para quen queira defender un platonismo forte.

O **Capítulo 3** contén unha análise do concepto de proba, onde son examinadas diversas perspectivas acerca do concepto de proba e a distinción entre as probas formais e informais, tomando a proba como a principal manifestación de explicación intra-matemática. Este estudo inclúe diversas visións de autores como Azzouni (2004, 2009), Hersh (1997), Lakatos (1998), Larvor (2012), Rav (1999) e Resnik (1998). Tómase a distinción de Brown (2012) entre dous significados do termo ‘explicación’ (un sentido *forte* e un sentido *débil* ou *epistémico*) para máis adiante aplicala ó estudo de casos. Por último, acrecéntase unha análise da cuestión de se hai un consenso á hora de avaliar as probas matemáticas, e tamén unha análise da matemática en contextos educativos e tamén algunhas cuestións de carácter estético no relativo ó valor estético das explicacións matemáticas.

O **Capítulo 4** consiste na presentación dos casos elixidos, que serán relevantes para o estudo de casos, como un dos puntos centrais do traballo. Os casos reflicten a práctica matemática no relativo ás probas matemáticas e as súas virtudes explicativas (ou a ausencia destas).

O estudo de probas concretas para establecer como diferentes probas teñen a capacidade para explicar contribúe á clarificación da noción (ou nocións) de explicación que entran en xogo na matemática. O estudo de casos na práctica matemática é usado para recompilar datos que logo serán usados como a base para examinar diferentes perspectivas sobre a explicación matemática. Isto contribúe ó estudo de (a) como as xeneralizacións e a xeneralidade son usadas para explicar, así como outras virtudes explicativas como a simplicidade e a prominencia cognitiva, (b) o papel que as explicacións matemáticas xogan en ciencia e na propia matemática e a súa relevancia, e (c) a análise de se hai unha teoría da explicación que poida acomodar tódalas clases de explicación matemática ou, de modo alternativo, as razóns polas que podemos aceptar diversas teorías ou diversos elementos de análise. Estes casos son as probas da completude da aritmética de Gödel e Henkin, o teorema das catro cores e unha *proba pictórica* da fórmula para os números triangulares, así como tamén unha pequena análise da indución matemática como método para probar resultados matemáticos e para explicalos.

Algúns destes casos son explicativos, algúns non explicativos e outros non poden ser considerados propiamente probas. Este estudo de casos pretende iluminar aspectos relevantes da explicación matemática, mais tamén servirán para testar a teoría contrafáctica e proporcionar ideas sobre que requisitos debe cumprir unha teoría unificada da explicación matemática. Esta cuestión enmárcase no debate sobre se é posible obter unha teoría unificada da explicación e como a explicación matemática interna e a externa funcionan, no máis xeral.

O **Capítulo 5** está dedicado por enteiro á cuestión de se pode haber unha teoría unificada da explicación matemática. A teoría contrafáctica da explicación (TCE) é analizada como a principal candidata para dar conta da explicación matemática. Knowles e Saatsi (2019) aplican a teoría á explicación extra-matemática, aínda que non exclúen outras aplicacións do marco teórico, polo que a porta queda aberta a adaptar a teoría contrafáctica para aplicala a explicacións matemáticas tanto internas como externas.

Co obxectivo de completar a tarefa de examinar a posibilidade de construír unha teoría unificada da explicación, volvemos ós casos concretos e aplicámoslles a teoría contrafáctica como marco de referencia, coa dirección de mostrar que o marco teórico pode dar conta do seu poder explicativo e ter éxito como teoría unificada da explicación matemática, dado que a súa aplicabilidade a explicacións extra-matemáticas xa foi analizada no Capítulo 2. Ademais, a teoría é aí complementada con elementos doutros marcos teóricos, como a concepción inferencial de Bueno e Colyvan (2011) e o test informacional de Baron (2017).

A idea dunha teoría unificada da explicación matemática é atractiva por diversas razóns filosóficas, das cales a central é que normalmente preferimos teorías e explicacións cun nivel alto de unificación, porque con elas obtemos entendemento sobre fenómenos diversos mediante a identificación de elementos comúns neles.

O estudo de casos permite analizar diversos aspectos da explicación matemática interna e, de xeito máis específico, a proba matemática. O caso da completude ilustra que, en ocasións, o feito de que teñamos unha proba dun teorema particular non frea a busca dos matemáticos de probas máis explicativas. O contraste entre a proba de Gödel e a de Henkin mostra que esta última é máis explicativa e presenta máis poder de unificación, xa que é aplicable a un espectro matemático máis amplo. O debate que concirne o teorema das catro cores puxo de manifesto que algunhas probas matemáticas resultan ser non explicativas. A proba pictórica abre o debate sobre se os elementos visuais como imaxes e diagramas poden constituír probas no sentido xenuíno, e o debate máis xeral sobre o uso de elementos visuais en probas. Tamén apunta ó feito de que os criterios para que algo constitúa unha proba non están claros e pode haber explicacións matemáticas con éxito que non poden constituír probas por si mesmas. Ademais, a análise da indución matemática como un método de proba en matemáticas introduce o debate sobre se esas probas poden ser consideradas explicativas, xa que moitos autores argumentan que non é o caso.

Neste estudo faise evidente que non tódalas probas matemáticas explican o fenómeno matemático en cuestión e que a explicación intra-matemática vai máis alá da proba. Algunhas probas poden ter resultados que van máis alá do de establecer a verdade dun teorema. Estes resultados poden incluír establecer relacións entre áreas diferenciadas mediante a unificación e a conexión das mesmas. Tamén contribúe á xa habitual posta en dúbida do concepto tradicional de proba, onde un concepto menos estrito (ou máis informal) entra en xogo. Da análise da proba matemática (no Capítulo 3, fundamentalmente, aínda que tamén presente nos Capítulos 4 e 5), extráese a conclusión de que a proba matemática vai máis alá da mera dedución de axiomas a conclusión. Neste concepto, os estándares son menos estritos e deixan espazo para a incorporación do pensamento diagramático, a informática, e outros métodos. Queda establecida a distinción ente as probas que están limitadas a establecer un feito matemático, e as que ademais explican ese feito matemático.

A partir da análise neste traballo, poden extraerse unha serie de resultados. Como xa foi sinalado, a TCE é examinada ó longo do traballo co fin de determinar se é unha teoría adecuada para dar conta da explicación matemática, tanto intra-matemática como extra-matemática. Despois da análise do papel da matemática nas súas diversas aplicacións, o marco teórico é adaptado para acomodar os casos de explicación intra-matemática, para comprobar se a TCE e un marco adecuado para ser estendido a todo tipo de explicación. De aí a relevancia de incluír un estudo de casos de probas matemáticas concretas e do seu poder explicativo, ou a súa carencia del.

Contar cunha teoría unificada da explicación pode considerarse vantaxoso en tanto proporciona unha visión máis completa da explicación matemática, que simplemente depender das virtudes explicativas para identificar o grao de poder explicativo que teñen os casos concretos. Nesta liña, perspectivas coma a TCE de Knowles e Saatsi semellan ser máis prometedores de éxito que a enumeración de virtudes explicativas ou outras teorías que teñen un enfoque moito máis estreito.

Neste sentido, dúas das maiores vantaxes deste marco teórico son o monismo e a *lixeireza* (*lightwheightness*) no plano ontolóxico, en tanto que esta visión non forza un enfoque platonista sobre a ontoloxía matemática. A TCE dá conta de como podemos pensar nas explicacións matemáticas en termos contrafácticos, e tamén das súas virtudes explicativas. Por outra banda, pode dar conta dos dous sentidos de 'explicación' que se distinguen no Capítulo 3, seguindo a caracterización de Brown: un sentido forte (máis obxectivo) ligado ó *porqué* do *explanandum*, onde explicar en termos da teoría contrafáctica acaba por ser particularmente útil, e un sentido máis débil da explicación, máis relacionado co entendemento e a facilidade coa que captamos conceptos e regularidades.

Aínda que, como se analiza no traballo, a TCE é unha boa candidata como teoría para dar conta da explicación matemática, pode ser mellorada. Nomeadamente, un resultado deste traballo é que a TCE pode beneficiarse dalgúns conceptos e ideas doutros enfoques, como a importancia que Bueno e Colyvan atribúen á selección e a interpretación dos modelos que usamos para explicar na súa concepción inferencial. Ademais do papel de proporcionar inferencias, estes autores apuntan a outros roles, como o da unificación (de teorías diferenciadas), auxiliar na obtención de novas predicións, proporcionar explicacións de fenómenos empíricos, etc. En particular, para o caso da explicación extra-matemática, esta perspectiva inclúe un papel salientable da escolla e a interpretación, así como a dependencia do contexto no sentido de que, cando aplicamos unha teoría matemática para explicar un fenómeno empírico, necesitamos elixir a estrutura ou *mapeo* (*mapping*) adecuado entre o punto de partida empírico e a matemática (cf. Bueno e Colyvan 2011).

Ademais, a perspectiva contrafáctica pode beneficiarse da contribución de Lange, de acordo coa que o factor relevante que aparece no punto de partida dunha proba é o que fai que a proba sexa explicativa. Esta idea, tal como se desenvolve no Capítulo 5, é compatible coa tese de que unha boa explicación establece patróns de dependencia contrafáctica entre o *explanans* e o *explanandum* mediante a identificación das variables relevantes, xa que os contrafácticos relevantes poden ligarse a un factor relevante na proba. Lange tamén apunta á simetría e a simplicidade como os principais exemplos deses aspectos, e tamén pon énfase na relevancia da unificación, do mesmo xeito que o fai a TCE.

Os contrafácticos son parte da nosa forma de razoar, e non resulta infrecuente razoar sobre situacións imposibles como parte do proceso de obtención de clarificación na área de coñecemento orixinal. Isto proporciona certas claves sobre que é o que ten o peso ou labor explicativo. Se os nosos contrafácticos relevantes para a explicación están na parte empírica da información, isto é, deixan o aparato matemático fixo, entón semella que non son as entidades matemáticas as que teñen o peso explicativo, senón que as estruturas matemáticas axúdannos a ver as relacións explicativas que xa estaban presentes nos feitos físicos.

Tamén obtemos claves sobre que fai que unhas probas matemáticas sexan máis explicativas que outras. Por exemplo a distinción ten que ver coa facilidade cognitiva que a proba teña, pero tamén coa capacidade que teñan as probas de responder preguntas ós nosos cuestionamentos en forma de *por que* a respecto dos fenómenos. Ademais, ser prominente cognitivamente garda relación co noso acceso a un feito matemático particular, isto é, a nosa capacidade de captar as *verdades* relevantes con certa facilidade. Aquí entran en xogo diversos factores, como coñecementos previos, o contexto, o equilibrio entre a simplicidade e o detalle da proba, etc. Por suposto, tamén ten que ver cun sentido máis *obxectivo* en que a matemática explica. Por exemplo, no aso da proba de Henkin para a completude, móstrase que o poder explicativo vai máis alá da prominencia cognitiva, en tanto que a proba tamén explica por que a lóxica de primeira orde é completa, isto é, explica nun sentido máis profundo.

No que ten a ver cos factores que fan que as explicacións sexan *boas explicacións* ou teñan éxito, atopamos que a inmensa maioría de enfoques á cuestión apuntan ás mesmas propiedades ou virtudes explicativas básicas que unha proba pode mostrar en diversos graos. Estes factores explicativos inclúen a simplicidade, a xeneralidade, a unificación e a prominencia cognitiva.

Outras temáticas abordadas, aínda que menores, teñen a ver cos diferentes estándares que os matemáticos semellan usar cando avalían probas. Colyvan (2018) examina diferentes probas do mesmo resultado para concluír que coexisten diferentes criterios e analiza diversas maneiras en que a matemática ten valor explicativo, o cal contribúe á idea de que se poden aceptar variadas virtudes explicativas en xogo na explicación, en diferentes graos. Ademais, do estudo da matemática desde un punto de vista relacionado coa educación, pode concluírse que a unificación de feitos matemáticos diversos baixo os mesmos patróns explicativos fai que sexan máis fáciles para o estudantado, así como salienta a importancia da dimensión do entendemento en contraste coa noción de rigor, habitualmente situada como a principal. Isto apoia a tese de que é preferible unha proba que é máis iluminadora que unha que simplemente é máis rigorosa. Ademais, analízanse diversas cuestións de natureza estética en relación coa nosa preferencia dunhas explicacións matemáticas sobre outras.

Cómpre salientar que hai motivos para considerar que a explicación matemática interna e externa teñen moitos aspectos en común. Non significa que sexan o mesmo fenómeno, mais hai un sentido en que tanto a explicación interna como a externa están unificadas baixo a mesma ferramenta explicativa, isto é, o aparato matemático. Ademais, explicación e entendemento poden considerarse algo *global* no sentido de que son parte da natureza humana. Hai, polo tanto, un sentido en que en que os feitos empíricos e matemáticos teñen elementos estruturais en común, xa que usamos a mesma estrutura matemática para explicalos, e esta estrutura pode dar conta de fenómenos diversos, o cal tamén é salientable a respecto do seu alto poder explicativo.

Por último, o enfoque elixido para dar conta da explicación matemática deberá ser o suficientemente amplo para acomodar os múltiples factores que poden facer que unha explicación sexa efectivamente explicativa, ou mesmo recoñecer que, posiblemente, o enfoque correcto á explicación matemática debe ser aberto, compatible coa idea de que varios marcos teóricos poden funcionar a hora de dar conta do papel explicativo da matemática. Se isto é correcto, entón a conclusión será a de aceptar algún tipo de pluralismo de cara á explicación matemática, o que resultaría nun escenario aberto no que, dependendo do caso concreto que esteamos analizando, un enfoque ou outro pode ser máis adecuado. Non semella problemático contarmos con diversas perspectivas que subliñen diferentes elementos de análise. Polo tanto, aínda que a TCE pode dar conta da explicación matemática tanto interna como externa e constituír unha boa teoría unificada, continúa habendo espazo para outros enfoques dentro do debate.

A disertación pretende proporcionar máis elementos de análise e unha clarificación dos diversos debates que teñen lugar arredor da explicación e ontoloxía matemática, así como dalgúns elementos clave na explicación como son a simplicidade e a unificación, mais queda moito traballo por facer no camiño da obtención dunha perspectiva satisfactoria para a explicación matemática, o cal é indicativo de que máis traballos nesta liña están aínda por vir.

Acknowledgements

Frist of all, many thanks to my PhD supervisors Concepción Martínez-Vidal and Mary Leng, for their i*nfinite and indispensable* patience, dedication and support.

To my family and friends for their unconditional kindness. Thank you.

To EPISTEME and all the professors and staff of the Faculty.

To everyone who asked questions and sent suggestions.

To my Logic students of the Faculty. *They* taught *me* my true vocation.

To my travel partners.

To María José Pérez Ferreira, for graphics.

To anyone who ever asked the question “How is that Thesis going?”, no matter how rude my response might have been. And to every “I’ll let you work”. Thank you.

This work was supported by FEDER/Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación/ Proyecto (FFI2017-82534-P) .

Contents

Introduction

There are seven species of cicadas that have life and reproductive cycles of 13 or 17 years, depending on the species. These are all species of the so-called ‘periodical magicicada’ genus. These insects remain during their nymphal stage in the soil for a long period of time, and then the adult cicada emerges every 13 or 17 years. In fact, this appearance is synchronized among the members of the brood. In a few weeks they reproduce, then die and the cycle continues (see Baker 2005, 229).

These life cycles can be considered evolutionary strategies to avoid predation. Apart from their durations, which will hardly overlap with those of their predators, each time they appear they do it in very large numbers so the predators will not be able to exhaust the species. This evolutionary strategy is known as “predator satiation” (see. Gould 1977, 101).

The natural question that arises here for everyone that encounters this case for the first time is: why prime number periods?

Apparently, it has something to do with the biological fact that it is an evolutionary advantage to minimize the intersection with predator species or even other periodical cicada species. The mathematical component supports this thesis and complements the biological claim, since prime periods actually minimize intersection. **Now, can a theorem of arithmetic explain a biological phenomenon?**

Bees build hexagonal honeycombs instead of any other polygonal figure or combination of them. Is it just a coincidence? That is very unlikely. Apparently, there are evolutionary facts that can explain this phenomenon. Bees that use less wax have a better chance at evolving via natural selection, because they spend less energy and economize their resources more efficiently. This explanation is completed by the geometrical fact that any partition of the plain into regions of equal area has perimeter at least that of the regular hexagonal honeycomb tiling. That is, the hexagon is the optimal shape with respect to dividing the plane into equal areas and minimizing the perimeter – thus, minimizing wax expenditure. This is what is known in the literature as the “honeycomb conjecture”, and was recently proved in Hales 2001. **Now, can a geometrical proof explain a biological fact? Or, moreover, can an explanation of a biological fact depend on a mathematical fact?**

Traditionally, the city of Königsberg was set on both sides of the Pregel River and had two islands connected to each other and the two mainland portions by seven bridges. There was at the time an unsolved problem about designing a walk through the city that would cross each of the seven bridges once and once only. The problem was solved with its now well-known negative answer by Leonhard Euler back in 1736, and led to the foundations of graph theory and prefigured the idea of topology**.**

Similarly, one can ask: why is it impossible to cross the seven Bridges of Königsberg crossing each bridge exactly once? The answer here is that the seven Bridges have the same structure as a graph that lacks an Eulerian Path. **Now, can graph theory actually explain a physical phenomenon?**

Examples such as the number theoretic explanation of cicada periods, the hexagonal honeycomb in geometry and the graph-theoretic explanation of the non-traversibility of Koenigsberg's bridges are commonly used in the literature on the so-called *Explanatory Indispensability Argument* (EIA). They are supposed (by platonists) to be particular cases where mathematics has a distinctively genuine explanatory power which shows or leads to the idea that we should be committed to the existence of mathematical objects and, more generally, abstract objects. Given that mathematics seems to have a central place in contemporary science, all these seem to be cases where mathematics is directly helping the scientist.

Indeed, it is undeniable that mathematics plays a relevant role in the three explanations described above, and also that it plays a huge role in all kinds of scientific theories and scientific explanations, up to a point where it is difficult even to imagine science or ordinary explanations of empirical phenomena without directly using mathematics. Moreover, in physics it can be hard to distinguish between the mathematical and the physical components of theories and explanations, due to the highly mathematized nature of this subject. Also, mathematics contributes to science from different areas, which becomes clear once we point out that the honeycomb case appeals to a geometrical theorem, whereas the cicada case appeals to an arithmetical theorem, and so on. Therefore, apparently mathematics can contribute to our ordinary and scientific knowledge in various ways.

There is no doubt that these cases pose very interesting questions regarding the applicability of mathematics to empirical sciences or empirical phenomena, in general, and also regarding the explanatory role of mathematics. The issue of how mathematics explains and how we use this tool to answer questions about the world constitutes a deep question joining our curiosity about what is outside our minds and our capacity to explain and understand, in connection with the resources we use for that task.

The fact is that we use (mathematical) explanations on a daily basis, but still we have little idea of how those explanations really work. However, it is clear that we prefer some particular explanations to others. This could be due to the fact that some explanations account for the cause of the phenomenon better, they are easier to grasp by a particular audience, they are simpler or more precise, they identify the cause of the phenomenon, or because they unify phenomena (in a way that diverse facts are explained by the same explanation). Or, for instance, an explanation can be better in providing more what-if questions about some issue.

Clearly, mathematics plays a central role in explanations of facts of the world. Scientific explanations including a prominent mathematical component are almost omnipresent in natural and social sciences. Once that is established, it is no surprise that this role mathematics plays in sciences different from mathematics itself interests us, from the viewpoint of wanting to understand the contribution of mathematics to explanations of empirical facts. Besides, how we apply mathematics to the different sciences is a rich field of research since it appears that mathematics and empirical science are of different natures and, still, large portions of mathematics are present in every area of scientific research. These applications can be found from economics to quantum mechanics and, just to provide an example, the use of difference equations and statistics is very common in biology.

The examples presented at the outline show that there are some *distinctively* mathematical components in the core of the explanations behind the cicada case, the Königsberg bridges and the hexagonal honeycombs. This relevance of mathematics might be just an aid in our getting simpler and more elegant theories, or it could be a more substantive (even indispensable) explanatory role of empirical phenomena.

This issue is interesting on its own. However, as has been pointed out (Baker 2006, 2009; Colyvan 2010, 2012), it is connected to a broader philosophical issue of great relevance: the ontological debate concerning abstract objects. In fact, it may be fair to say that the central debate in philosophy of mathematics is currently the issue of the connection between mathematical explanation and mathematical ontology. The question of whether mathematical entities exist or not is behind most of the literature concerning mathematics and mathematical explanation in, at least, the last three decades, especially as far as the so called EIA debate is concerned. Moreover, the subject can be considered part of the general debate concerning mathematical explanation, and it is behind most studies (from both philosophers and mathematicians) about what it means for mathematics to have an explanatory role.

That is the (certainly broader) orientation of the present work. Mathematics produces explanations and provides us with understanding of both empirical phenomena and mathematical facts in an internal sense. This could even be its fundamental role. So, our main aim is to analyse mathematics' explanatory role in general.

With this in mind, the present PhD dissertation addresses the issue of explanation and understanding in mathematics, by focusing on two main aspects (or kinds of explanations): the first one, concerning extra-mathematical explanation (Chapters 1-2) and the second one, concerning intra-mathematical explanation (Chapters 3-5). The first one concerns the application of mathematics to the diverse sciences in order to obtain explanation of empirical phenomena. The second one is the use of mathematical explanation within mathematics, in which mathematical proof has a predominant role.

One of the central issues is the concept of explanation itself, where several features of the concept are addressed, such as the distinction between two senses of explanation (explanation in a *weak* and in a *strong* sense), the study of several explanatory virtues or features, and the assessment of different accounts of mathematical explanation and their success. All these accounts of explanation have in common the aim to account for how explanation and understanding work for the case of mathematics.

Mathematical applications and the indispensability of mathematics occupy a central position in the analysis, specially focused on the classical Quine-Putnam IA and the Enhanced Indispensability Argument (EIA), on one side, and the nominalist responses to it, such as Field's program and *easy-road nominalist approaches*. There is an analysis of the switch from representational indispensability to explanatory indispensability as the platonists' strategy to argue for a realist approach to mathematical ontology. Many accounts are addressed, such as the ones put forward by Baker (2005, 2009, 2017), Balaguer (1996, 1998), Baron (2016, 2019), Colyvan (2010, 2012), Field (1980, 1989), Knowles and Saatsi (2019), Lange (2013), Leng (2010, 2020), Maddy (1995, 1997) and Melia (2000, 2002), and some more classical approaches, such as Hempel's (1965) and Salmon's (1989), among others. In addition, a detailed presentation of the Counterfactual Theory of Explanation (CTE) is provided and an assessment of the ontological debate is addressed.

The study of intra-mathematical explanation includes the presentation and analysis of different cases of mathematical proofs. Indeed, proofs are probably the main kind of intra-mathematical explanation. With this in mind, a significant part of the dissertation is dedicated to the concept of proof, as the central means of intra-mathematical explanations, providing a classification of mathematical proofs and trying to elucidate what makes (and what does not make) a proof explanatory. Besides, the explanatory power of proofs by mathematical induction, which constitutes a philosophical debate in its own right, is also addressed in this work.

This study includes the classical accounts of intra-mathematical explanation developed by Steiner (1978) and Kitcher (1981), both of them top-down approaches (from the general model to case studies) as well as models of extra-mathematical explanation, such as the deductive-nomological model or the CTE.

Some other, perhaps secondary, issues are addressed. For instance, there are sections devoted to the role of mathematics of making proofs easier for us to follow in educational contexts, or some aesthetic concerns regarding mathematical proofs. As for mathematics and education, it has been pointed out that mathematicians might find a certain proof convincing, but they might be unsatisfied with regard to its explanatoriness, which might lead them to pursue new *easier* or *more illuminating* proofs of an already established result. This is linked to the explanatory features of particular mathematical explanations, and it raises the question of whether internal mathematical explanations are considered explanatory because they are *user-friendly* (that is, they make things easier to grasp) or because they get at the fundamental reasons why the mathematical result is the case (which connects with the strong meaning of ‘explanation’).

For our study of the explanatory role of mathematical proof, this work provides an analysis which the methodology of case studies is employed. For this purpose, several cases are presented: the completeness theorem (in the versions of Gödel and Henkin), the Four-Colour Theorem and an example of a *picture proof* of the formula for triangular numbers, followed by a brief analysis of mathematical induction. Some of these cases are explanatory, some non-explanatory and some perhaps do not constitute proofs. This case study includes relevant examples of mathematical proofs and explanations that might help us illuminate relevant aspects of their explanatory role and the concept of proof, and they are used to test the CTE and provide ideas as to what an account of mathematical explanation needs to be. This is immersed in the debate about whether a unified theory of mathematical explanation is possible and how internal mathematical explanations work, in general. Besides, this debate will help us see some relevant features concerning the educational and pedagogical sides of mathematical proof and mathematical understanding.

The idea of a unified theory of mathematical explanation is appealing for several philosophical reasons, the central one being that we usually prefer accounts and explanations that present a higher level of unification, because with them we obtain clarification regarding diverse phenomena by identifying common principles in them. A similar idea is present in the interest of many philosophers to investigate the possibility to construct a model of explanation that accommodates all areas of scientific research. This is arguably a sign that it is possible that internal and external mathematical explanations are two occurrences of a similar phenomenon, since explaining is present in all areas of research and our ordinary lives, and so is mathematics, which already provides some unity to our ability to explain as a part of human nature. Therefore, this is one of the main questions addressed in the present work, where the Counterfactual Theory of Explanation is examined as a candidate to be that unified account of mathematical explanation. Here, the task of the case study is completed, by accounting for their explanatoriness from the perspective of the counterfactual framework.

Methodology and aims

As for the methodology of research, the broad guidelines include a great deal of revision of bibliography that is related to the central issues. The result is that the study provides a quite detailed revision of how the most relevant issues concerning mathematical explanation and mathematical ontology were addressed in the literature. In the process of the research project, this was completed with frequent meetings with both advisors, attendance to research seminary sessions of the research group and workshops organized in the Faculty of Philosophy, as well as events in other Universities or Research Centres. This was accompanied by the production of presentations and papers related to this project, in the different stages of the research, the content of which was included as part as the present PhD Thesis. Different philosophical questions are addressed using the traditional methodology of analytical philosophy, which includes conceptual analysis, contrast of viewpoints, the use of formal languages and analysis of arguments, etc. It is worth mentioning that a significant part of this thesis is a case study concerning internal mathematical proofs and explanations with the aim of examining their explanatory power and the different explanatory features they present in a detailed and comprehensive way.

As has been pointed out, the main aim of the present work is to provide an analysis of the role that mathematical explanations – and mathematical entities as abstract entities – play in explanation, both in scientific explanation and within mathematical explanation. In order to test whether the CTE is appropriate as a unified theory of internal and external mathematical explanations, the case study methodology is employed, by analysing various relevant cases.

Some of the hypotheses in this work include the following: mathematics must have a crucial role in scientific explanation; mathematical explanation, as a whole, can be thought of in counterfactual terms, so the counterfactual theory of explanation is a good candidate to account for mathematical explanation in several areas of research; that accepting mathematics as genuine explanatory does not imply that the platonistic approach to mathematical entities is correct; and that not all mathematical proofs constitute explanation of the relevant results.

More specifically, here is an outline of the contents in this dissertation:

Chapters 1 and 2 constitute a discussion of the state of the debate concerning extra-mathematical explanation, that is, the application of mathematics in scientific explanation. This includes an assessment of the ontological question concerning mathematical objects, providing some clues to understand the debate better and also some ideas of how it should be approached. Several options are considered in relation to the nature of abstract entities (more in particular, mathematical entities), including platonistic, nominalistic and deflationist approaches. The issue of elucidating whether the explanatory role of abstract objects is enough to argue for their existence, as platonists claim, or if this explanatory role can be addressed without carrying a commitment to the existence relevant entities is addressed.

Chapter 4 contains an analysis of the concept of proof, where several approaches to the concept of proof and the distinction between formal and informal proofs is examined. This includes diverse views of authors such as Azzouni (2004, 2009), Hersh (1997), Lakatos (1998), Larvor (2012), Rav (1999) and Resnik (1998). We take Brown's (2012) distinction between two meanings of explanation (a strong sense and a epistemic sense) in order to later on apply it to the case study, and there is a brief analysis of whether there is consensus in evaluating proofs, mathematics in educational contexts and aesthetical concerns.

Chapter 3 consists of the presentation of the cases from our case study. These cases reflect the practice of mathematicians dealing with proofs and their explanatory features (or lack thereof). The analysis of particular proofs in order to establish how the different proofs explain contributes to a clarification of the notion (or notions) of explanation in play in mathematics. The case study in mathematical practice is used to collect data that is later on used as the basis for examining accounts of mathematical explanation. This contributes to the study of (a) how generalizations and generality are used to explain, as well as other explanatory features such as simplicity and cognitive salience and (b), the role that mathematical explanations play in science and within mathematics and their relevance, the analysis of a theory of explanation that can accommodate all kinds of mathematical explanation or, alternatively the reasons why we can accept several theories or diverse elements of analysis. The cases are Gödel's and Henkin's completeness theorem, the Four-Colour Theorem and a *picture proof* of the formula for triangular numbers, and a brief analysis of mathematical induction

Chapter 5 is devoted to the question of whether there can be a unified theory of mathematical explanation. The Counterfactual Theory of Explanation (CTE) is analysed as the main candidate to account for mathematical explanation. Knowles and Saatsi (2019) applied the theory to extra-mathematical explanation, but they do not exclude other applications of the theory, so the door is open to adapt the theory and apply it to both internal and external mathematical explanations. In order to complete the task of examining the possibility to construct a unified theory of explanation, we go back to the cases and applied the CTE to them, with the aim to show that the framework can account for their explanatoriness and succeed as a unified theory of mathematical explanation, given that its applicability to extra-mathematical explanations had already been analysed in Chapter 2. Besides, the theory will be complemented with features from other frameworks, such as the inferential account (Bueno and Colyvan 2011) and Baron's (2017) informational test.

CHAPTER 1:
Indispensability and mathematical explanation

1. Introduction: theories of scientific explanation

In philosophy of mathematics, there is a remarkable interest in issues concerning mathematical explanation, in particular, the search for a suitable account for mathematical explanations of scientific or empirical phenomena. Consequently, any study about mathematical explanation has to address how mathematics is applied to empirical sciences. This leads us to one of the fundamental aims of the present chapter, which is to review the state of the question regarding the search for a suitable account for mathematical explanations of scientific or empirical phenomena. Even though our primary aim is to provide a suitable account of the explanatory role of mathematics, inevitably, the ontological debate will be a constant in the presentation of explanation for the case of mathematics.

In the literature, there has been a predominance of causal accounts of scientific explanation, according to which the sciences explain by identifying the causes of and mechanisms for the phenomenon that is to be explained.

This situation has changed since the mid-2000s, when a significant number of philosophers of science have argued that there are explanations in the sciences that did not fit into the concept of causal explanation. There are numerous examples of scientific explanations whose explanatory power does not derive from the identification of causes and mechanisms. The conclusion is that there are non-causal explanations.

This has led to two possible attitudes. The first one involves denying the non-causal character of those explanations (strong attitude). The second involves considering that there is a need for an account of non-causal explanations complementing an account of causal explanations.

Taking Reutlinger’s (2006) analysis, among the various examples of non-causal explanations there is a special kind that is “distinctively mathematical”, and these examples have been used in the debate around the existence of mathematical objects, and abstract objects more generally. They include precisely the examples cited at the outline, that is, the cicada case, the Königsberg bridges, the honeycomb cells and so on.

The list of examples, as Reutlinger (2018, 74-75) points out, is taken to include different kinds of “purely” or “distinctively” mathematical explanations – such as number-theoretical (Baker, 2009a), graph-theoretic (Lange, 2013a; Pincock, 2012), topological (Huneman, 2010; Lange, 2013a), geometric explanations (Lange, 2013a), abstract explanations (Pincock, 2012, 2015), structural explanations (Bokulich, 2008), and statistical explanations (Lange, 2013b; Lipton, 2004). Other kinds of non-causal explanations in physics are explanations based on symmetry principles and conservation laws (Lange 2011, 2013a, 2017; French and Saatsi 2018), kinematic principles (Saatsi, 2016), renormalization group theory (Reutlinger, 2014, 2016; Batterman, 2000; Morrison, 2018), dimensional analysis (Lange, 2009a; Pexton, 2014), laws of coexistence (Kistler, 2013), structural explanations in special relativity (Lange, 2013c), variational principles, laws of composition (Lange, 2009b), and inter-theoretic relations (Batterman, 2002; Bokulich, 2008). Furthermore, the recent debate identifies examples of non-causal explanations in the special sciences, such as in neuroscience (Chirimuuta, 2014, 2017, 2018), and in the sciences of complex systems (Morrison, 2018).

Facing this situation, we can adopt (again, following Reutlinger 2018) three different strategies or approaches:

- **Causal reductionism.** This is the thesis that there are no non-causal explanations. Therefore, the alleged examples of non-causal explanations can ultimately be understood as causal explanations.

Lewis (1986) and Skow (2014) have presented attempts to provide an account with this strategy by weakening the causal account with the idea that explanations do not necessarily need to identify the cause but only provide some information about the causal history of the explanandum. Therefore, allegedly non-causal explanations turn out to be causal explanations if we adopt this weakened account of causal explanation (see Reutlinger 2018, 75-76).

- **Pluralism.** Within a pluralist approach, causal and non-causal explanations are covered by two (or more) different theories of explanation.

Reutlinger points to two different pluralist approaches to mathematical explanation. The first one is Salmon's claim about the peaceful coexistence of the ontic causal account and the epistemic unification account. According to this account, phenomena can have two kinds of explanation, causal bottom-up explanations and unificationist top-down explanations (Salmon 1989), and there is no single theory that can capture both types of explanation. Second, Reutlinger points to Lange's approach (Lange 2011, 2013a, 2016). Lange defends a modal account, according to which many non-causal explanations operate by showing what constrains the explanandum phenomenon, where constraining means showing why the explanandum had to occur. He argues that, ultimately, distinctive mathematical explanation in science appeal to facts that are modally stronger than ordinary causal laws (see Lange 2013a, 491.). So, the ontic causal account applies to causal explanation, whereas this modal account applies (at least) to distinctly mathematical explanations, hence, the pluralist approach.

- **Monism.** There is one single philosophical account capturing both causal and non-causal explanations, since both kinds of explanation share a feature that makes them explanatory.

The classical monist account was put forward by Hempel (1965), according to which causal and non-causal explanations are explanatory by virtue of having one single feature in common: nomic expectability. This theory is known as the covering-law account or the Deductive-Nomological Model. This account is more detailed in section 5.1 of the present chapter.

Another monist account, which is currently one of the most popular ones, is the counterfactual theory of explanation, to which we will dedicate a detailed analysis in the chapters to follow. In fact, the counterfactual theory of explanation will be addressed as a good candidate to account for all kinds of mathematical explanation (including its applications to science but also explanation within mathematics) and, hopefully, one that will be ontologically neutral or, at least, not committing to a strong platonist ontological position.

According to Reutlinger, there are two main reasons to prefer a monist account: (a) the fact that there are compelling examples of non-causal explanations in science, so monism would be superior to causal reductionism in the sense that it does not imply denying the existence of non-causal explanations, and (b) general philosophical theories are often preferred to less general theories, which would make monism in principle more attractive than pluralism since monism offers one general theory for both causal and non-causal explanations.

Apart from these approaches, we should note that there is a coexistent tendency (as we will see in Baker 2005 and others) to analyse explanatory virtues of mathematical explanations without necessarily making them fit into a particular general theory of explanation.

- **The plan**

The fundamental aim of the present chapter is to provide a review on the history of the debate around the alleged indispensability of mathematics and how we got to the debate on mathematical explanation.

For this, we will first examine the original indispensability argument (Section 2) for the mathematical case, focusing on the concept of indispensability itself, to then review some of the most relevant objections to it.

Section 3 will be devoted to the Easy-Road Nominalism as a conglomerate of different nominalist (or, at least anti-platonist) views, which all have in common that they oppose the indispensability argument.

In Section 4, the *Enhanced Indispensability Argument* (EIA) will be introduced, in order to justify the need to evaluate mathematical explanation, since the EIA appears to be a sort of improvement of the classical indispensability argument by making use of the idea that mathematics' role in the sciences cannot be just that of describing phenomena and thus there is a deeper sense of mathematical involvement of some sort of *genuine* or *distinctive explanatory role*.

Since the study of how mathematics explains becomes an important task, Section 5 will address two classical and well-known theories of explanation: the Deductive-Nomological Model and the Causal Mechanistic Model.

With the same idea in mind, in Section 6 we will find a presentation of some other more recent approaches to mathematical explanation, from the viewpoint that there are two different perspectives in this task: we can either (a) provide a general theory of explanation that addresses the role of mathematics in explanations, or (b) just list and evaluate mathematics in the light of explanatory virtues that mathematical explanations show in variable degrees.

The project of analysing the role of mathematics in explaining empirical phenomena will be completed in Chapter 2, where the Enhanced Indispensability Argument debate will be addressed more thoroughly, in order to clarify both the concept of mathematical explanation and the ontological implications it carries.

2. Indispensability arguments

As pointed out above, one of the most interesting philosophical topics regarding mathematics is its applicability to empirical science since, apparently, mathematics and empirical science have very different natures and, however, every branch of natural science uses or contains large and diverse portions of mathematics. These range from the use of basic geometry almost everywhere to more sophisticated mathematical results such as Hilbert spaces in quantum mechanics.

The use of mathematics in physics or statistics may not surprise us at all. Still, big portions of mathematics are immersed also in other empirical sciences such as biology, where the use of difference equations and statistics is very common, as is in other empirical and also social sciences.

The relevance of mathematics can be that it helps making empirical predictions or just an *economical* use, allowing elegant statements and simplifications of theories. Or, perhaps, it could have a more important, even indispensable, role in explaining what is going on in the world. There has even been a tendency to see predictions and explanations to be two sides of the same coin.

The widespread use of mathematics in science makes it hard sometimes to imagine advanced theories such as quantum mechanics or general relativity without the role mathematics plays in them, leading us to think that they would in fact have been impossible without mathematics.

This use of mathematics in science might make us reflect on the question of which is the correct metaphysical picture for the case of mathematics (and abstract entities, more in general). From this, we might think that the fact that mathematics has a relevant role in predictions and explanations in science forces a platonist view on mathematical objects. The current debate seems to take us to a position where the focus has turned to explanation, and questions regarding scientific predictions containing a substantial mathematical involvement do not appear in the ontological debate. However, it seems reasonable to think that the requirement of a platonist approach works in the same way for predictions and explanations, given their closeness.

Among the main aims of the present work is that of clarifying the ontological question regarding mathematical entities as much as possible. As we shall see, there might not be a clear answer as to whether mathematical entities exist. However, some conclusions will include the thesis that a platonistic view on abstract objects is not necessary to account for the explanatory role of mathematical predictions and explanations.

Much of the discussion in the debate between nominalism and platonism in philosophy of mathematics has focused on indispensability arguments, both the original Quine-Putnam Indispensability Argument and the now well-known *Enhanced Indispensability Argument* by Baker (2005, 2009a, 2017), Colyvan (1998, 2002) and other authors. The debate leads us to a series of Indispensability Arguments in philosophy of mathematics.

- What is an indispensability argument?

An indispensability argument is a special kind of inference to the best explanation that attempts to establish the truth of a claim based on the indispensability of such claim for certain purposes. In the case addressed here, the indispensability arguments have been seen as attempts to justify our knowledge of an abstract mathematical ontology by using a strictly empiricist epistemology.

There has been an enormous amount of literature on the discussion concerning the indispensability of mathematics. Applied to mathematics, an indispensability argument is an argument that aims to establish the truth of mathematics based on its indispensability in empirical sciences for some important purpose. For instance, when focused on the explanatory role of mathematics, as we shall see in the context of the EIA debate, indispensability arguments require indispensability for *explanatory* purposes. In contrast, in the original indispensability argument, representational indispensability was considered sufficient.

The classical version of the Indispensability Argument (IA) for the mathematical case is attributed to Quine and Putnam. It has been seen by some as the best argument for mathematical platonism or realism. Many platonists rely on this argument to justify their belief in the existence of mathematical entities.

As a consequence of the appearance of the IA, mathematical nominalists who wish to be realists about scientific entities (quarks, electrons, black holes…) face a difficulty, since the IA is based on the same grounds that justify scientific realism. This perspective (mathematical nominalism and scientific realism) is what Quine (1980, 45) calls holding a “double standard” with regard to ontology, and it has led to the attempt of many nominalists to identify where this argument goes wrong. Quine has never explicitly stated an Indispensability Argument as such, but he alludes to it in several places (1939, 1948, 1955, 1958, 1960, 1986).

Contrary to Quine, Putnam (1971, 1975) has provided some more direct discussion of the indispensability argument[1]. Since his presentation is the standard source for the reconstructed versions of the argument, such as the one displayed below, let us briefly reference the original one. In Putnam’s words, it is “an argument for realism along roughly the following lines: quantification over mathematical entities is indispensable for science, both formal and physical; therefore we should accept such quantification; but this commits us to accepting the existence of the mathematical entities in question” (Putnam 1975, 57). When discussing the indispensability argument, Putnam himself references Quine’s interest in the indispensability of quantification over mathematical entities and the “intellectual dishonesty” (or *double standard*, as was mentioned above) of denying the existence of those entities we presuppose in our scientific theories (see Putnam 1975, 57). He also discusses and rejects some reasons why philosophers might deny the validity of indispensability arguments.

In order for us to get a clearer version of the Indispensability Argument, a reconstruction of the Quine-Putnam Indispensability Argument could go as follows[2]:

(P1) We ought to have ontological commitment to all and only the entities that are indispensable to our best scientific theories.

(P2) Mathematical entities are indispensable to our best scientific theories.

---

[1] Note that in “Indispensability arguments in mathematics”, Putnam (2012) rejects having endorsed the indispensability argument attributed to Quine and Putnam. He argues that mathematics’ contribution to science does not need to be interpreted in platonistic terms and it can have a modal interpretation. In fact, he states that his position was, rather than platonism, some sort of semantic realism. For a discussion on Putnam’s view on the IA and this apparent shift of position, see: **Martínez-Vidal, C. (2018),** “Putnam and contemporary fictionalism”, *THEORIA. An International Journal for Theory, History and Foundations of Science*, *33*(2), 165-181.

[2] This reconstruction can be found in: Colyvan, Mark, "Indispensability Arguments in the Philosophy of Mathematics", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/mathphil-indis/>. (Accessed on the 1st of August, 2019.)

(C) We ought to have ontological commitment to mathematical entities.

Let us now clarify and discuss some of the relevant features of the argument.

- **The concept of indispensability**

The first premise (P1) of the so-called “Quine-Putnam Indispensability Argument” is supported by the doctrines of naturalism and holism. Let us briefly review their main ideas.

According to **naturalism**, there is no first philosophy and philosophical research is continuous with the scientific enterprise (Quine 1981). Science is taken to be the complete story of the world, and this idea is loaded with a deep respect for scientific methodology and its success in answering fundamental questions about the nature of things. From a naturalist approach, there is no need for any supra-scientific authority or any justification beyond observation and the scientific method.

From a metaphysical perspective, therefore, we should look at our best scientific theories to determine what we ought to believe exists. Still, it remains debatable as to whether naturalism justifies the belief in *all* the entities of the best scientific theories.

Confirmational holism is the thesis that theories are confirmed or disconfirmed as wholes (see Quine 1951): “The dogma of reductionism survives in the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all. My countersuggestion, issuing essentially from Carnap's doctrine of the physical world in the *Aufbau*, is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body” (Quine 1951, 38).

This thesis entails that our mathematical theories and our scientific theories are linked, and that our justifications for believing in science and mathematics are not independent. When new evidence conflicts with our current scientific theory, we can choose to adjust either scientific principles or mathematical ones. This also implies that the same evidence justifies the mathematical and the empirical portions of the theory.

Quine’s epistemology allows for beliefs in mathematical objects despite their abstractness, since it rejects positivism’s requirement for reductions of scientific claims to sense data, so we do not need sensory experience of mathematical entities in order to justify our mathematical beliefs. It suffices to show that mathematical objects are indispensable to our best theory, which gives us an idea of the strength of the claim put forward by the indispensability argument. We simply extend our belief in a theory to the objects which it posits.[3]

- **Objections to the Indispensability Argument**

There have been multiple objections to the classical Indispensability Argument. Let us stop to see a few of them.[4]

- **Hartry Field and the Hard-Road Nominalism**

---

[3] Quine also defends a semantic holism: the unit of meaning is not the single sentence, but systems of sentences (in extreme cases the whole of language). This thesis is highly controversial. Most commentators argue that confirmational holism is enough for the IA. See, for instance, Colyvan (1998); Field (1989, 14–20); Maddy (1992).

[4] Most of these objections can be found in Colyvan, Mark, "Indispensability Arguments in the Philosophy of Mathematics", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/mathphil-indis/>. (Accessed on the 1st of August, 2019)

Perhaps the first objection to the IA that comes to mind is Hartry Field's nominalist program, which is considered the "hard-road" or the most radical version of the nominalist approach to the ontological question regarding abstract objects.

Field (1980, 2016), as is well-known, denies the second premise of the IA by suggesting that *mathematics is not indispensable* to science.

He not only argues that mathematical sentences cannot be considered literally true, but goes on to state that all mathematical discourse must be thought of as false.

Field's radical posture is salient in the thesis that, since we have no reason to believe that mathematical objects exist, a more moderate *agnostic* view regarding the truth value of the mathematical discourse would also make no sense (see Field 1989).

Therefore, the right question to ask would be about how the mathematical discourse is applicable to the physical world. His thesis is that, in order to explain these applications to the physical word, there is no need to take mathematical statements as true, and considering it *little more* than being consistent would suffice (see Field 1980, 1-7).

According to Field (1980, 1989), mathematics' role is that of facilitating inferences from some assertions to other assertions in empirical sciences. Departing from acceptable assertions about empirical data, mathematics would work as a vehicle to help making predictions, so the mathematical discourse need not be true to be useful in its applications. Nevertheless, mathematics must be conservative, that is, it cannot introduce any new ontological implications. A mathematical theory is conservative if, when we add it to a nominalistic theory *T*, no nominalist consequences follow that would not follow from the *T* alone:

> In the first place, a conservative mathematical theory might facilitate inferences from nominalistic theories. That is, if M is conservative, N + M doesn't imply A unless N itself implies A (where N is a nominalistic theory and A a nominalistic assertion); but it might be much easier to see that A follows from N + M than it is to see that A follows from N alone. As we will see later, mathematics really does serve to facilitate inferences in this sense, and that is certainly a large part of its value. (Field 1989, 58-59)

Field's strategy is, then, to demonstrate that we can nominalize our theories. The main idea is that we can eliminate quantification over mathematical entities and the resulting theory would still be reasonably attractive, so the aim is "to take physical theories stated in terms of numerical functors and try to restate them in terms of comparative predicates instead" (Field 1989, 130).

Field actually succeeded in the nominalization of a large fragment of Newtonian gravitational theory. From here, the idea is that, once we see how the elimination of reference to mathematical entities can be achieved for a typical physical theory, it would seem plausible that the project could be completed for the rest of science. That is, once we see how we can provide a nominalistic version of the differential fragment of classical gravitational theory, we should be able to see how we could use the same method to do the same to the rest of science, or at least that is the hope.

In Field's account, arithmetic would be some sort of a widely known story, and we accept mathematical propositions as long as they fit into that story, which would be just like any other fictional tale – hence the name *fictionalism* of his proposal. So, '5 is prime' and '5 is even' are only different in that the first one conforms to the story and the second one does not, but none of them is literally true, given that mathematical entities do not exist and, consequently, the mathematical discourse is false. Therefore, the acceptability of mathematical assertions comes from the fact that they follow from accepted axioms, a view that is also shared by Leng (2010).

This view does not include a rejection of mathematics altogether. On the contrary, Field recognizes the value of mathematics, speaking *as if* there were mathematical entities, although we do not believe in their existence. Therefore, it is not about disregarding any mention of mathematical objects in our scientific or ordinary discourse. In practice, mathematics is extremely useful to draw conclusions from our nominalistic theories, and going without its use would make our deduction procedures too hard and perhaps practically impossible.

Moreover, if we end up concluding that mathematics is indispensable *to scientific practice,* according to Field, this still does not mean that the indispensability argument is right, since we can present the assumptions of our best scientific theories and their consequences in non-mathematical terms, even if we use mathematics as the means to get from the former to the latter: "its special status arises from its utility, and since we've shown that it is always in principle eliminable (i.e. you don't get any results with it that you couldn't get without it), its utility is no grounds for believing it true" (Field, 1980, 24).

There has been a big debate about the likelihood of the success of Field's program. However, Field did not succeed at providing translations of more scientific theories, and some authors have pointed out the difficulties that extending Field's nominalization project to contemporary physical theories might carry.

For example, Alasdair Urquhart (1990) remained pessimistic about extending Field's program based on the convenience of Newtonian science, which makes references to space-time quite easy to make in terms of real numbers, since it assumes they have the same structure. In contrast, contemporary theories do not present an isomorphism between space-time and R^4, the four dimensional manifold of real numbers, and this lack of an isomorphism poses some difficulties for finding nominalizations for those theories.

In the same vein, David Malament (1982) points out that the Newtonian theories concern properties of space-time points, which is also convenient, whereas our current best scientific theories (such as Hamiltonian phase space theories or, *even worse*, quantum mechanics[5]) contain laws expressing relations between possible states of physical systems. This may constitute a problem since it carries an ontology of *possibilia*, which could be as problematic as the commitment to the existence of mathematical objects. In quantum mechanics, propositions and eventualities are given mathematical measure. This means that these propositions and eventualities are the basic objects with which we work, so the approach Field uses in *Science without Numbers* will carry some unacceptable ontological commitments for the nominalist approach, as Malament (1982, 534) ironically points out by asking: "What could be worse than *propositions* (or *eventualities*)?"

Of course, these objections are not necessarily the end of Field's program, but they do give an idea of the difficulties it may carry and of the fact that just "getting rid of numbers" is not such an easy task. They also stress the fact that the strategy followed by Field in *Science without Numbers* will not do the job applied to our current best scientific theories, which means that the fictionalist who wishes to keep Field's project must find a way to nominalize theories where an analogy with Newtonian physics is shown not to work. In conclusion, given the crucial role that mathematics plays in science, it is doubtful that every single one of our best scientific theories can be nominalized. If this is the case, then we need other strategies from the nominalist in order to account for the role of mathematics in scientific explanations without carrying an ontological commitment to abstract objects[6].

From here, and taking into account the technical difficulties to Field's program (analysed in Burgess and Rosen 1997), some authors began a search for nominalistic alternatives to his project, that is, nominalist accounts that do not require a complete nominalization of scientific theories. This is what came to be known as the *Easy-Road Nominalism*.

---

[5] It is worth noting that Balaguer (1997, Chapter 6) argues for the possibility of nominalizing quantum theory, even though it is not the main view on the subject.

[6] Nevertheless, there have been some attempts to continue the task. For instance, Mark Balaguer (1996a; 1996b; 1998) presented steps toward nominalizing quantum mechanics.

- **Other objections**

Here is a small selection of other objections to the Indispensability Argument in its original form.

To start with, Penelope Maddy (1995; 1997) follows a different strategy by pointing out that if P1 is false, then Field's project may turn out to be irrelevant to the question, so she has directed her analysis to P1, instead of P2. Maddy suggests that, if we take seriously the naturalist requirement to look to scientific practices to decide what we have reason to believe there is, then we ought not to have ontological commitment to all the entities indispensable to our best scientific theories, so P1 is not true, by arguing that there are problems of reconciling naturalism with confirmational holism. A holistic view of scientific theories has problems explaining the legitimacy of certain aspects of scientific and mathematical practices.

What this means is that we could believe a theory while remaining agnostic or instrumentalist about whether its objects exist. Physics is full of fictional idealizations, like infinitely long wires, centres of mass and uniform distributions of charge. Other sciences also posit objects that we do not really think exist, like populations in Hardy-Weinberg equilibrium (biology), perfectly rational consumers (economics), and average families (sociology). Positing ideal objects sometimes facilitates using a theory and scientists often do it to simplify analyses. It is established that we can believe a theory and at the same time recognize that some of the objects to which it refers are only ideal. If that is our attitude towards average families and infinitely long wires, it could be our attitude towards sets, numbers and squares as well.

Also, Maddy presents some objections concerning methodological consequences of accepting the Quinean doctrines of naturalism and holism.[7]

1. The actual attitudes of working scientists towards the components of well-confirmed theories vary "from belief, through tolerance, to outright rejection" (Maddy 1992, 280). Maddy suggests that we should side with naturalism and not holism here (since naturalism would point at respecting the methods of working scientists, and yet holism is apparently telling us that working scientists ought not to have such differential support to the entities in their theories).
2. Mathematical portions of theories could fall within the true elements of the confirmed theories or within the idealized elements. Maddy suggests the latter. Scientists themselves do not take the indispensable application of a mathematical theory to be an indication of truth in mathematics. Scientists would take any portion of mathematics that would do the job, without regard to the truth of the mathematical theory in question (Maddy 1995, 255).

   Since we have no reason to believe that the mathematical theory in question is true, we have no reason to believe in the existence of the entities posited by the theory.
3. It is hard to make sense of what working mathematicians are doing when they try to settle independent questions (that is, questions that are independent of the standard axioms of set theory, ZFC axioms). New axiom candidates have been proposed, but the problem is that the arguments that were advanced seem to have nothing to do with applications in physical science (they are intra-mathematical arguments).

---

[7] Part of this presentation of Maddy's objections is taken from Colyvan, M., "Indispensability Arguments in the Philosophy of Mathematics", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/mathphil-indis/>. (Accessed on the 1st of August, 2019)

> Confirmational holism again seems to be advocating a revision of standard mathematical practice, which is at odds with naturalism (Maddy 1992, 286–289).

Elliot Sober's objection (1993) is similar to Maddy's second and third objections. He argues that mathematical theories are not being tested in the same way as the empirical theories are tested in science. Rather than confirming entire theories, what we do is compare contrasting hypotheses, that is, hypotheses are confirmed relative to competing hypotheses. Following this line of thought, if we accept that mathematics is confirmed along with our best empirical hypotheses, as the IA seems to sustain, there must be non-mathematical competitors that allow that contrast. However, the mathematical core is common to all scientific theories and there are no such competing hypotheses. Thus, Sober claims, mathematics does not receive confirmational support from the evidence in the same way that other scientific hypotheses do. This gives some support to the idea that confirmational holism is wrong, and it is also an objection to Quine's view that mathematics is part of empirical science.

As for other objections, we should note that Parsons (1980) argues that the obviousness of basic mathematical statements is left unaccounted for and, similarly, Kitcher (1984, 104–105) contends that the indispensability argument does not explain *why* mathematics is indispensable to science. Furthermore, Mark Balaguer's plenitudinous platonism claims that mathematics provides a theoretical apparatus which applies to all possible states of the world (see Balaguer 1998). The purpose is to explain the applicability of mathematics to the natural world in a non-miraculous fashion, since any possible state of the natural world will be described by some mathematical theory.

It is also worth noting that the IA approach is an attempt to justify our mathematical beliefs without appealing to some sort of rational insight, also known as "mathematical intuition". Other platonist philosophers, who approach mathematical explanation from different perspectives (outside of the IA debate) have appealed to mathematical intuition or "seeing with the mind's eye". Examples of this kind of approach can be found in Brown (2012) and Gödel (1944)[8].

Lastly, these anti-holist arguments seem to suggest that the presence of references to mathematical objects in a scientific theory is not by itself enough to support the thesis that those objects exist, given that theoretical posits can sometimes contribute to the success of those theories even if they do not exist. This is precisely the central thesis of the easy-road nominalists, as will be discussed in the next section.

---

[8] Gödel's platonist approach is connected with his mathematical realism. According to this perspective, we do not construct or invent mathematics, we rather *discover* mathematical facts and these facts, consequently, have an existence that is independent from our mental states in the same way that the empirical world is independent from our sensory experiences. Gödel (1944) even compared the way we get knowledge of mathematical facts with the way we sense the empirical world: "It seems to me that the assumption of such objects [classes and concepts] is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a satisfactory theory of our sense perceptions" (Gödel 1944, 137. This quote can be found in Parsons, Ch., (1995): "Platonism and mathematical intuition in Kurt Gödel's thought", *Bulletin of Symbolic Logic*, 1(1): 44–74).

3. Easy-Road Nominalism

As suggested above, on the nominalist (or ontologically “neutral”) side of the debate around the existence of mathematical entities, Field’s program is not the only choice available. There is an alternative –in fact, a series of alternatives – to avoid mathematical platonism that does not involve the strategy of nominalizing all our scientific theories, which makes it an *easier* path to account for the role of mathematics without committing us to mathematical ontology. This set of accounts is known as the *Easy-Road Nominalism*, as Colyvan (2010) labelled it.

There are many different accounts or analyses within what we may label as Easy-Road nominalism, among which we can cite Jody Azzouni (1997; 2004; 2012), Mark Balaguer (1996a; 1996b; 1998), Otávio Bueno (2012), Mary Leng (2010; 2012), David Liggins (2012), Penelope Maddy (1995; 1997), Joseph Melia (2000; 2002) and Stephen Yablo (2002; 2005; 2012).

The ones mentioned –and many more– are all different approaches to the role of mathematics in empirical science and different nominalistic alternatives to Field’s program. That is, precisely, what they all have in common: the rejection of mathematical Platonism with attempts to provide an easier path to nominalism than that of Field’s. Moreover, these authors grant that mathematics may be indispensable to our best scientific theories, but this is not enough to force an ontological commitment with the existence mathematical entities. They are *no-nominalization responses, that is, they do not involve avoiding quantification over mathematical entities.*

From the other side of the debate, Mark Colyvan (2010) analyses the Easy-Road nominalist approach and argues that it is not the right approach to the debate. In his 2010 paper, Colyvan focuses mainly on the proposals of Azzouni, Melia and Yablo. The three of them suggest that it is a mistake to read our ontological commitments from the range of the quantifiers of our best theories (see Colyvan 2010, 286), but acording to Colyvan, they all fail for the same basic reason: the proposals would ultimately need to presuppose the success of Field’s program in order to work, which would make them *uneasy* approaches anyway and would ultimately collapse with the *hard road* (see Colyvan 2010, 286-7). Two years later, following some responses from nominalists, Colyvan (2012a) strikes again in a paper titled “Road Work Ahead: Heavy Machinery on the Easy Road” (2012a), in which he replies to Azzouni, Bueno, Leng, Liggins and Yablo. This series of papers gives the idea that we are facing some sort of a very recent and direct debate that went on between the nominalists on one side and Colyvan on the other trying to make the case against any form of an *easier* way to nominalism.

Since it would be impossible to examine all the easy-road nominalist proposals here, the focus of this section will be on three accounts that show some relevant features for the purposes of this work: Balaguer (1996a; 1996b; 1998), Leng (2010; 2012), and Melia (2000; 2002). As we will see, they raise similar arguments, for they agree that the value of mathematics comes from its nominalistic content. In the process, we will take some of Colyvan’s concerns in our analysis of some of the *easy* approaches to the case of mathematical explanation[9]. Ultimately, what Colyvan (2010; 2012a) and Baker (2005; 2009a) argue against the easy-road is that

---

[9] Note that Baker (2005, 2009a) also responds to the Easy-Road Nominalism. We will see more on Baker’s work on mathematical explanation in Chapter 2.

mathematics does not play just a descriptive role in science, but also an indispensable explanatory role.

3.1. Mark Balaguer and the distinction between physical and mathematical content

Balaguer’s (1996a; 1996b; 1998) main thesis is that we should believe the “physical content” in our theories and remain agnostic about its “mathematical content”.

Of course, one of the challenges of this account is providing a criterion to distinguish between “physical content” and “mathematical content”.

Balaguer grants that there are some indispensable applications of mathematics to empirical science. What he denies is the idea that we ought to have ontological commitment to the entities that are indispensable to our best scientific theories (P1). His strategy is, then, to account for these applications while avoiding any ontological commitment to the existence of mathematical entities.

So, the argument would be something like the following[10]: In the case of there being abstract objects, then they would be causally inert. If this were the case, it follows that the truth of a scientific theory would depend on two sets of facts that hold or do not hold independently one of another. One set contains only purely platonistic and mathematical facts, and the other contains only purely physical facts. Given that scenario, fictionalists can maintain that it is coherent to believe that these purely physical facts (those required to make science true) while at the same time not believing any of the purely platonistic facts (since there are no abstract objects). This would keep fictionalism consistent with a form of scientific realism – which Balaguer calls ‘nominalistic scientific realism’ – involving the belief in the physical or nominalistic content of our scientific theories.

In Balaguer’s picture, empirical theories use mathematical-object talk in order to construct theoretical apparatuses to make assertions about the physical world, so mathematics plays a descriptive role. In such a picture, mathematics provides an *easy way* to express our claims about facts of the physical world. A fictionalist perspective would be consistent with scientific realism because fictionalists can defend that, even if there are no mathematical entities (and, therefore, mathematics cannot be strictly true), our theories are still accurate in their representation of the world. It is not even needed for our mathematical discourse to be true in order to have this kind of representational role.

Balaguer is not himself a nominalist; his aim is simply to show that a nominalist attitude to our scientific theories is coherent. Thus, in 1998, he presented arguments for both *plenitudinous platonism* and *fictionalism*. Plenitudinous platonism states that for every logically possible mathematical theory, there is a portion of mathematical reality that the theory truly describes (so that the mathematical realm is ‘plenitudinous’ in containing all the mathematical objects there logically possibly could be). It follows from this plentitudinous picture that the mathematical realm is full enough to provide a theoretical apparatus that is applicable to possible states of the world. This means Balaguer can provide a non-miraculous explanation to

---

[10] Part of the presentation of Balaguer’s strategy is taken from: Balaguer, Mark, "Fictionalism in the Philosophy of Mathematics", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/fictionalism-mathematics/>.

the applicability of mathematics to the natural world, by arguing that any possible state of the natural world will be described by a mathematical theory. After discussing the arguments for both approaches to mathematical ontology, Balaguer concludes that there is no fact of the matter about which one is right. This “there is no fact of the matter” thesis will come back to the present work in several occasions, as we will see.

3.2. Melia and the *weaseling* strategy

Joseph Melia (2000) claims that we can assert our scientific theories and just take back the platonistic consequences in them. For this, he employs the resource of *weaseling* when accounting for the mathematical case. According to this strategy, just as when we say that “all Fs are Gs, except b”, it is not contradictory to assert claims such as “There exists a differentiable function that maps from the space-time manifold to the real numbers, but there are no mathematical objects” (Colyvan 2010, 294).

Melia establishes an analogy with our ordinary discourse by arguing that sometimes weaseling is the only way we can say what we really mean. In the case of mathematical entities, this is exactly what is going on when we speak of our ontological commitments to abstract objects.

Colyvan finds some problems with Melia’s account. The main problem is that “if we cannot say what we want any other way except by weaseling, it is just not clear what we are saying” (Colyvan 2010, 295). He argues that there are limits to how much we can retract, that is, there are limits to the weaselling strategy, by pointing out that if we take mathematics back from our sciences, we no longer know exactly what we are talking about. Furthermore, if we ask for a translation of sentences with references to mathematical sentences, then the weasel is committed to providing a translation which does not contain any commitment to mathematical objects (see Colyvan 2010, 295). So, once again, Colyvan’s bottom line is that the easy-road ends up depending on the success of the hard-road.

Liggins (2012) defends Melia’s thesis and his strategy of the weasel. He points out that Melia’s account does not render the contents of our best scientific theories obscure and that it overcomes the technical difficulties of Field’s program that we discussed above. Liggins objects to Colyvan’s claim that the only plausible way to make the content clear after using the strategy of the weasel is to provide a translation, which, according to Colyvan, would go back to the hard road nominalism. Colyvan summarizes Liggins’ claims from his 2012 paper in the following way: “more specifically Liggins argues that (i) my demand for a nominalistic restatement of all our best scientific theories is unreasonably strong; (ii) Melia has already met a more reasonable demand; (iii) as a consequence Melia’s view does not render the contents of our best scientific theories obscure; and (iv) the view does not give rise to technical difficulties such as those facing Field’s (1980) nominalisation program (Colyvan 2012a, 6).

Liggins argues that it is not reasonable to demand a translation for every sentence. The request would be unreasonable for sentences like “There exists a differentiable function that maps from the space-time manifold to the real numbers, but there are no numbers” (see Colyvan 2012a for his discussion of the issue), given that not all the mathematical content of our scientific theories can be expressed in a nominalistic way. However, Liggins does grant that we should be able to provide translations of the contents of the nominalistic theory, which, he argues, can be done by providing translations of less problematic sentences.

Daly and Langford (2009) also defend Melia's approach when they claim that the role of positing concrete unobservables in science allows the explanation of the behaviour of observable facts. Mathematics, in contrast, does not show this explanatory value, for it merely represents or indexes physical facts. Therefore, the presence of the concept of primeness in the explanation of the cicada case is a product of our choice to measure the periods in years instead of any other method (months, seasons…). If this observation is correct, then Melia's approach can also be applied to the cicada case, and 13 and 17 years would be just a way of indexing the empirical phenomenon.

3.3. Leng's fictionalist proposal

A particularly illuminating account of the ontology of mathematical objects was advanced Mary Leng in *Mathematics and Reality* (2010). In this section, we will also take into account some ideas on her paper "Is there a fact of the matter about the existence of abstract objects?" (2020), a question that goes straight to the issue we attempt to tackle in this work.

Leng (2010) argues for taking a naturalist approach to ontology, which means that we should look to natural sciences to tell us what there is and what there is not. In the debate between scientific realism and constructive empiricism, Leng offers a third option: mathematical fictionalism. If realists argue that we should believe our best scientific theories in their entirety and constructive empiricists argue that we should only believe their observational content, fictionalists hold that we should believe only our best scientific theories' nominalistic content, a feature this account shares with Balaguer's.

In a scenario where mathematics is indispensable and, thus, Field's strategy does not work for all mathematical explanations, Leng takes a non-realist point of view when she argues that the concepts of mathematical "truth" and "existence" do not carry *genuine* truth and existence. Moreover, the actual existence of mathematical objects does not affect mathematical practice at all. Then, mathematical practice would be ontologically neutral, so it would in principle be compatible with both Platonist and anti-Platonist accounts of mathematical ontology.

Leng's thesis in *Mathematics and Reality* (2010) is that scientific practice shows that scientists see themselves as committed to the existence of objects when the hypothesis of their existence is required to explain the success of the theories in which they occur. This leads to the need for an account of the success of mathematics in scientific explanation in order to settle the question of the existence of mathematical objects. In this task, the most challenging case is the application of mathematics in thoroughly mathematized theories such as physics, where it is often hard to distinguish what is purely physical and what is mathematical. The nominalization of our current theories is problematic since these theories appear to make stronger assumptions about space and time than scientists can take to be warranted (for instance, they model space-time as continuous), and apparently they even make false assumptions for the purpose of simplification, for instance, by using continuous functions to model properties such as temperature.

In this sense, Leng puts forward the idea that the pretence that "real objects are like the objects in our model can allow us to provide a good explanation of the phenomena to be explained, even though its literal content (that there are ideal abstract objects that resemble concrete objects in some respects) may be false." (Leng 2010, 221). The explanatory value of appeals to mathematical objects is not a result of the existence of such objects, but a result of the aptness of the pretence that such objects are related to non-mathematical objects in the ways our explanations suppose.

In the context of the debate around the cicada case, Leng (2010) argues that the theorem in question is applied to the years in which cicadas appear (with them being numbered), and a fictionalist about mathematical objects would not in principle have any problems with being realist about time and periods of time. The explanation based on the concept of primeness is a good explanation in the sense that "our modelling of the succession of years as set with a set-theoretic relation satisfying the Dedekind–Peano axioms is an *apt* one, only because it *respects* facts about the years and their succession, that our explanation of the behaviour of cicadas in terms of facts about prime numbers is a good one" (Leng 2010, 248). Although we appeal to mathematical objects such as sets, set-theoretically defined relations, etc., in the explanation of the cicada case, what is actually happening there is that our mathematical model is a suitable one, for it respects fundamental facts about the succession of years and about the behaviour of cicadas.

More recently, Leng (2020, 112) elaborates on her 2010 proposal to develop "a more thoroughgoing naturalism" as the approach to ontology that can save the post-Quinean ontological project from some recent sceptical concerns put forward by authors like Yablo (1998) or Azzouni (1997). This strategy involves looking to science itself for an account of ontological commitment in mathematics.

Looking at scientific practice, it seems that scientists do make a distinction between those parts of the theories that are taken literally as oppose to those that are taken instrumentally. In particular, there is a difference between assertions that 'there exist *F*s' and 'there really exist *F*s' (see Leng 2020, 123). Also, as Leng already put forward in 2010, the focus must be in uncovering which genuinely ontological claims are justified given scientific standards of evidence.

In line with this, Leng's criterion for ontological commitment would be the following: "if our best explanation of the success of a theoretical posit requires the assumption that the object posited really exists, then this will support ontological commitment to that object. On the other hand, if we can explain the success of a theory that quantifies over Φs from a perspective which sees Φs as merely useful fictions, then the empirical success of our theory does not constitute evidence for Φs" (Leng 2020, 124).

Leng (2020) introduces a relevant question regarding the sense in which we use the predicate "exists" and the fact that when we try to answer questions of existence, part of what determines whether we accept the answer as correct or not is the concept of existence we are using. Even in informal contexts, Carnap's internal/external distinction seems to have a place in the debate. When we ask about the existence of numbers in an internal "non-worrying" "according to standard mathematics" sense, the answer seems to be an obvious yes. However, when we push the issue a bit further and ask whether numbers "really" exist, spatiotemporally, the answer is a clear no. Even in academic contexts this seems to be the most natural attitude towards the question, if 'exists' means 'exists spatiotemporally'. But this raises the question of what we should mean by 'exists' when answering ontological questions. Whereas Carnap thinks that there is no meaningful answer to this question, Leng accepts Quine's response to Carnap that ontological questions about what 'really' exists should be answered with reference to 'exists' as it is used in the natural sciences.

In this context, as Leng (2020) points out, it looks like Maddy is right when she claims that, taking "exists" in the sense at work in the natural sciences, the question of whether we should argue that numbers exist or are just fictions ultimately has no deep answer. This is supported by the idea that scientific practice seems to leave open more than one way of going on with the term ‘exists’, which makes it compatible with both thin realist or arealist approaches. Having discarded *Robust Platonism*, Maddy argues that “Thin Realism and Arealism are equally accurate, second-philosophical descriptions of the nature of pure mathematics. They are alternative ways of expressing the very same account of the objective facts that underlie mathematical practice” (Maddy, 2011, 112). The correctness of mathematical claims is, according to Maddy, grounded in objective facts related to mathematical depth.

Maddy contrasts *Thin Realism* and *Arealism* with a more traditional *Robust Realism*, according to which the facts about what mathematical objects there are not grounded in facts about mathematical depth. Her aim is to show that the empirical evidence and our empirical concepts cannot decide between Thin Realism and Arealism, which share the assumption that the goodness of mathematics is grounded in the objective depth of mathematical concepts and theories, disagreeing only about whether mathematics is also true or not.

Even if this is the case and we cannot decide whether Thin Realism or Arealism are right, we can discard Robust Realism. Therefore, there may be some facts of the matter about the existence of abstract objects, in the sense that we have no reason to believe that mathematical objects exist in the strong sense envisaged by the ‘robust realist’ picture.

To sum up, Leng’s central idea is that we should treat mathematics as some sort of *if-then* theory. Mathematics just tells us what must be true of any system that has the relevant mathematical structure. In this picture, there is no reason to take the explanations in question to involve mathematical objects. Therefore, we are committed only to the physical instantiation of mathematical structures, not to the system of abstract mathematical objects itself (see Leng 2012, 993). Therefore, the acceptability of mathematical assertions comes from the fact that they follow from accepted axioms, and the indispensability argument does not refute fictionalism because fictionalists can provide an adequate account of the success of science.

As for our purposes in this chapter, the key ideas to highlight from Leng’s account of mathematical explanation are that (1) mathematics’ role is to represent phenomena, and (2) this representational role does not carry a commitment to the existence of mathematical entities or the truth of mathematical statements that figure in those representations. In addition, (3) it is reasonable for scientists to make use of mathematics without believing in the existence of the entities involved, so the indispensability argument is flawed for naturalistic reasons.

From this section, we can conclude that *easy-roaders* all agree that mathematics can be useful without being true. One issue to take into account is that these authors do not elaborate easy-road alternatives *just* because they do not involve the nominalization of our empirical theories, that is, they do not do it *only* because Field’s strategy is too difficult. Namely, Melia, Yablo Balaguer and Leng argue that their view is superior to Field’s view in that they account better for scientific practice. The grounds for their reasoning are that, since they reject holism, they can also reject the existence of mathematical entities and accept the existence of physical theoretical entities, in that the concrete and mathematical realms work in different ways with regard to the ontological question.

There is a sense in which these theories seem to have provided a very *reasonable* way to account for mathematics’ role in science and to comply with our beliefs with respect to mathematical objects. It is plausible that scientists and mathematicians may work without even considering the question of mathematical ontology or mathematical truth. Perhaps it is more reasonable to limit ourselves to taking mathematics to be merely useful, which is a thesis shared by the three authors considered in this section.

Another question is whether the ordinary speaker (or even the scientist, or even the mathematician) would be willing to accept the thesis that the mathematical *discourse* cannot be true and that statements such as “2 + 2 = 4” are false or, at least, *not literally true.* However, going against ordinary speech does not worry the easy roader, since their account of mathematical correctness accounts for why it is natural to treat some claims as ‘true’ even though they are not literally true.

There is no doubt that the easy-road approaches look quite promising in accounting for mathematical explanation without leading to an ontological commitment to the existence of the relevant abstract entities for anyone wishing to avoid such ontological commitments. However, Baker and Colyvan have come back with new arguments against such approaches. Basically, as a response to nominalists who claim that mathematics has a merely representational role and that this role can be played by fictions, what Baker and Colyvan argue is that mathematics is doing something more than just representing phenomena, and that nominalist approaches, even *easy-road* ones, fail to account for these other roles, namely the *genuine* explanatory role. This has led to the current situation, where the debate is taking place in other terms. In order to overcome the challenge presented by the *easy-road* proposals and capture this genuine explanatory role of mathematics, platonists have presented what we can call *the Enhanced Indispensability Argument* (EIA), which is currently in the centre of the debate.

This process takes us to the current state of the debate. Hopefully, we will be able to see some advances in the debate in Chapter 2.

4. The Enhanced Indispensability argument (EIA): an introduction[11]

Faced with the problems that arise within the debate of the classical Quine-Putnam Indispensability Argument, some mathematical platonists such as Baker (2005) and Lyon and Colyvan (2008) have argued for the *Enhanced Indispensability Argument* (EIA).

This new wave of indispensability arguments departs from Quine in that it seeks the justification of our mathematical beliefs in scientific explanations which rely on mathematics, rather than in scientific theories generally. Moreover, it relies on case studies in order to convey the idea that mathematics is explanatorily indispensable in a way that carries ontological commitment with mathematical objects. In addition, the EIA follows the idea that mathematics does something more than describing, and this *substantive* or *genuine* role can be used to justify an ontological commitment to the existence of mathematical entities.

This shows a tight connection between scientific explanation and mathematical explanation, or between philosophy of mathematics and philosophy of science.

One of the versions of the EIA that Baker formulates is the following:

> **The Enhanced Indispensability Argument**
>
> (1) We ought rationally to believe in the existence of any entity that plays an indispensable explanatory role in our best scientific theories.
>
> (2) Mathematical objects play an indispensable explanatory role in science.
>
> (3) Hence, we ought rationally to believe in the existence of mathematical objects.
>
> (Baker 2009a, 613)

This constitutes again an argument for mathematical platonism, as our best science appeals to mathematical entities in explanatory contexts. The basic idea is that since mathematical entities arguably feature in our best scientific explanations on a par with unobservable entities, and since rational belief in the unobservable entities is defended via inference to the best explanation, the scientific realist should also be a mathematical Platonist regarding the explanatory role of mathematics.

While nowadays few philosophers defend the Quine-Putnam version, given its limitations, Baker's version of the argument does not depend on holism. It is rather supported by case studies, such as the life-cycle of periodical cicada from evolutionary biology.

The ways this argument allegedly enhances the original are the following (see Saatsi 2011): first, it focuses on the indispensable explanatory role of mathematics—not just on indispensability simpliciter. Second, it appeals to the role of inference to the best explanation in the defence of scientific realism.

If we apply this argument to the cicada case (as we shall see in Chapter 2) Baker's conclusion is that "[t]he explanation makes use of specific ecological facts, general biological laws, and number theoretic results. My claim is that the purely mathematical component [prime

---

[11] This section intends to be just an introduction on the subject of the EIA debate in order to clarify the history and the natural reasoning behind the importance of examining the role of mathematics in explanation. A more detailed analysis of such debate takes place in Chapter 2.

periods minimize intersection (compared to non-prime periods)] is both essential to the overall explanation and genuinely explanatory on its own right. In particular it explains why prime periods are evolutionary advantageous in this case" (Baker 2005, 233).

The EIA raised many criticisms, of which many have focused on premise (2), and on questioning the claim that mathematics plays the *right* kind of role. As authors who have posed challenges to the argument, we can mention Leng (2005, 2010), Saatsi (2007, 2011, 2016), Bangu (2008), Daly and Langford (2009), Rizza (2011) and Pincock (2011), among many others.

The approaches vary. Just to take a few, while Leng (2010) argues that both Colyvan and Baker infer illegitimately from the existence of mathematical explanation that the statements grounding the explanation are true and, since mathematical explanations need not have a true *explanans*, the objects posited by such explanations need not exist; Saatsi (2007) focuses on inference to the best explanation to also resist the implication that scientific realism carries a commitment to mathematical platonism; and Rizza (2011) is deflationary about the alleged ontological consequence of the role of mathematics in the cicada case, but he does not deny the centrality of mathematical concepts in constructing explanations.

The general idea is that to reject the EIA by refuting one of the premises of the argument either that mathematical objects play an indispensable explanatory role in science or that it commits us with the existence of such entities.

Since this debate will reappear in Chapter 2, we can leave it for now by saying that focus is not on the indispensability of the reference to abstract entities in our explanations of empirical phenomena anymore, but on the explanatory indispensability of mathematics. That is, what is now in question is whether we can explain certain phenomena without the reference to those abstract (mathematical) entities.

5. The focus on explanation. Classical theories of explanation

The debate of the Enhanced Indispensability Argument (EIA) brings the question of what we mean when we say that mathematics explains. Seeing that the ontological debate cannot be settled taking just mere indispensability of the reference to mathematical entities into the picture, there has been a shift in the research to focus on mathematical explanatory indispensability. Here, the ontological debate meets with the debate about mathematical explanation and mathematical application to other disciplines.

In this context, it becomes necessary to go to the literature for a brief digression into some classical models of explanation applied to the mathematical case, in order to see the options of theories of explanation *in the market*. We will begin by examining two classical approaches to explanation in philosophy of science: the Deductive-Nomological Model and the Causal-Mechanistic Model.

More recently, we find the Counterfactual Theory of Explanation (CTE) as one of the most promising general theories of explanation, and indeed one that will receive a great deal of attention in this work. Moreover, there has been an attempt to adapt the DN model to the specific mathematical case by Baron (2016, 2017) and to overcome the classical DN model's problems. A more detailed analysis of this approach will take place in Chapter 2.

5.1. The Deductive-Nomological (DN) Model

According to the Deductive-Nomological model, a scientific explanation consists of two constituents: the *explanandum* and the *explanans*. "By the *explanandum*, we understand the sentence describing the phenomenon to be explained (not that phenomenon itself); by the *explanans*, the class of those sentences which are adduced to account for the phenomenon" (Hempel 1965, 247). There are two subclasses in the *explanans*: the specific antecedent conditions $C_1, C_2, \ldots, C_k$ and the set of sentences $L_1, L_2, \ldots, L$, which represent general laws (*ibidem*).

A *sound* explanation needs to satisfy certain logical and empirical conditions of adequacy, which we briefly state below (see Hempel 1965, 247-8).

As logical conditions, (R1) the *explanandum* must be a logical consequence of the *explanans*, (R2) the *explanans* must contain general laws which are required for the derivation of the *explanandum* (here is the "nomological" component of the theory), and (R3) the *explanans* must have empirical content.

In the set of empirical conditions, (R4) the sentences constituting the *explanans* must be true. This corresponds to a veridicality constraint, or a condition of "factual correctness" in Hempel's terminology. He, however, later points out that it is more correct to state it as a requirement that the *explanans* be "highly confirmed by all the relevant evidence available rather than it should be true" (Hempel 1965, 248).

The relation holding between the *explanans* and the *explanandum* is a relation of logical deduction. That is, the explanation should take the form of a sound deductive argument in which the *explanandum* follows as a conclusion from the premises in the *explanans*. This is the "deductive" component of the model.

Hempel points out (see 1965, 248) that the same formal analysis and the four necessary conditions also apply to scientific prediction as well as to explanations, which shows that in this approach to explanation there is a strong connection between explanation and prediction. Indeed, Hempel even goes one step forward and argues that an explanation of a particular event is not fully adequate unless its *explanans* could have served as a basis for predicting the phenomenon in question (see 1965, 249).

The kind of explanation Hempel is referring to is causal explanation: "If E describes a particular event, then the antecedent circumstances described in the sentences $C_1, C_2, \ldots, C_k$ may be said jointly to "cause" that event, in the sense that there are certain empirical regularities, expressed by the laws $L_1, L_2, \ldots, L_k$, which imply that whenever conditions of the kind indicated by $C_1, C_2, \ldots, C_k$ occur, an event of the kind described in B will take place. Statements such as $L_1, L_2, \ldots, L_k$, which assert general and unexceptional connections between specified characteristics of events, are customarily called causal, or deterministic, laws. They must be distinguished from the so-called statistical laws which assert that in the long run, an explicitly stated percentage of all cases satisfying a given set of conditions are accompanied by an event of a certain specified kind" (Hempel 1965, 250).

- Problems within the DN model

This account has been shown to present several problems and is currently discarded by the vast majority of philosophers of science. These problems were analysed by Wesley Salmon (1989: 46-50).

Temporal relations

First, there is a concern regarding the temporal relations between the explanatory facts and the facts-to-be-explained. The statements of antecedent conditions are given no temporal constraints in the formal setup of the theory. The issue, as Salmon points out, has been treated in terms of the explanation of an eclipse.

An occurrence of a total lunar eclipse can be explained by deducing it from the relative positions of the Earth, sun and moon in conjunction with the laws of celestial mechanics. However, it is also possible to deduce the occurrence of the eclipse from those positions at some time after the eclipse has occurred with the very same laws (see Salmon 1989, 46). The latter will hardly constitute an explanation of the phenomenon.

Asymmetry

Another issue is related to the role of causality in scientific explanation. It is the problem of asymmetry. There are many examples – rather, counterexamples – that present this problem. We will now focus just on the first one, the well-known example of the flag and the shadow.

A vertical flagpole of a certain height under the sun casts a shadow of a certain length. Given either the height of the flagpole or the length of its shadow, we can deduce the other one. The deduction of the length of the shadow from the height of the flagpole may be accepted as a legitimate DN explanation. However, the deduction in the other direction can hardly be considered an explanation, that is, it is unlikely that the height of the flagpole can be legitimately explained by the length of its shadow. As Salmon (1989, 47) points out, the reason for the asymmetry is that a flagpole of a certain height causes a shadow of a certain length so the height can explain the length and not the other way around, though it would fit the DN model.

Therefore, the DN model is insensitive to these kinds of directional or asymmetric features shown by some explanations.

Other examples concern the case of the barometer and the case of the moon and the tides. From the barometric reading, we can predict a storm, but the reading is not an explanation of the storm. In the same way, pre-Newtonians could predict the behaviour of the tides from the position and phase of the moon, but they had no real explanation of them. There are certain correlations that provide bases for prediction and still do not constitute or provide explanations[12].

The key feature of these examples is that they challenge the symmetry of explanation and prediction, which constitutes one of the core theses of the D-N model defended by Hempel and Oppenheim. According to it, prediction is construed as "inference from the known to the unknown" (see Salmon 1998, 48).

The symmetry thesis can be thought of as a two-part thesis (*ibidem*):

(1) Every D-N explanation is a prediction (given the right pragmatic situation)
(2) Every nonstatistical scientific prediction is a D-N explanation.

Other counterexamples include Scriven's syphilis and paresis example. Paresis can occur only in individuals who go through the primary, secondary and latent stages of syphilis without being treated with penicillin. The explanation for a case of paresis is that it is due to latent untreated syphilis. Nevertheless, only about 25% of these will develop paresis, which means that it is not correct to predict that individuals with latent untreated syphilis will develop paresis. This is a counterexample of the broader symmetry thesis.

Relevance

Another fundamental problem concerning the D-N model has to do with relevance[13]. It is illustrated with two counterexamples: the hexed salt and the birth-control pills examples.

Taking one of them, in the case of the birth-control pills, the example is that John Jones is a male who has regularly taken birth-control pills and has not become pregnant. The law that any male who regularly takes contraceptives avoids becoming pregnant is sufficient to conform to the requirements of the DN model, but intuitively fails to be an explanation, since the fact that John takes birth-control pills is irrelevant to the fact that he has not become pregnant.

Therefore, it can be argued that the DN model fails to establish sufficient conditions for explanation, so explaining a phenomenon cannot be just a matter of showing its nomic expectability.

To sum up, the Deductive-Nomological Model faces some serious problems that led to the vast majority of researchers in philosophy of science and mathematics to consider it a discarded theory. However, we should note that there have been some attempts at providing an improved approach based inspired by the D-N model, such as Sam Baron's (2017) Deductive-Mathematical theory of explanation. This account will be discussed in Chapter 2.

Interestingly, problems such as asymmetry have motivated accounts of explanation based on causation, that is, causal theories of explanation. However, as we shall see, mathematical explanation does not fit the causal picture.

---

[12] See Salmon 1998, 47 for details.

[13] As we shall see in Chapter 2, Baron (2016) is also concerned with this issue.

5.2. The Causal-Mechanistic Model

- Explanation and causation

Even though nowadays almost the entire philosophical community takes for granted that not all (scientific) explanations are causal, this idea is not at all an old one. As Mancosu (2008) points out, there has been a shift from the idea that all explanations in natural sciences are causal to the questioning of this idea and the conclusion that it is not true, which today we sometimes take as a starting point. Mathematical explanation had a crucial role in this switch.

Indeed, the contributions in the analysis of non-causal explanations have become more and more prominent in the last two decades. Many recent papers in philosophy of science and philosophy of mathematics are concerned with compelling (already paradigmatic) cases of non-causal explanation. The cicada case, the honeybee comb and Königsberg bridges are just a few of many examples that are dominating the most recent literature on mathematical explanation[14].

In the second half of the twentieth century, however, the dominant accounts of scientific explanation were causal approaches. Causal accounts present the central idea that to provide an explanation of a given scientific fact is to provide its cause. Among these accounts, furthermore, we find mechanistic accounts, according to which providing an explanation of a given fact is to provide the mechanism that yields the fact in question.

Causal approaches to scientific explanation seemed to overcome the problems that were found in the classical DN model. However, they also present some problems, as we shall see.

- Salmon's Causal Mechanistic Model

Let us take Salmon's (1984) development of the Causal Mechanical (CM) model of explanation as the paradigm of causal accounts, for our purpose of seeing the main consequences of adopting a causal approach to scientific (and mathematical) explanation.

After presenting the Statistical-Relevance model (the analysis of which goes beyond our focus here), Salmon came up with what he called the CM (Causal-Mechanical) model. It can be understood as a tentative to capture the feature of explanations that goes beyond mere statistical relevance within a Humean framework.

The CM model involves the notion of process, which includes spatio-temporal continuous entities, such as waves and material objects. In particular, the notion of *causal process* is central. A causal process is a physical process that is characterized by the ability to transmit a *mark* in a continuous way. A *mark* is a local modification to the structure of a process, and a process is capable of transmitting a mark if it persists to other spatio-temporal locations in the absence of further interaction once it is introduced at one particular spatio-temporal location. Causal processes transmit causal influence in the form of information, energy or structure, in contrast to pseudo-processes, which are different from causal processes in the sense that they lack the ability to transmit marks.

Woodward (2004)[15] uses the example of the shadow of a moving physical object as a pseudo-process. If we try to mark the shadow by modifying its shape at one point, the modification will not persist unless we continually intervene to maintain it as the shadow occupies successive spatio-temporal positions. This feature differentiates this case from genuine causal processes.

---

[14] Reultinger and Saatsi (2018) edited a volume devoted to this issue.

[15] Part of this presentation is taken from the entry titled "Scientific Explanation", https://plato.stanford.edu/entries/scientific-explanation/#CauMecMod, accessed on the 9th of July 2019.

In Salmon's account, the notion of *causal interaction* is also crucial. A causal interaction involves spatio-temporal intersection between two causal processes which modifies the structure of both.

According to this account, an explanation traces the causal processes and interactions leading to the event in question. The idea underlying the approach is that to explain a phenomenon is to locate it at some point in the net of causal processes, which are the physical mechanisms responsible for the phenomena of the world (see Salmon 1984, 123).

- **Limitations to the theory**

The first concern has been pointed out by Hitchcock (1995) and is related to explanatory relevance. He points out that there is nothing in the notion of mark transmission or causal process that allows us to distinguish between the explanatorily relevant features of an explanation and the explanatorily irrelevant ones. More formally, those features of a process *P* in virtue of which it constitutes a causal process may not be the features of *P* that are causally or explanatorily relevant to the outcome we are trying to explain, that, is, the mark may be irrelevant to the event with some other property *R* of *P* being the property which is causally relevant to the event.

There is another worry related to the application of the CM model to systems departing from simple physical paradigms, when these explanations involve action at a distance or causal interactions that do not include intervening spatio-temporally continuous processes or transfer of energy and momentum from cause to effect. This would put Newtonian gravitational theory as unexplanatory, for instance.

The CM model has been criticized for its limits in application, since it is arguably only adequate with respect to physical and chemical causation, and very direct and obvious forms of causation. Campaner[16] points out that the view has been accused of imposing too strong requirements, providing a network of processes and interactions that is only applicable to idealized or simple cases and lacking of indications of how to identify the explanatorily relevant causal processes and interactions (see Campaner 2012).

More details on criticism of the CM model can be found in Kitcher 1989, Woodward 1989 and Hitchcock 1995.

- **A note on counterfactuals**

In Salmon's characterization, the ability to transmit a mark is a counterfactual notion. Indeed, the notion is counterfactual in that a process may be causal even if it does not transmit any mark, provided that if it were appropriately marked, it would transmit the mark. Moreover, the notion of marking involves a counterfactual contrast between how the process behaves and how it would behave if left unmarked. As Campaner (2012, 200) points out, Salmon (1984) appeals to counterfactuals with some philosophical regret (since it was seen as a problem given the Humean structures the theory was supposed to satisfy) and is glad to later abandon them, instead adopting Phil Dowe's "conserved-quantity theory" (Dowe, 2000), an approach that does not rely on counterfactuals.

After some concerns brought up by Kitcher (1989), Salmon (1994) tried to adapt his theory in order to avoid any appeal to counterfactuals. The new version, partly inspired by Dowe's (2000) conserved process theory of causation, defines a causal process as a process that

---

[16] From the review: Commentary Scientific Explanation and the Causal Structure of the World Wesley Salmon Princeton University Press, Princeton, 1984 Raffaella Campaner (Humana.Mente Journal of Philosophical Studies, 2012, Vol. 21, 197 – 204) https://pdfs.semanticscholar.org/f686/8974f46f57e78846f6947f49f605adcee185.pdf

transmits a non-zero amount of a *conserved quantity* at each moment in its history. Thus, the concept of *conserved quantity* becomes one of the central concepts of the theory. Without going into much detail about this new version, interestingly enough, it does not avoid either the dependence on counterfactuals or the problem related to causal or explanatory relevance mentioned earlier. The relevance problem is still present in that the transmission of the conserved quantity in question may not indicate the features of the causal process that are relevant to the phenomenon to be explained. There are, without a doubt, a number of conserved quantities in the process of John's taking birth-control pills. However, this does not determine which features in the process are relevant and which are irrelevant to his failure to get pregnant, so the notion of a causal process does not account for the notion of relevance, either related to causes or explanation.

5.3. Non-causal explanation

Within causal approaches to explanation, the sciences explain by identifying the cause(s) of the phenomenon to be explained. More particularly, according to the mechanist version of causal accounts, the sciences explain by identifying the causal mechanisms for that phenomenon.

As Reutlinger (2017) points out, there are several features that can account for the attractiveness of causal accounts. First, as is pointed out above, the approach focused on the casual mechanical aspects of explanations has had some success in response to the limitations detected within the covering-law model. These include issues mentioned above such as the asymmetry and directionality problems, which can be overcome by appealing to causes. Second, the proponents of causal theories have analysed several case studies taken from real-life explanations, instead of just using abstract examples or toy examples. In the same vein, they have also incorporated case studies from life and social sciences, avoiding the tendency to focus only in physics and forget about the whole spectrum of other sciences. Finally, the interest in casual theories of explanation has also something to do with the fact that many paradigmatic explanations actually rely on information about causes and mechanisms.

As has been pointed out at the outline of the present chapter, nowadays, the consensus there was about scientific explanation being a matter of providing information about the causes of the *explanandum* can be thought of as broken. There has been a shift of attention to non-causal approaches to explanation, with the idea that a big part of explanations in the sciences and in ordinary life cannot be accounted for in terms of causal explanations. Mathematics has been one of the key "suspects" for not fitting into a causal account. The motivation of this shift is that there are many compelling, real-life examples of non-causal explanations that causal accounts of explanation seemingly fail to capture. As we have seen, the natural sciences, physics in particular, are a good source for such examples, which range from explanations involving symmetries and inter-theoretic explanations, to more abstract explanations that rely on renormalizations group techniques. There are also examples of non-causal explanations in ordinary life and social sciences, such as mathematical, statistical, computational, network explanations and so on.

After the unavoidable fact that there is more to explanation than causation, instead of speaking of causal theories of explanation, we should rather be speaking of theories of causal explanation and, even so, we should bear in mind that they will not tell the full story. There is already a vast literature on non-causal ways of explaining, among which we can find unificationist accounts (Friedman 1974; Kitcher 1984, 1989), pragmatic accounts (Van Fraassen 1980; Achinstein 1983), analyses of asymptotic explanations in physics (Batterman 2000, 2002), statistical and geometrical explanations (Lipton 1991/2004; Nerlich 1979) and many others. In this context, the question of whether there can be a unified theory of explanation (that is, a theory that accounts for *all* explanations) constitutes a relevant and, potentially, fruitful field of research.

Although there are some reductionist approaches, that is, approaches to mathematical and scientific explanation that argue that all explanations can ultimately be accounted for within a causal approach, according to Reutlinger (2017), there are several reasons to prefer a monist account (see Section 1 of the present chapter). In the present work, the counterfactual theory of explanation will be analysed as a possible candidate to constitute a monist account of mathematical explanation.

6. More recent approaches to mathematical explanation: explanatory virtues and general theories

As we have seen, the rejection of the Indispensability Argument in its original formulations led platonists in the last decades to rethink their approach and so they began to focus on the *explanatory indispensability* of mathematics rather than the mere indispensability of the reference to mathematical (and, in a broad sense, abstract) objects.

Apart from the Deductive-Nomological model (by Hempel) or the Causal-Mechanistic Model (by Salmon), there are some newer approaches that are currently gaining much interest among researchers, such as the Counterfactual Theory of Explanation (CTE), now presented in new versions adapted to the case of mathematical explanation. Besides the CTE, we can find other attempts to account for mathematical explanation, such as Baron's (2017) Deductive-Mathematical account, or even authors who prefer to appeal to explanatory virtues of particular explanations instead of appealing to a general theory of explanation. This last one is the strategy followed by authors like Baker and Colyvan.

Let us review these two possible attitudes before this issue: (1) that of accounting for mathematical explanations in terms of explanatory virtues, and (2) that of providing a general theory of explanation to account for the case.

6.1. Explanatory virtues

Before the difficulty of bringing up a general theory of explanation that overcomes all the problems that have been raised against the *DN* and the *CM* models, some authors seem to have given up on that idea. Instead, they have been directing their interest to independent explanatory virtues that explanations can show in variable degrees.

Firstly, it is worth noting that Michael Keas, in a very recent paper titled "Systematizing the theoretical virtues" (2017) has done an impressive job in providing a detailed picture of the explanatory virtues that function in our widely accepted theories across many disciplines, namely in the natural sciences. In his paper, he also mentions previous attempts to understand and systematize the theoretical virtues, by authors such as Kuhn, Laudan, Douglas or McMullin (see Keas 2017, 2763-2764).

Keas defines "explanatory virtues" as "the traits of a theory that show it is probably true or worth accepting" (Keas 2017, 2761). He classifies them into four groups, each of which contains at least three virtues that sequentially follow a repeating pattern of progressive disclosure and expansion (see Keas 2017, 2762-2763).

Here is Keas's picture of the theoretical virtues in form of a list of a total of twelve explanatory virtues, divided into four categories, and much summarized definitions of each one of them[17].

---

[17] See Keas 2017, 2765- 2787 for details.

Evidential virtues:

They indicate different facets of how well a theory accounts for the entities, events and regularities of the world.

- Evidential accuracy (TV1). It is instantiated in a theory when it fits the empirical evidence well.

- Causal adequacy (TV2). It is instantiated when a theory specifies causal factors that plausibly produce the effects in need of explanation.

- Explanatory depth (TV3). A theory exhibits explanatory depth when it excels in causal history depth (for the CM model) or in the range of counterfactual questions that its law-like generalizations answer regarding the item being explained (in the case of the counterfactual theory of explanation). This second sense refers to the number of what-if-things-had-been-different questions an explanation can answer.

Coherential virtues:

- Internal consistency (TV4). This virtue is exhibited in a theory when its components are not contradictory. It is, therefore, a logical property.

- Internal coherence (TV5). We say that a theory possesses internal coherence when its components are coordinated into an intuitively plausible whole. This kind of coherence is more extensive and subtler than the logical principles of internal consistency.

- Universal coherence (TV6). It is present if a theory sits well with, or is not obviously contrary to, other warranted beliefs.

Aesthetic theoretical virtues:

- Beauty (TV7). A beautiful theory evokes aesthetic pleasure in properly functioning and sufficiently informed persons. Some degree of cultural and individual variation of aesthetic experience is contemplated. Among the factors that trigger the experience of beauty Keas points at symmetry, aptness and surprising inevitability.

- Simplicity (TV8). A simple theory explains the same facts as rival theories, but with less theoretical content.

- Unification (TV9). A unified theory explains more kinds of facts than rival theories with the same amount of theoretical content.

Diachronic theoretical virtues:

These theoretical virtues can only be instantiated as a theory is cultivated after its origin, that is, they present an extended temporal dimension.

- Durability (TV10). A theory exhibits durability if it has survived testing by successful prediction or by plausible accommodation of new unanticipated data.

- Fruitfulness (TV11). A theory is fruitful if it generates additional discovery by means such as successful novel prediction, unification, etc. This property is sometimes referred to as fertility or fecundity.

- Applicability (TV12). A theory is applicable when it is used to guide successful action (e.g., prepare for a natural disaster) or to enhance technological control (e.g., genetic engineering).

There are several aspects to note with regard to this classification of explanatory virtues.

First, here explanatory depth (TV3) has a clear meaning and definition –even though it can acquire several specifications – whereas it is quite common to find the concept used in very vague or unspecified senses, which can lead to confusion. Indeed, sometimes “explanatory depth” equals explanatory power or explanatoriness in general. As used in Keas’s approach, it seems that “explanatory depth” is equivalent to scope generality (in Ylikoski and Kuorikoski’s 2010 terminology[18]).

Second, with regard to beauty (TV10), there is some new research questioning the aesthetic value of scientific and mathematical theories. It is worth going through Giaquinto’s (2016) paper titled “Mathematical Proofs: The Beautiful and The Explanatory”[19].

In addition, Lange (2016) compares explanatory virtues found in mathematical proofs and argues that beauty and explanatoriness need not appear together in a mathematical explanation, that is, there are features of a proof that can contribute to its explanatory power and also to its aesthetic value, but a beautiful proof need not be explanatory.

Third, even though I do not intend to question the value of having a systematization of explanatory virtues, I have some concerns with this classification. For instance, simplicity and unification should probably not be counted as aesthetic virtues, in the sense that a theory being simpler or more unified seems to be something objective that does not fit into an aesthetic category. As much as I agree with the idea that philosophers and scientist often express their engagement with theories appealing to their aesthetic value, that is, a value related to the beauty of an explanation, the epistemic virtues of simplicity and unification can be accounted for in non-aesthetic and non-subjective terms, as shown below in the definitions of the most common explanatory virtues found in the context of the EIA debate.

Specifically, Keas treats simplicity and unification as “specific complementary aesthetic theoretical virtues” and “special cases of beauty” (see Keas 2017, 2775). However, what can be found in the literature is that they are often treated as the central and most important explanatory virtues for the case of mathematics. That is, there is reason to think that mathematics’ role in explanation is, at least partly, that of presenting information in simpler and more unified ways, given that it contributes to a better understanding of the phenomenon in question. Besides, arguing that those are aesthetic virtues on the grounds that we prefer to live in a simpler and more comprehensible world does not suffice, since that is an argument that could be applied to any other explanatory virtue and, following this path, every single explanatory virtue would be aesthetic just because it increases explanatoriness. In this sense, I am even tempted to state that simplicity and unification are no more “aesthetic” than any other explanatory virtue and I would rather treat them as useful goals to go for, than necessarily aesthetic ones. In fact, the aesthetic side of the properties probably stems from their increasing explanatoriness. However, at least as far as the debates about explanation and mathematical ontology are concerned, it is not particularly relevant whether we treat simplicity and unification as aesthetic or explanatory virtues, since both *sides* of these properties would work in the same direction.

Independently from these comments, Keas’s paper does show that we sometimes use explanatory virtues more or less vaguely in order to account for what is going on in mathematical explanations and explanations of all sorts.

A classification and definition of explanatory virtues can also be found in Ylikoski and Kuorikoski‘s (2010) presentation of the counterfactual theory of explanation, and those were later on used by Knowles and Saatsi’s own version of the theory (2019), as we shall see in Chapter 2. Ylikoski and Kuorikoski point at five different explanatory virtues –or dimensions of explanatory power – that explanations can show in degrees: non-sensitivity, precision, factual accuracy, degree of integration and cognitive salience (see Ylikoski and Kuorikoski 2010, 208-215). Let us briefly review the meaning of these five virtues.

---

[18] See below.

[19] See Chapter 4 for a presentation of Giaquinto’s views on mathematics and aesthetics.

1. **Non-sensitivity** refers to the explanatory relationship with respect to changes in background conditions. The more sensitive the explanation is to changes in background features, the less powerful it is, which means that the same answer would be correct for a smaller number of what-if questions.
2. **Precision** is an attribute of the *explanandum*, related to how precisely the explanation characterizes the *explanandum* phenomenon. The more detailed the explanation, the better.
3. **Factual accuracy** refers to the apparent consensus that explanation is factive, and, thus, false explanations are not really explanations.
4. The **degree of integration** relates to unification or connectedness to a larger theoretical framework: "there is a sense in which an explanation that is integrated into a larger body of knowledge contributes to explanatory understanding more than do these bodies of knowledge considered separately" (see Ylikoski and Kuorikoski 2010, 213-214). This expands the range of answers to different *what-if questions* in two ways. First, because of the inferential connections to an already existing body of knowledge, dependencies between factors in the background theory and different aspects of the *explanandum* phenomenon may open up unforeseen dimensions in which contrastive what if-questions concerning the *explanandum* can be answered. Second, the explanation itself may bridge previous gaps within the existing theory and thus enable answers to new what if-questions not directly concerning the original *explanandum* phenomenon (see Ylikoski and Kuorikoski 2010, 213-214).
5. **Cognitive salience** refers to the ease with which we grasp a given explanation, how easily the implications of the reasoning can be seen and how easy it is to evaluate the scope and limitations of the explanation. Ylikoski and Kuorikoski (see 2010, 214) argue that it would be a mistake to dismiss this property on the basis of it being pragmatic, arguing that "[s]ome of the factors underlying cognitive salience are species-relative, and some of them are highly personal, but many are related to disciplinary traditions and training" (Ylikoski and Kuorikoski 2010, 214).

There is no doubt that both pictures of explanatory virtues are clearly different, but so are their purposes in systematizing those virtues. Keas attempts to provide a global systematization of the theoretical virtues that would be applicable to all main accounts of explanation, whereas Ylikoski and Kuorikoski's aim is to account for what constitutes a good explanation and the concept of explanation at use in the context of the counterfactual theory of explanation. However, expectedly, both approaches point at some common features of explanation, such as the importance of the sharpness of an explanation, its applicability or how it fits into the set of accepted beliefs.

More specifically, (at least) as far as the literature on the EIA debate for the case of mathematical explanation is concerned, I found that there are fundamentally *four explanatory virtues* that are repeated in several accounts of explanation. These refer to the particular case of mathematical explanation and its role in accounting for empirical phenomena. It is worth highlighting and defining them:

- **Scope generality.** It is equivalent to non-sensitivity. Taking a toy example, scope generality is the kind of generality that allows us to explain the fact that we cannot divide 23 cherries evenly among five people, and also the fact that we cannot divide them evenly among six people by appealing to the same principle (division, in this case).

- **Topic generality.** It is connected with unification. Using the toy example above, topic generality is the kind of generality that allows us to use the same principle to explain the impossibility to divide 23 cherries evenly among five people, but also pens, books, etc. Baker (2017) argues that topic generality allows for unificatory power.
- **Simplicity**. As an explanatory virtue, simplicity is probably referenced in every account of mathematical explanation. It is thought to be one of the central advantages of using mathematics in explanations of phenomena. However, its meaning is not that clear, for it can be taken two ways. On one hand, simplicity can refer to the number of premises an argument presents, where the smaller the number of premises, the better the argument. On the other hand, in a more epistemic sense, simplicity could refer to the ease with which we can follow an argument, where the easier it is to follow, the better the explanation. This second sense of simplicity collapses with “cognitive salience” in Ylikoski and Kuorikoski’s account.
- **Explanatory depth.** This is probably the vaguest explanatory virtue in the literature. In Baker’s papers (2017) it refers to the explanation providing the *why* of the *explanandum* in some metaphysical (ontologically committing) sense. However, this would not be different from *explanatoriness* or *explanatory power* itself. It is not clear, then, that it constitutes an explanatory virtue by itself, but it is worth mentioning because it repeatedly appears in the literature around the EIA debate. As can be seen above, Maddy (2011) also refers to mathematical depth as one of the features of mathematical explanations.

Hopefully, this analysis of explanatory virtues will help get a clearer picture of the kind of role mathematics plays in scientific and even in ordinary-life explanations. It is also worth noting that many authors who provide formal or systematized theories of explanation in order to account for the mathematical case also refer to these explanatory virtues when characterizing mathematics’ role in explanations.

7. Conclusions

The aim of this chapter was to place the debate around mathematical explanation and mathematical ontology by providing an introduction to the state of the question, along with some comments on the different initial approaches that we can find in the literature.

With this purpose, we began by introducing the debate about mathematics' role in explanations with Reutlinger's illuminating analysis. Then, the original (Quine-Putnam) Indispensability Argument was introduced, and some objections to it, with Field on the most radical side and some other responses that did not demand a nominalization strategy. With respect to Field's program, the (near) consensus is that his requirements are too strong it is doubtful that we can reformulate every single mathematical statement in explanations of empirical phenomena in order to obtain a nominalization.

This led to the presentation of a selection of Easy-Road Nominalist approaches, namely, those proposed by Balaguer, Leng and Melia, which showed the feature that they all have in common: the idea that mathematics need not be true to be useful in representing physical phenomena.

The original indispensability argument for mathematics is currently discarded by philosophers and researchers. Platonists, such as Baker and Colyvan, have changed their strategy by arguing that mathematics does more than just representing phenomena, and that there is an indispensable explanatory role of mathematics that can support the platonist ontological thesis. This is what we know as the Enhanced Indispensability Argument, which is now the real challenge for the mathematical nominalist or moderate platonist.

The focus on explanation directed the attention to theories of explanation, so in Section 5 an account of the Deductive-Nomological Model and the Causal-Mechanistic Model are provided, as classical accounts of explanation in philosophy of science, together with some serious problems they present.

Continuing with the research on mathematical explanation, the present chapter ends with a presentation of the explanatory virtues that are commonly attributed to mathematical explanation –and explanation in general. These will help guide our research of the issue of what constitutes a good mathematical explanation and mathematics' role in science.

From here, Chapter 2 will address the EIA debate in its most recent terms. It remains an open question whether there can be a successful monist account of mathematical explanation. One of the most popular current proposals along this line is the counterfactual theory of explanation (CTE), but we can also find the Deductive-Mathematical theory of explanation (by Baron) and some others. The next chapter will provide a deeper analysis of these.

CHAPTER 2: In search of a monist account of mathematical explanation in the context of the EIA debate

1. What are we discussing? A brief introduction

As we have seen in Chapter 1, (easy-road) nominalists who faced the Indispensability Argument have attempted to provide an approach to mathematics that does not carry a commitment to the existence of mathematical entities, in order to avoid the need to accept a platonist picture of abstract objects. Their general strategy, with variations, is to argue that mathematics has a merely representational role in our scientific theories, and that this representational role does not support by itself the thesis that the platonist approach to mathematical entities is correct. In response, platonists have argued that mathematics does more than just represent, in that it *explains* certain scientific results in an indispensable way (that is, we would not have an explanation of those phenomena without referring to mathematical entities). This has led to the EIA debate.

The EIA (Enhanced Indispensability Argument) debate arises as a new and more *advanced* form of the traditional indispensability argument debate around whether the alleged indispensability of mathematics constitutes good grounds for supporting mathematical platonism or not. In recent years, the debate has switched to this framework. The assumption in the EIA is that mathematical structures are not just representational, so there is *something more* that mathematics is doing that needs to be examined. That *something more* has been referred to as the *substantial* or *genuine* explanatory role of mathematics.

In this sense, platonists argue that mathematics goes beyond expressing or representing phenomena and its role is *genuinely* or *distinctively explanatory*, in their attempt to avoid nominalists' criticism. On the other side of the debate, nominalists need to account for this *extra* that mathematics provides in explanations in a way that does not carry a commitment to the existence of mathematical entities in order for their argument to succeed.

This is the context in which the arguments –and, sometimes, even dialectic *dialogues* – take place among authors such as Baker (2017), Saatsi (2011), Bueno and Colyvan (2011), Baron (2016a; 2017), Liggins (2016) and Knowles and Saatsi (2019), among others. Their terminologies differ as to what corresponds to the mentioned substantive role of mathematical explanation: while Baker uses terms such as "explanatory generality" that is specific to mathematics, Knowles and Saatsi (2019) refer to this kind of generality as 'distinctive' mathematical explanation and, for instance, Baron writes in terms of 'genuine' mathematical explanation.

Therefore, we find at least two assumptions that all of these authors share in the debate. First, there are cases in which there is some kind of substantial or genuine mathematical involvement in explanations of empirical phenomena, which is decisive in relation to whether platonism or nominalism is right and also in relation to the terms in which the debate should advance. Second, the debate is currently at an *impasse*, since no account seems to provide a breakthrough in the aim of making some substantial claim that solves – or comes closer to a solution of – the issue. All these authors (and many more) suggest strategies to overcome said *impasse* by providing new and different arguments.

Even though our interest in this work is to examine the explanatory role of mathematics in a rather broad sense, it is clear from the state of the question we have presented in Chapter 1 and, very briefly, in the present introduction, that any study concerning mathematical explanation needs to address the EIA debate. In this context, it also becomes necessary to address the relation between the question of the explanatory role of mathematics and the question of whether mathematical objects exist.

The plan

Bearing this in mind, this chapter will devote Section 2 to a quite detailed review of the EIA debate, including a discussion of the newest proposals regarding the EIA debate. With the aim to evaluate if there is an account that succeeds at overcoming the *impasse* in the debate, Section 3 will provide an analysis of the Counterfactual Theory of Explanation as a good (in principle) candidate to account for the explanatory role of mathematics. Some other alternative accounts will be suggested, such as Baron’s approach (The Deductive-Mathematical Theory of explanation) in Section 4 and Lange’s pluralist approach in Section 5. Finally, some conclusions will be put forward as to which ontological picture seems more plausible, which will constitute Section 7.

2. The Enhanced Indispensability Argument

As we have seen in Chapter 1, the on-going debate about mathematical ontology was, at the beginning, in terms of whether mathematics plays an indispensable role in our scientific theories and, that being the case, whether that carries an ontological commitment with mathematical entities. However, after the development of so many easier roads to nominalism and the criticism of the classical Indispensability Argument, the debate as it was originally formulated lost its predominance.

Baker (2009a, 613) rightly points out that the debate between Melia (2000, 2002) and Colyvan (2002), despite their opposed views (Melia with a nominalist approach and Colyvan with a platonist one), has led to some kind of a breakthrough. That breakthrough is the acknowledgement that, in order for the indispensability argument to succeed, something more needs to be said regarding the role that mathematics plays in science. That is, it is right that mathematics needs to be indispensable (in order for the platonist to win the debate), but it needs to be so *in the right way*. The result is that what needs to be shown is that the reference to mathematical objects plays an *explanatorily* indispensable role in science.

Hence, authors like Colyvan or Baker changed their strategy. As a result of this shift of focus, instead of arguing for the indispensability of mathematical entities, they argue for the indispensable explanatory role of mathematics in explanations of physical phenomena. This is what we know as *Enhanced Indispensability Argument* (EIA).

The EIA is a special kind of an inference to the best explanation. As we will see, the first premise adds that “explanation” component, restricting our attention to those cases where mathematics is playing an indispensable explanatory role, and thus appealing to an inference to the best explanation to conclude that we should believe in the existence of mathematical entities.

So, the EIA presented by Baker (2005, 2009a), in contrast to the Quine-Putnam IA, does not include conformational holism but, rather, it relies on mathematics’ explanatory role in scientific explanations. In order to analyse the EIA in more detail, let us take one of its possible formulations. Here is the argument by Baker (2009a, 613):

> **The Enhanced Indispensability Argument**
>
> (1) We ought rationally to believe in the existence of any entity that plays an indispensable explanatory role in our best scientific theories.
>
> (2) Mathematical objects play an indispensable explanatory role in science.
>
> (3) Hence, we ought rationally to believe in the existence of mathematical objects.

The main difference with respect to the original argument based on Quine’s work is that it aims at getting the justification for our mathematical beliefs in scientific explanations relying directly on mathematics, rather than in scientific theories in general. The focus on explanation is clear, and it highlights the importance of studying mathematical explanation and the theories that have arisen to account for it.

However, as potential problems, we can point out that the argument leaves open the question of how to determine the commitments of an explanation. The second premise does not specify how we determine which are the posits of an explanation. If we set the criterion in the entities posited by our best scientific theories, we run the risk of ending up back in the classical Quinean indispensability argument. If that were not the case, then an account of the commitments of mathematical explanations and the criteria to identify them would need to be provided.

2.1. The EIA and the *cicada* case

In order to illustrate the argument in its EIA version and make his case, Baker introduces the well-known example of the periodical cicada. Let us review the case before we get into Baker's several arguments within the EIA strategy.

The cicada case, presented in the outline of this work, has become a paradigmatic case of extra-mathematical explanation, where the life cycle of the periodic cicada is explained by a mathematical fact, that is, the fact that periodic cycles minimize the intersection with potential predators. The case has been studied by Baker (2005, 2009a, 2017 and more) and used to defend a platonist approach to mathematical entities, that is, the claim that we ought to commit to the existence of mathematical objects. It has since been one of the most discussed cases in de debate around mathematical explanation.

The evolutionary strategies of the cicada species, both the fact that they emerge every 13 or 17 years and predator satiation[20], give rise to a quite natural question for everyone that encounters this case for the first time: *why prime numbers?* As mentioned in the introduction, this has something to do with the biological fact that it is an evolutionary advantage to minimize the intersection with predator species or even other cicada species. That is, part of the explanation biologists give for the case is that the cicadas have evolved to avoid overlapping with other periodical creatures. The mathematical component supports this thesis and complements the biological claim, since prime periods actually minimize intersection. Now, can a theorem of arithmetic explain a biological phenomenon? And, perhaps more importantly, what makes the arithmetical explanation a good one?

There is a sense in which mathematics and, in particular, the concept of primeness explain what is going on with the cicada case. As we will see, the controversy concerning the role of mathematics in this example led to a great deal of literature on mathematical explanation and its alleged indispensability in explaining the case.

2.2. Baker's series of *enhanced indispensability arguments*

Since the paper from 2005, Baker has developed several *enhanced* arguments for the explanatory indispensability of mathematics (and, thus, the correctness of mathematical platonism), some of them in a direct debate with Saatsi (2011)[21]. With this in mind, let us go through his different presentations of the arguments.

---

[20] "Predator satiation" is an evolutionary strategy according to which each time the cicadas emerge, they do it in very large numbers so the potential predators cannot exhaust the species (see Gould 1977, 101).

[21] It is relevant to note that in "Mathematical Spandrels" (2017) Baker aims at opening a new front in the debate by introducing the term 'mathematical spandrel' to describe penumbral properties, which would show that extra-mathematical explanations involve mathematics. In particular, taking the cicada case, Baker argues that "if there is a collection of *n* concrete objects, and if *n* has the property of being the sum of two perfect squares (S2S), then this has non-mathematical consequences" (Baker 2017b, 7). However, he does not clarify what he means by "non-mathematical consequences". In terms of Bueno and Colyvan (2011), we have no relevant interpretation of the property as Baker presents it, so there is no way to determine whether the

Below is the main sketch of the argument's application to the cicada case as presented by Baker (2009a, 614, emphasis in original)[22]:

> **The Cicada Example**
>
> (4) Having a life-cycle period that minimizes intersection with other (nearby/lower) periods is evolutionarily advantageous. (*biological law*)
>
> (5) Prime periods minimize intersection (compared to non-prime periods). (*number theoretic theorem*)
>
> (6) Hence organisms with periodic life cycles are likely to evolve periods that are prime. (*'mixed' biological/mathematical law*)
>
> (7) Cicadas in ecosystem-type E are limited by biological constraints to periods from 14 to 18 years. (*ecological constraint*)
>
> (8) Hence cicadas in ecosystem-type E are likely to evolve 17-year periods.

The main thesis here is that this case is an example of "an *indispensable, mathematical explanation* of a *purely physical phenomenon*. Hence, applying the Enhanced Indispensability Argument, we ought rationally to believe in the existence of abstract mathematical objects" (Baker 2009a, 614, emphasis in original). This presentation of the argument reappears in "Mathematics and Explanatory Generality" (2017) with the name "MES" ("Mathematical Explanation in Science").

Baker opposes the picture where mathematics only plays a purely representational role (the view named "explanation through representation" in Knowles and Saatsi's work)[23], where physical facts being represented by mathematics is the only feature that carries explanatory power. Instead, he argues for a *genuine* explanatory role of mathematics, where this explanatoriness goes beyond mathematics' power to represent phenomena. In a way, Baker argues that what mathematics is doing is picking out structural facts in the explanation that are not (and go beyond) the physical facts of the explanation.

With that in mind, Baker reformulates his previous argument of 2009a, arguing that in the original presentation the premise (5), "prime periods minimize intersection" was misleadingly labelled as a 'number-theoretic theorem', because the notions of period and intersection are not number-theoretic *per se*, so it is more accurate to think of that premise as a 'mixed' claim with both mathematical apparatus and facts about temporal duration (see Baker 2017, 196).

In his 2017 paper, Baker compares his version of the argument for the cicada case with Saatsi's (2011) nominalistic version of the argument ("NES" in Baker's paper), where the only mathematical concepts are specific numbers and there are no references to primality, thus removing the ineliminable mathematical content of the original argument. The main problem with this argument, as Baker sees it, is that it lacks the scope generality that was packed into the original formulation. That is, it lacks enough explanatory power to explain cases where the life-cycle duration is not 13 or 17 years. This would have the disadvantage of requiring a new

---

property in question has the consequences Baker argues. Besides, with a suitable interpretation, the nominalist could provide a non-mathematical version of the case. Therefore, his argument based on "mathematical spandrels" does not look as strong as he puts it to be and, overall, his argument does not change much with respect to other articles of his, analysed with more detail in the present work.

[22] The numeration continues from the basic argument that can be seen above.

[23] As we will see, Knowles and Saatsi (2019) also argue that mathematics' role cannot be only representational, so their view is not committed to the idea that mathematics explains through representation. Also, they argue that this principle (or premise) is shared by most previous accounts of mathematical explanation in the EIA debate, both platonist and nominalist.

argument for each different case. Let us examine both versions of the argument to place the discussion[24].

"MES" (Mathematical Explanation in Science) is the platonist version presented by Baker, whereas "NES" is the nominalistic adaptation. Both are applied to the cicada case, starting with premise number (4), to then compare these versions of the argument with the (superior) versions he calls 'Cicada MES_{GEN}' and 'Skid Patch MES_{GEN}'[25], which, Baker argues, increase generality in a substantive way.

MES[26]

(4) Having a life-cycle period which minimizes intersection with other (nearby/lower) periods is evolutionarily advantageous. [biological law]

(5) Prime periods minimize intersection (compared to non-prime periods). [number theoretic theorem]

(6) Hence organisms with periodic life-cycles are likely to evolve periods that are prime. ['mixed' biological/mathematical law]

(7) Cicadas in ecosystem-type E are limited by biological constraints to periods from 14 to 18 years. [ecological constraint]

(8) Hence cicadas in ecosystem-type E are likely to evolve 17-year periods.

NES*[27]

(4) Having a life-cycle period which minimizes intersection with other (nearby/lower) periods is evolutionarily advantageous. [biological law]

(5/6)** There is a unique intersection minimizing period T_x for periods in the range $[T_1, ..., T_2]$ years. [fact(?) about time]

(7)* Cicadas in ecosystem-type E are limited by biological constraints to periods in the range $[T_1, ..., T_2]$ years. [ecological constraint]

(8)* Hence cicadas in ecosystem-type E are likely to evolve T_x-year periods.

As we can see, the key modification NES* introduces is premise (5/6)**, a general argument pattern. This reformulation substitutes the problematic mathematical premise with a schema: "(5/6)** There is a unique intersection minimizing period Tx for periods in the range [T1, . . . , T2] years. [fact(?) about time]". The key feature of this modification is that there is an instantiation of (5/6) for any given range [T1, . . . , T2] of cicada life cycles. If the instantiation of the schema is true, that is, if there is indeed a unique intersection minimizing period, then it can be used to explain the life-cycle length of the given cicada subspecies. Baker concedes that this version achieves the desired level of scope generality, in that it is applicable for every particular life-cycle length. However, he objects that it provides a less unified explanation and has less explanatory depth than the standard cicada argument –that is, MES. In Baker's account, the fact that NES* provides less explanatory depth is not unimportant, since explanatory depth refers to the *why* of the *explanandum*, and thus what Baker is arguing is that NES* leaves unexplained why this fact about time holds for some ranges of duration and not for others.

As for the concept of explanation at use in Baker's account, mathematics explains by providing very general information that cannot be understood in terms other than involving

---

[24] With "the discussion" I am fundamentally referring to Baker (2009a, 2017) and Saatsi (2011).

[25] See below.

[26] Baker 2017, 195-198.

[27] Baker 2017, 198.

mathematical facts about empirical entities (MES). These mathematical explanations carry information about features of the world. This makes mathematical statements indispensable to explanations of scientific phenomena and, in conclusion, this indispensability carries ontological commitment regarding the status of mathematical entities.

2.3. Explanatory virtues by Baker: topic and scope generality

Before we proceed to the version of the argument that achieves the highest level of generality, it is worth stopping to recall that Baker points at the explanatory virtues of topic generality and scope generality to account for the explanatory role of mathematics (in extra-mathematical explanation)[28]. Since Baker links his analysis of the EIA with these virtues as two different senses in which mathematics achieves the highest level of generality in explaining phenomena, let us see their application to the cicada case. As we shall see, Knowles and Saatsi (2019) also refer to these two senses of generality in their aim to refute Baker's argument.

- **Scope generality**

Each of the specific durations of life-cycles can be easily expressed without quantifying over mathematical entities by using first-order logic with identity. However, this is not enough to pick out the entire set of durations without using mathematical concepts and mapping durations onto the infinite set of natural numbers. Here, mathematics provides some kind of generality which, it is argued, is more useful for expressing facts about durations and intersections. This is what Baker and Knowles and Saatsi calls *scope generality*.

Mathematics allows us to express facts about durations and intersection minimization that are unlimited in their scope (see Baker 2017, 197). It has to do with abstracting from the actual physical structure of the entities involved or the actual laws of nature that underlie the physical processes (see Jansson and Saatsi 2017). This corresponds to the first dimension of explanatory abstraction, the one Jansson and Saatsi (2017, 4) characterize as "a matter of relative independence of the *explanans* from the actual physical structure of the entities involved". Hitchcock and Woodward (2003, 181) refer to this as "generality with respect to objects and systems other than the one that is the focus of explanation". This is the kind of generality emphasized by representational nominalists, since they contend that scope generality can be guaranteed without using mathematics.

This issue, according to a certainly simplistic distinction by Baker, has prompted three kinds of response from the representational nominalist (see Baker 2017, 197-200).

The first one is to argue that scope generality is not required in order to explain the cicada life-cycle phenomenon, so the mathematical apparatus is dispensable.

A second response is to concede the value of scope generality but to argue that the desired level of scope generality can be achieved in a nominalistic framework, so the mathematical apparatus is ultimately dispensable.

A third way to go, according to Baker, is *easy-road nominalism*, which concedes that the mathematical apparatus here is indispensable to the explanation of cicada life cycles, but maintains that it is solely the physical facts about durations that carry the explanatory load. This 'road' has the advantage that it frees the nominalist from the burden of providing a mathematics-free alternative explanation (see Baker 2017, 199-200).

---

[28] For a discussion of explanatory virtues, see Chapter 1, Section 6.

Of course, as we have seen in Chapter 1, there are many different approaches within the Easy-Road when it comes to explanatory uses of mathematics and this is nothing more than a simplification made by Baker in order to classify the different stages that nominalists seem to have gone through in relation to the indispensability argument issue and the evolution of the debate on this subject.

- **Topic generality**

Topic generality refers to the fact that mathematical explanations achieve a level of generality that allows their application to different topics, that is, the explanation is valid for a large range of areas or subject matters. It has to do with abstracting from the properties of objects. Topic generality is essential to the substantial mathematical involvement in explanation. As Baker characterizes it (see 2017, 200-201), *topic generality* is orthogonal to scope generality. In this context, he argues for the following four theses[29]:

(1) The optimal version of the Cicada MES has an explanatory core that is topic general, so it is not 'about' any designated class of physical facts, such as facts about time, or facts about durations.

(2) The topic generality exhibited by the Cicada MES is of scientific value. To support this thesis, he argues that the explanatory core of the Cicada MES is shared by other explanations of non-temporal, non-biological phenomena.

(3) This level of topic generality cannot be matched without using mathematical apparatus, hence the mathematics is indispensable.

(4) There are no specific physical facts in the core explanation for the mathematics to pick out, which indicates that the role of the mathematics in the Cicada MES is explanatory, and not 'merely representational'.

With this in mind, Baker formulates a new argument for the cicada case, but using a generalization that renders the core mathematical component non-temporal, abstracting from the specific case and using the more general pattern of unit cycles (UCs). He argues that the new version is the best explanation of what is going on with the periodical cicada life cycles, with the idea that, by achieving such a high level of generality, it can be used to account for explanations from different topics using the same explanatory core from the cicada case. This general argument is the MES_{GEN}.

2.4. The most general argument: unit cycles MES_{GEN}

In the transition from MES to Cicada MES_{GEN}, Baker departs from picking out specific facts about time periods and uses the more general notion of 'unit cycles'. His motivation is that the traditional version of the argument, MES, encourages representational nominalism by masking the genuine explanatory contribution made by the mathematical facts. According to "representational nominalism", as Baker characterizes it, "mathematical concepts (in this case the concept of primality) are merely used to pick out aspects of the physical situation, and it is these physical facts that are doing all the explanatory work" (Baker 2017, 196).

Cicada MES_{GEN} intends to show that mathematics goes beyond representation in genuine mathematical explanations of empirical phenomena. It goes as follows:

---

[29] Baker 2017, 201.

Cicada MES_{GEN}[30]

(M_1) The lowest common multiple (LCM) of two numbers, m, n, is maximal if and only if m and n are coprime. [pure mathematical fact]

(UC_1) The gap between successive co-occurrences of the same pair of cycle elements of two unit cycles is equal to the LCM of their respective lengths.

(UC_2) Hence any pair of unit cycles with periods m and n maximizes the gap between successive co-occurrences of the same pair of cycle elements if and only if m and n are coprime.

(M_2) All and only prime numbers are coprime with all smaller numbers.

(UC_3) Hence, given a unit cycle, pm, of length m and a range of unit cycles, qi, of lengths shorter than m, pm maximizes the gap between successive intersections with each qi if and only if m is prime. (from UC_2, M_2).

(1) For periodical organisms, having a life-cycle period that maximizes the gap between successive co-occurrences with periodical predators is evolutionarily advantageous. [biological law]

(2) Periodical organisms with periodical predators whose life cycles are restricted to multiples of a common base unit can be modelled as pairs of unit cycles.

(3) Hence organisms with periodic life cycles that are exposed to periodic predators with shorter life cycles, and whose life cycles are restricted to multiples of a common base unit, are likely to evolve periods that are prime. [‘mixed’ biological/mathematical law] (from 1, 2, UC_3)

(4) North American periodical cicadas fit the application conditions stated in premise (3).

(5) Hence periodical cicadas are likely to evolve periods that are prime. (from 3, 4)

(6) Cicadas in ecosystem-type E are limited by biological constraints to periods from 14 to 18 years. [ecological constraint]

(7) 17 is the only prime number between 14 and 18. [pure mathematical fact]

(8) Hence cicadas in ecosystem-type E are likely to evolve 17-year periods. (from 5, 6, 7)

As we can see, from M_1 to UC_3, there has been no reference to the particular case of the cicadas’ life cycles. What follows is precisely the application of this argument to the cicada case. The strategy is completed when, afterwards, Baker presents another argument, named the Skid Patch MES_{GEN}. This new argument shares the explanatory core (the first five premises) with the Cicada MES_{GEN}, and yet it is an explanation of a different phenomenon in a different area. The Skid Patch MES_{GEN} refers to the rear wheel of brakeless fixed-gear bicycles (BFGBs) and gear ratios. It goes as follows:

The Skid Patch MES_{GEN}[31]

(M_1) The lowest common multiple (LCM) of two numbers, m, n, is maximal if and only if m and n are coprime. [pure mathematical fact]

(UC_1) The gap between successive co-occurrences of the same pair of cycle elements of two unit cycles is equal to the LCM of their respective lengths.

(UC_2) Hence any pair of unit cycles with periods m and n maximizes the gap between successive co-occurrences of the same pair of cycle elements if and only if m and n are coprime.

---

[30] Baker 2017, 8-9.

[31] Baker 2017, 204.

> (1’) For brakeless fixed-gear bicycles, having a gear set-up that maximizes the gap between successive co-occurrences of the braking position of the front cog and a given position of the rear cog tends to make the rear tire last longer. [mechanical law]
>
> (2’) The positions of the front cog and the rear cog of a bicycle that are linked by a fixed chain can be modelled as a pair of unit cycles whose lengths correspond to the number of teeth on each cog.
>
> (3’) Hence brakeless fixed-gear bicycles whose front-cog and rear-cog tooth numbers are coprime will tend to have rear tires that last longer. (from 1’, 2’, UC$_2$)
>
> (6’) A particular group of fixed-gear bicycles have 14-tooth rear cogs and either 46-tooth, 47-tooth, 48-tooth, or 49-tooth front cogs.
>
> (7’) Of the pairs (14, 46), (14, 47), (14, 48), (14, 49), only (14, 47) is a coprime pair. [pure mathematical fact]
>
> (8’) Hence, among this group of bicycles, the bicycle with a 47-tooth front cog is likely to have the longest-lasting rear tire. (from 31, 61, 71).

Baker uses this strategy to show that the explanations of the cicada case and the BFGBs have the same mathematical explanatory core, given that Cicada MES$_{\text{GEN}}$ and Skid Patch MES$_{\text{GEN}}$ share the statements of the argument that refer to mathematical facts about unit cycles (see Baker 2017, 202-203). His conclusion is that topic generality involves non-physical entities and allows *unificatory power* in a way that is impossible to achieve by the nominalistic versions of the argument so the correct perspective with regard to mathematical entities is mathematical platonism.

Therefore, here mathematics would be playing a role that goes beyond representing, it would be picking up aspects of the world and genuinely explaining them. *How?* The fact that the BFGBs argument shares its first three premises with Cicada MES$_{\text{GEN}}$, Baker argues that it is evidence for the scientific value of *topic generality*. This property is what makes disparate phenomena such as insect life cycles and bicycle mechanics unified under a single explanatory pattern about unit cycles (see Baker 2017, 11). Hence, topic generality and unification are maximized in this version, to the point that Baker doubts that an analogous nominalistic argument that preserves this level of topic generality can be formulated.

It should be pointed out that Baker does not reduce mathematics’ explanatory value to the two kinds of generality discussed earlier. In addition to scope generality and topic generality, Baker makes use of the notion of “explanatory depth”, which has to do with how mathematical explanation shows a certain pattern of relations that NES* lacks. Since NES* lacks the mathematical aspects of the explanation, it does not explain why this fact about time holds for some ranges of durations and not for others” (Baker 2017, 6), while the MES versions of the argument (MES, Cicada MES$_{\text{GEN}}$ and Skid Patch MES$_{\text{GEN}}$) do account for it, by using the concept of primeness and thus tackling the *why* of the phenomenon.

2.5. Philosophical implications and Heavy-Duty Platonism

As Baker puts it (2017, 205), his Cicada MES$_{\text{GEN}}$ constitutes a difficulty for Easy-Road nominalism because it is not clear that unit cycles can be considered purely physical entities. A unit cycle can be manifested in various ways: temporally, spatially or even virtually. This concept of unit cycles can be problematic for the nominalist, since it disconnects from any specific instantiation. On the other hand, the platonist can take unit cycles as a class of abstract entities and use the mathematical apparatus of number theory to account for the case.

Baker’s challenge to nominalism is formulated, basically, in terms of topic generality, that is, it is ultimately topic generality the crucial feature that would defeat any nominalist strategy, since the unificatory power of mathematics cannot be achieved by the nominalistic versions of the argument.

The Cicada MES_{GEN}, the most general argument, illuminates how mathematics goes beyond representation by adding a series of premises including the relevant mathematical facts and how those apply to the cicadas’ case. With this in mind, together with the application of the argument to the case of rear wheels of BFGBs and gear ratios, Baker obtains the Skid Patch MES_{GEN}, so both phenomena from separate areas share the same mathematical explanatory core.

As for the ontological picture that follows from Baker’s account, from the analysis of these arguments, he suggests that heavy-duty platonism is right. Heavy-duty platonism is the view according to which how concrete entities are fundamentally depends on how mathematical entities are[32]. The resulting metaphysical picture is one where the mathematical and the physical realms cannot be orthogonal, so they are not metaphysically independent. More specifically, Baker argues that mathematics explains by unifying (and thus exhibiting its generality) because how concrete things are depends on how mathematical facts are, that is, there is something in the mathematical structure that, in a way, is indispensably explanatory of the empirical phenomena in question. This explanatory indispensability is what Baker uses to argue for a realist (platonistic) picture of mathematical ontology.

This, as we know, carries an ontological commitment to abstract entities that many are not willing to accept.

For instance, Liggins (2016) rejects the ontological picture of heavy-duty platonism for two main reasons. The first one is that our pre-theoretical intuitions are not compatible with the idea that the way physical objects are depends on how mathematical objects are. The second one is that there is no decisive reason to choose one scale or another (since there are various ways of measuring magnitudes) as the correct one that will determine how physical things are. Liggins’ conclusion is that heavy-duty platonism is wrong and the platonist should go for an account of the relation between the mathematical and the physical realm in terms of a grounding relation.

Baker (2003) rejects this objection by arguing that it relies on an analogy between the mathematical and the physical real that does not hold, since it begins by pointing out the acausal character of abstract entities and then demands that these could be relevant to the physical world by having causal consequences.

Moreover, Knowles (2015) devotes a whole paper to the thesis that none of the objections to heavy-duty platonism is successful. He bases his paper on five different kinds of arguments against HDP, of which here is a (necessarily) brief summary:

The first argument (see Knowles 2015, 3-6) is based on the Lewisian distinction between intrinsic and extrinsic properties (some physical magnitudes are intrinsic properties; according to HDP, all physical magnitudes are extrinsic properties, so HDP is false). Knowles objects that the (initially intuitive) distinction between intrinsic and extrinsic properties must be based on our theories about their nature and physics, in particular, challenges the distinction as was put forward by Lewis. Next, he presents two arguments alluded to by Crane (see Knowles 2015, 6-7): (1) that HDP must come together with the claim that there is a metaphysically privileged measurement scale or that “a physical magnitude is a case of an object being related to all numbers the magnitude property is measurable by”, and (2) that HDP implies the contradiction

---

[32] This characterization is taken from Knowles and Liggins 2015; Liggins 2016.

that physical objects have some of their causal powers by being related to non-causal objects. Knowles claims that this is not contradictory on the grounds that it can be argued that there is a metaphysically privileged measurement scale, thus undermining Liggins' second reason for rejecting HDP (see above). Lastly, he points at arguments by Field (see Knowles 2015, 14-15); based on the idea that HDP is inconsistent with relationalism, both in its letter (only relations between aggregates of matter can be posited in our fundamental theory, whereas HDP involves relations between aggregates of matter and numbers) and its spirit (only unproblematic relations should suffice whereas HDP introduces problematic relations which would have no explanation). However, Knowles argues, HDP establishes a robust connection (in the metaphysical level), between physical objects and numbers, and numbers are explanatorily relevant to physical phenomena, since the fact that physical magnitudes are partly constituted by causally inert entities does not stop them from being causally relevant to those phenomena. From these observations, Knowles concludes that all main objections to the thesis of HDP fail.

Of course, this does not mean that HDP is right. It just points at some doubts as to whether it can be easily rejected, by pointing out that the main arguments put forward in the literature do not succeed at that task. Knowles's paper deserves some attention in the debate if we are to be cautious with our arguments for and against the HDP approach to mathematical entities. Perhaps it is not the end for heavy-duty platonism. However, this does not mean that the EIA debate is the one that will get the thesis through. As we know, after several versions and refinements of the argument put forward by Baker, there is still no conclusive evidence that heavy-duty platonism is right on the grounds of mathematics' indispensability.

Apart from this doubt about the perspectives of the heavy-duty platonist thesis, there are some other preliminary concerns with Baker's view so far.

First, while it is clear that the Cicada MES_{GEN} is the most topic general version of the cicada case argument that can be found in the literature, it is doubtful that it is the most explanatory version. As we shall see, an unlimited increase of topic generality in an explanation does not always render the explanation more powerful, as it can compromise cognitive salience and ultimately *lose* some of its explanatory power in the process.

Second, the power of Baker's argument relies not just on indispensability itself, but in whether its explanatory virtues are actually as powerful as he characterizes them. It is not clear that an argument based on explanatory virtues can demonstrate an ontological claim. Without further argumentation, the step can be thought of as being unwarranted.

Third, it is difficult to distinguish between the mathematical apparatus that is representationally indispensable and the mathematical apparatus that is genuinely explanatory (and indispensable). In the absence of a criterion demarcating the two, it may seem that it remains a question of pure intuition. Besides, the fact that unit cycles can model the empirical system does not necessarily imply that the latter depends on the mathematical structure.

Fourth, the notion of explanatory depth is unclear. Baker does not define it and it even seems that it is just a synonym for explanatoriness itself, so the notion does not clarify what we are talking about when we characterize mathematical explanation. Baker argues that NES* leaves unexplained why this fact about time holds for some ranges of duration and not for others, but he does not specify the notion any further, which suggests that explanatory depth and explanatory power are the same thing. If this were the case, the argument based on explanatory depth would not add much to Baker's point as a whole.

However, it is highly plausible that mathematics is actually providing us with something more than just representation of empirical phenomena in those distinct mathematical explanations, so the need to account for this feature is still there, especially if we conclude that Baker's approach is not the right one.

3. The Counterfactual Theory of Explanation

3.1. The CTE, an introduction

The Counterfactual Theory of Explanation (CTE) has arguably become one of the most popular theories of explanation in the search for an account of mathematical explanation[33].

It is a monist account of explanation, that is, a theory according to which both causal and non-causal explanations share a feature that makes them explanatory. In the case of the counterfactual theory, causal and non-causal explanations are explanatory by virtue of revealing counterfactual dependencies between the *explanandum* and the *explanans*.

The starting point of this theory is Woodward's interventionist proposal, which was initially intended to capture causal explanation. Although Woodward (2003) originally designed the theory to account for causal explanations, the core of the theory is not necessarily tied to a causal interpretation, for analysing explanatory relevance in terms of counterfactual dependence can be applied to non-causal explanations as well. As Woodward (2003) puts it, an explanation "must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways" (Woodward 2003, 11).

In the counterfactual framework, an explanation has two parts: *explanans* and *explanandum*. The *explanans* includes generalizations and auxiliary statements (such as initial conditions, boundary assumptions, theorems…). The *explanandum,* as well as the generalizations and auxiliary assumptions in the *explanans* have to be (approximately) true. Moreover, the *explanans* has to logically entail the *explanandum* and the covering generalization has to establish counterfactual dependencies between the auxiliary hypothesis and the *explanandum*. In order to introduce the CTE more formally, below are the five necessary conditions in terms of which Reutlinger characterizes the theory[34]:

1. STRUCTURE CONDITION: Explanations have a two-part structure consisting of a statement E about the occurrence of the (type or token of the) explanandum phenomenon; and an explanans including nomic generalizations $G_1, \dots, G_m$, statements about initial (or boundary) conditions $IC_1, \dots, IC_n$, and, typically, further auxiliary assumptions $A_1, \dots, A_0$ (such as Nagelian bridge laws, symmetry assumptions, limit theorems, and other modeling assumptions).
2. VERIDICALITY CONDITION: $G_1, \dots, G_m$, $IC_1, \dots, IC_n$, $A_1, \dots, A_0$ and E are (approximately) true.
3. INFERENCE CONDITION: $G_1, \dots, G_m$, and $IC_1, \dots, IC_n$ allow us to deductively infer E, or to infer a conditional probability $P(E \mid C_1, \dots, IC_n)$. This conditional probability need not be high, in contrast to Hempel's covering-law account; it is merely required that $P(E \mid C_1, \dots, IC_n) > P(E)$.
4. DEPENDENCY CONDITION: $G_1, \dots, G_m$ support at least one counterfactual of the form: if the initial conditions $IC_1, \dots, IC_n$ had been different than they actually are (in at least one specific way deemed possible in the light of the nomic generalizations), then E, or the conditional probability of E, would have been different as well.

Reutlinger adds a fifth possible condition: the minimality condition. That is, no proper subset of the set of *explanans* statements $\{G_1, \dots, G_m, IC_1, \dots, IC_n, A_1, \dots, A_0\}$ satisfies all of the

---

[33] We can find counterfactual accounts, for instance, in: Bokulich, 2008; Kistler, 2013; Pexton, 2014; Pincock, 2015; Reutlinger, 2016, in progress; Rice, 2015; Saatsi & Pexton, 2013; Saatsi, 2016, Knowles and Saatsi (2019).

[34] Reutlinger 2018, 78-79.

conditions 1-4 of the CTE. The purpose of this condition is to avoid including irrelevant factors into the *explanans*, that is, to guarantee relevance. With this fifth condition, the set of conditions becomes sufficient.

- **The concept of explanation in the CTE**

As for the concept of explanation at stake, according to the theory, explaining becomes a matter of providing information of systemic patterns of counterfactual dependence (see Woodward 2003, 191). Explanatory counterfactuals capture objective, mind-independent modal connections that show how the value of the *explanandum* variable depends on the value of the relevant *explanans* variable(s). They indicate how the explanandum would have been different, had the *explanans* been different. For instance, in the cicada case, (mathematical) abstractness allows us to extend the scope of the counterfactuals to any possible duration and provides information about the dependencies among the different variables, since prime periods minimize the intersection.

If explaining is a matter of providing information that answers what if-questions, then it is natural to regard as more powerful those explanations that provide more such answers (with respect to a given explanation). As Knowles and Saatsi (2019) have argued, the counterfactual account is not committed to "explanation through representation" (ETR) or the claim that successful explanations carry ontological commitments. We will discuss this more in detail later.

3.2. Knowles and Saatsi's version of the CTE

Knowles and Saatsi (2019) provide a version of the CTE applied to mathematical explanation of empirical phenomena. These authors adopt the counterfactual theory of explanation as their framework in the EIA debate regarding distinctively mathematical explanations. Their perspective is that of appealing to a well and independently established theory of explanation in order to account for the cases of "distinctive" mathematical explanation and thus overcoming the impasse of the debate around this issue.

According to them, neither side (platonists or nominalists) has provided a detailed, satisfactory account of explanatory generality, and how mathematics contributes to it. Taking the paper as a response of Baker's analysis[35], one of its aims is to defend that nominalists need not deny the explanatory role of mathematics, since the kind of contribution of mathematics can be accounted for in nominalist terms. Moreover, platonists are right when they argue that mathematics plays a genuine explanatory role but, they claim, this does not justify the ontological commitment to abstract objects.

Knowles and Saatsi point out that most accounts put forward by both nominalists and platonists share the principle of *explanatoriness through representation*, which is the idea that explanatoriness is a matter of representing explanatory features of reality. By this principle, the difference in explanatoriness lies on the relevant mathematical explanations representing explanatory features of reality that their nominalistic alternatives do not. This challenges the nominalists to point at the features of reality that are behind this increase in explanatoriness (if not mathematical features). Moreover, nominalists must address the task of denying that the

---

[35] See above.

kind of explanatory generality at stake is a genuine explanatory virtue, or they must indicate what non-mathematical features of reality underlie this increase in explanatoriness.

Knowles and Saatsi argue that, although mathematics contributes to explanatory generality in a way that allows us to say that it is itself explanatory (and not merely representational), this is still compatible with nominalism (see Knowles and Saatsi 2019, 2). Moreover, their claim is that their analysis is ontologically neutral, so both platonist and nominalist perspectives are compatible with it.

So, as we can see, one of the crucial features introduced by Knowles and Saatsi, from a nominalist – or, at least, ontologically neutral – perspective, is their claim that mathematics' role cannot be only representational, and still this does not commit us to mathematical ontology. If their theory is successful, this could be a breakthrough in the debate, since one of the claims of Baker and Colyvan against nominalist accounts was precisely that these did not account for that *something more* that mathematics is doing in explanations beyond representation.

Let us go through the main features of Knowles and Saatsi's account of mathematical explanation in terms of counterfactuals. The basic idea of the concept of explanation at stake is that explanatory counterfactuals capture objective, mind-independent modal (logical) connections. This corresponds to a dimension of abstractness (characterized by Jansson and Saatsi, 2017) that does the explanatory work by showing the relations of dependence among variables; how the *explanandum* would have been different had the *explanans* been different.

Knowles and Saatsi take three distinctive explanatory virtues in their account of what is going on in the periodic cicada case. These are non-sensitivity (corresponding to what Baker names 'scope generality'), degree of integration ('topic generality' in Baker's terminology) and cognitive salience (about how the information is conveyed). In this theory, mathematics explains by providing information about the explanatorily relevant features of the world, which are objective explanatory dependencies (see Knowles and Saatsi 2019, 17).

In this analysis, they follow Petri Ylikoski and Jaakko Kuorikoski (2010), who also suggest a counterfactual framework for comparing explanatory virtues. Ylikoski and Kuorikoski consider *five different explanatory virtues*: non-sensitivity, precision, factual accuracy, degree of integration and cognitive salience[36]. Here is a characterization of the three explanatory virtues selected by Knowles and Saatsi:

- **Non-sensitivity** is the range of values that the explanans variables can take without breaking the explanatory relationship. It corresponds to the *scope generality* of an explanation.

  As was explained in Chapter 1, Section 6, taking a toy example in which we are unable to divide 23 cherries evenly among 3 people. What if things were different and we had 21 instead of 23? 17? 15? There is nothing special about the number 23, as opposed to any other positive integer not evenly divisible by 3. Scope generality allows us to consider any natural number *n* as a possible value of the *explanans* variable quantifying the number of things being divided. Arithmetic undoubtedly makes it easier to abstract from the particular case.

- The **degree of integration** refers to the unification or connectedness to a larger theoretical framework.

  Again, with the previous toy example, the explanation has nothing to do with the nature of the things involved (that is, the particular things being divided). The arithmetical resource also explains why we cannot divide 23 strawberries evenly among

---

[36] For a general presentation of the five explanatory virtues, see Chapter 1.

3 children or 23 pairs of shoes among 3 men. To achieve topic generality, the explanation must abstract away from the particular topic we are considering.

- **Cognitive salience** is how easily a given explanation may be grasped. It is related to our being explaining creatures with limited cognitive capacities, and these limitations (partly) determine which explanations are preferable depending on the ability to enable explainers to draw counterfactual inferences for different values of the *explanans* variables.

Knowles and Saatsi argue that this last one is the crucial property of mathematical explanations, so they are explanatory and better than nominalistic alternatives in the sense that they increase cognitive salience. So, ultimately, it is cognitive salience what makes mathematics explanatory. Non-sensitivity and the degree of integration are about *worldly issues*, while cognitive salience is about how the information is conveyed.

It might be worth advancing Knowles and Saatsi's theses regarding the explanatory virtues. They argue that we can account for scope generality in counterfactual terms without including mathematical vocabulary. Mathematical explanations have the virtue of increasing scope generality and sometimes the mathematical apparatus is even indispensable in the description. However, there is a limit in the value of increasing scope generality, in the sense that the increase of information is not relevant past a certain reasonable point (Knowles and Saatsi 2019, 11). According to this theory of explanation, mathematics' indispensability in the procurement of scope generality is a matter of improving cognitive salience.

In the case of **topic generality,** according to this account, it is not always required to achieve the relevant level of generality, for there are linguistic resources to express the generalization in a neutral way and still succeed in accommodating all the relevant subject matters. Still, they consider the possibility that mathematics is sometimes indispensable to express an explanation that maximizes topic generality in order to examine the kind of explanatory contribution of the degree of integration (Knowles and Saatsi 2019, 6). Still, the authors emphasize that increasing topic generality does not make an explanation more explanatory unless it increases scope generality or cognitive salience.

Therefore, mathematical explanations present the relevant information in a better and more comprehensible way. Therefore, mathematics' usefulness is ultimately a matter of maximizing cognitive salience, which makes this a view that is compatible with nominalism. Mathematics' indispensable role in empirical explanations provides no reason to believe in mathematical objects.

Explanatory power is, then, a measure of how much modal explanatory information it provides to an explainer. This is a function both of the amount of information that the explanation provides about the relevant worldly features, and of the way in which this information is presented (see Knowles and Saatsi 2019, 17).

- **Application to the cicada case**

As we know, there is a sense in which mathematics (in particular the concept of primeness), explains what is going on with the cicada case. If we apply the CTE to the cicada case (as Knowles and Saatsi do), we will find that (mathematical) abstractness allows us to extend the scope of the counterfactuals to any possible duration and provides information about the dependencies among the different variables, since prime periods minimize the intersection. The mathematical explanation based on arithmetic is the one that provides us with more counterfactual information about what would have happened had things been different, in the sense that not only does it explain why thirteen and seventeen year life-cycles minimize intersection for the case of the cicadas, but also we can extend the explanation to account to other life-cycle lengths and other species if needed. The explanation provides a high level of topic and scope generality without compromising cognitive salience.

As we can see, Knowles and Saatsi consider cases where, though there is a nominalistic explanation available, the nominalistic alternative is never as explanatory as the mathematical original explanation. This is the case because, even though they arguably have the same scope and topic generality, the nominalistic counterparts often compromise cognitive salience or, as I would put it, are less efficient in satisfying our urge for understanding.

The generality that the mathematical explanations achieve is not a reflection of counterfactual dependencies in features of the world, which would be ontologically committing. In this account, we need not commit to any entity that does not feature in objective explanatory dependencies. Those dependencies only take place among facts of the (empirical) world, so mathematics is outside these counterfactual dependencies.

On these grounds, and given that mathematics' role is merely that of providing cognitive salience, Knowles and Saatsi argue that mathematics' indispensable role in empirical explanations provides no reason to believe in mathematical objects. In the same vain, they drop the principle that mathematics explains by representing phenomena as their response to the mentioned challenge of explanatory generality. So, although mathematics contributes to explanatory generality in a way that allows us to say that it is itself explanatory, this is compatible with nominalism, bearing in mind that mathematics' explanatoriness does not obtain through representation. If the approach is right, it would overcome Baker and Colyvan's concern that nominalists fail to acknowledge the features of mathematical explanation that go beyond just accurately representing phenomena.

- *A limit to generality*

One of the salient features of the account is that poses a limit to generality, in that the mere fact that an explanation is more general by itself does not necessarily help understand the *explanandum* better (Knowles and Saatsi 2019, 12). That is, a biologist would not prefer a particular explanation of the cicada case over another just because it integrates more domains if that does not illuminate the case in her own domain, though that more general explanation would undoubtedly broaden the explanatory landscape. Moreover, making an explanation more abstract in order to account for different areas of research can even carry a loss in domain-specific detail, which would ultimately result in a loss of explanatoriness. As Ylikoski and Kuorikoski (2010, 213-214) point out, this could mean a loss of other explanatory virtues such as factual accuracy or cognitive salience. If this is correct, Baker's emphasis on unification would ultimately end on a return to an implausible Quinean confirmational holism. Therefore, there is a limit to the idea that, given that explaining is a matter of providing information that answers what if-questions, then it is natural to regard as more powerful those explanations that provide more such answers (with respect to a given explanation).

In this analysis, the emphasis is on mathematics' role in increasing cognitive salience. Behind this notion lies the idea that human beings as explaining creatures have limited cognitive capacities, and our preferences in this domain will be determined by these limitations and depend on the capacity to enable particular explainers to draw counterfactual inferences for different values of the *explanans* variables.

This feature of the account is relevant, since it is related to the concept of explanation. Knowles and Saatsi see explaining as a human activity, the goal of which is the provision of explanatory understanding. This introduces epistemic and pragmatic aspects to explanation. From this, it follows that their notion of explanation is predominantly an epistemic one, related to understanding, as opposed to views that go for an objective concept of explanation, where explanatoriness would be mind-independent.

However, there is also nothing in the counterfactual account that rules out mathematical objects or properties bearing objective, explanatorily relevant relations to physical *explananda*.

There is the worry that, when they argue that “mathematics is not indispensable for generating a desirable level of scope generality” (Knowles and Saatsi 2019, 13), it seems that they are committing themselves to the need to provide formulations for the different cases with sufficient scope generality (degree of integration) and no mathematics in them, that is, formulations that look like NES*. The problem with this idea is that they would need to show that there is a nominalization (or a way to provide one) of every case of genuine mathematical explanation, even if mathematical generalizations just need to be cognitively salient (they need not be ontologically committing or even true). However, as we know, the possibility of reformulating every single explanation in which mathematics appears playing a genuine or substantive role is still in question.

4. A different approach: Baron's account

An alternative approach may come from Baron's work, in which he modifies the DN model to adapt it to extra-mathematical explanations, in order to account for what he calls *genuine* mathematical explanations. His aim is to provide an account of mathematical explanation that is in correspondence with a lightweight platonist approach to mathematical ontology, as we shall see.

4.1. In search of a suitable theory of explanation: Baron (2016, 2017)

In the paper "Explaining Mathematical Explanation" (2016), Sam Baron analyses several theories of explanation in order to account for the case of mathematics. His aim is to develop a metaphysically lightweight theory of mathematical explanation by appealing to resources available in the literature. He assumes some sort of mathematical platonism is correct and that the enhanced indispensability argument (EIA) is right. So, even though his proposal is quite different from Knowles and Saatsi's and the general perspective in the present work, it is worth examining it for it may provide some illuminating ideas to the debate.

In order to fit into Baron's requirements, the strategies in the analysis have to respect two constraints, a modal and a genuineness constraint. The **modal constraint** is a desideratum concerning the robustness of explanations, where the *explanandum* is shown to hold of necessity, for some modality. Therefore, a theory of extra-mathematical explanation should say something about this modal dimension of explanations of this sort (see Baron 2016, 459). The **genuineness constraint** refers to the fact that not every use of mathematics in empirical explanations is genuinely explanatory. Here, Baron notes that nominalists (such as Saatsi 2011 and Leng 2010) maintain that mathematics does not play a genuinely explanatory role, whereas platonists such as Baker (2005, 2009a) and Colyvan (2011) maintain that it does.

Baron (2016) assumes that there are cases of what he calls "genuine mathematical explanation" and searches for an independently established theory of explanation that allows a characterization of them and is compatible with a lightweight platonist approach.

In his search for such an account of mathematical explanation, Baron goes through at least five different well-established accounts of mathematical explanation available in the recent literature on the topic.

1. Laws. The first possibility is the appeal to laws of nature, resulting in a nomic theory of extra-mathematical explanation. Here, the explanatory relation between mathematical and empirical facts in virtue of which the former explain the latter is a relation of lawful determination. Two facts F and F^* are connected by a relation of lawful determination just when there is some law L that takes F as input and yields F^* as output.

The main problem with going this way, as Baron himself points out, is that metaphysical laws of the relevant kind are very similar or identical with the dependence relations posited by

Heavy Duty Platonism. Therefore, extending the nomic theory in this fashion will not result in a genuinely lightweight alternative.

2. Morphisms. An alternative is to consider that mathematics is applied within science via structure-preserving mapping relations – morphisms – between mathematics and physical systems. According to such an account, the explanatory relation between mathematical and empirical facts is a morphism of some kind. Morphisms have the potential to underwrite the modal guarantee that mathematical facts exert over empirical facts.

From Baron's perspective, there are two difficulties concerning this account:

- First, it would be too easy to model the broad structure of a physical system with some mathematics or other, and establish a structural relationship between the two structures. So for any explanation of an empirical fact featuring these continuously varying properties, there is a version of that explanation that includes a structural morphism between the real numbers and the continuously varying properties at issue.
- The second, more serious, problem for the structural mapping account is that mathematics is playing a representational role in scientific explanation when it is being used to 'represent' or 'index' underlying physical facts (see Baker and Colyvan (2011), Melia (2000), Saatsi (2011)). Baron is precisely trying to avoid the possibility that his theory of mathematical explanation yields the result that in every putative case of extra-mathematical explanation, the relationship between the mathematical facts and physical facts at issue is a representational one[37]. In short, his worry is that if the relation between mathematical and physical facts is always a representation relation, then the door is open for a Field-style nominalist to swoop in and undermine the indispensability of mathematics for explanation.

3. Entailment. The third alternative is an account according to which the construction of a sound argument in which an empirical fact is derived from partly mathematical premises is sufficient for extra-mathematical explanation. An entailment-based view would present no problems to satisfy the modal constraint, since the entailment of one fact by another is undeniably a strong modal relationship. An example of a view of this sort would be the one developed by Baker (2005, 2009a) applying it to the cicada case.

Baron's worry with this view is related to the irrelevance problem. That is, if we adopt an entailment-based approach without supplementation, it will face the difficulty that mathematical facts are necessities, and with respect to classical logic, however, for any conditional $A \rightarrow B$, if $A \rightarrow B$ is true, then for any necessary truth C, $(A\&C) \rightarrow B$ (see Baron 2016, 466).

4. Counterfactual dependence. According to accounts based on counterfactual dependence, as Baron characterizes them, mathematical facts explain empirical facts just when there is some counterfactual dependence of the empirical facts on the mathematical facts.

In the cicada case, the analysis would go as follows: "Suppose 13 had, in addition to 13 and 1, the numbers 2 and 6 as factors. From this supposition, it follows that a cicada with a 13-year life cycle would overlap with predators that have 2 and 6-year life cycles. Accordingly, a 13-year life cycle would no longer be an optimal life cycle for avoiding predation, since it would be about as good as a 14-year life cycle" (Baron 2016, 468).

Then, we can understand the explanatory relation between the mathematical and empirical facts via the following two counterfactuals (Baron 2016, 468):

---

[37] Note that some nominalists would probably have no problems with this worry raised by Baron. Recall that Baron here is focused on finding an account that would yield a lightweight platonist position in the ontological debate.

(CF1) If 13/2 = 6, North American magicicada would not have had 13 year lifecycles.

(CF2) If 17/2= 6, North American magicicada would not have had 17 year lifecycles.

According to Baron's analysis, the counterfactual theory of extra-mathematical explanation fails to satisfy the genuineness constraint. If mathematics is playing a merely representational or calculational role there is going to be a counterpossible available in which we *change* the mathematical facts so that different results are yielded for whatever calculation we might be undertaking (see Baron 2016, 469).

In order to illustrate this, Baron compares this case to the case of the train, where we use mathematics to represent the distance between stations as well as the speed of the train and a very basic mathematical calculation is deployed, namely: 10/10 = 1. He uses the following counterpossible: (CF3) IF 10/10 = 2, then T would not have arrived at the station at 3:00 pm. Baron sees no difference between the counterpossibles in both cases. In both of them, we find the expression of a relationship between a basic mathematical fact involving division and some empirical fact. If one of the counterfactuals expresses an explanatory relationship, it is hard to resist the same conclusion with respect to the other one (Baron 2016, 469). Such a case will always be analogous to the cicada case, if we accept Baron's counterfactual analysis of the case.

For obvious reasons, we will come back to the issue of counterfactuals and how they are differently conceived by Knowles and Saatsi, on one hand, and Baron, on the other. This will constitute Section 5.1.

5. Pragmatic Theories. The pragmatic approach to explanation establishes that explanations are answers to 'why' questions. The explanatory relation between mathematical and empirical facts, then, is just the relation between 'why' questions and their answers. In the cicada case, for instance, we could ask why North American magicicada have prime-numbered life-cycles. The answer to this 'why' question would be that prime numbers minimise the frequency of intersection between North American magicicada and predators with periodic life-cycles (see Baker 2005, 235). Similarly, one can ask why it is impossible to cross the seven Bridges of Königsberg crossing each bridge exactly once. Here, the answer would be that the seven Bridges have the same structure as a graph that lacks an Eulerian Path.

However satisfying and straightforward this approach may seem, Baron points at the issue that the options for elucidating the relevance relations seem to be exactly the same options already considered for giving a theory of extra-mathematical explanation (see Baron 2016, 476). The relevance relations can be of several kinds and they might involve "the counterfactual dependence of empirical facts on mathematical ones, or the entailment of empirical facts by mathematical facts, or the determination of empirical facts by mathematical facts via the laws of nature, or the structural relevance of mathematical structures for physical ones and so on" (Baron 2016, 476). One advantage is that there is no need to pick a particular kind of relation, since it can take them all or some of them depending on the particular context. However, at this point Baron had already discarded each and every one of the relations considered as good candidates to be the explanatory relation between mathematical and empirical facts for the case of extra-mathematical explanations.

4.2. Baron's proposal: The Deductive-Mathematical Theory of explanation

After rejecting all five theories analysed in the 2016 paper, Baron sketches a possible account of mathematical explanation that would be compatible with his preliminary constraints. Taking the idea of entailment, he explains that the problems raised for entailment-based pictures suggest that classical logic is not the one we should be using as a basis for a theory of extra-mathematical explanation and suggests that some sort of relevance logic should be taken into account.

In his 2017 paper, titled "Mathematical Explanation by Law", he elaborates the proposal: the Deductive-Mathematical Theory of explanation.

According to the basic Deductive-Nomological Theory**,** an explanation is a deductive argument where the premises are the *explanans* and the conclusion is the *explanandum*, and said *explanans* is a set of initial conditions and laws of nature (See Woodward and Hitchcock 2003). Baron intends to apply the main ideas behind the theory to the mathematical case.

A statement of the resulting theory would be the following[38]:

> **The Full-Blown DM Theory of Extra-Mathematical Explanation:**
>
> 1. Extra-mathematical explanations are sound ℝ-Arguments, where an ℝ-Argument is an argument in which all of the information contained within the conclusion of the argument is contained within the premises, and each premise contributes some part of the information contained within the conclusion.
> 2. The conclusion of an extra-mathematical explanation is a proposition stating the physical phenomenon to be explained.
> 3. Among the premises of an extra-mathematical explanation, there must be at least one mathematical claim.
> 4. An extra-mathematical explanation must satisfy the genuineness condition.
> 5. Mathematical claims contain information about physical systems either indirectly, via structural mappings, or directly in virtue of physical objects possessing mathematical properties.

Baron's move here is to adapt this framework to the mathematical case in the following way. The theory presents some features that deserve our attention.

First, he chooses relevant logic as his background logic. His motivation is that classical logic is monotonic so, since mathematical theorems are true in every model, it is possible to add any mathematical theorem to our set of premises and we still get a sound explanation, with the risk of there being irrelevant information. Still, he acknowledges the possibility of using classical logic as the background logic providing the addition of the following containment constraint: "The premises of an extra-mathematical explanation must contain all of the information contained within the conclusion and each premise must contribute some part of that information" (Baron 2017, 27).

The obtaining mathematical explanation must satisfy the genuineness constraint, and for it to be genuine it must also comply with the "Razor-Sharp Essential Deducibility Constraint (REDC)" and pass the informational test. These constraints are designed to guarantee that genuine mathematical explanations do not collapse in descriptive uses of mathematics. If that were so, the EIA would ultimately collapse with the Quinean version of the indispensability argument.

---

[38] Baron 2017, 36.

The Razor-Sharp Essential Deducibility Constraint (REDC) states that explanations including among its premises mathematical information are better than any possible nominalistic alternative.

As for the criterion for determining what a good theory is, Baron appeals to the explanatory virtues of simplicity and unity. These features, as has been seen, are among the most common explanatory virtues that philosophers go for in mathematical explanation. So, for Baron, the best explanations are the ones that obtain an optimal balance between simplicity and unity. In this account, simplicity refers to the number of premises of the argument, where the smallest the number of premises, the better the argument.

Given two arguments A and A* with the same conclusion C, A is better than A* if A's premises can be used to establish a greater range of conclusions other than C[39]. Baron compares different explanations in terms of the consequences they have exclusively for physical claims. This is related to his intention to characterize extra-mathematical explanation, and thus disregarding mathematical explanations of mathematical facts in this analysis.

The informational test is a mechanism designed to ensure that the mathematical proposition includes non-descriptive information. So, when we remove the part of the information *I* that is descriptive about a physical system while holding the rest of the informational content of *M* fixed, there are two possibilities: (1) If the information in *I* exhausts the information that *M* carries regarding the *explanandum*, then *M* fails the test. Failing the test means that *M* is only contributing descriptive information to the explanation. (2) If, on the other hand, the information in *I* does not exhaust the information that *M* carries regarding the *explanandum*, then *M* passes the test, so it carries non-descriptive information regarding the *explanandum*. This information is used to derive the *explanandum* and thus explain it. In this case, *M* plays a non-descriptive role in the explanation (Baron 2017, 703-704).

In the case that there are no alternative derivations of the *explanandum*, *M* is indispensable to explaining it. This, however, does not affect the verdict of the informational test: in the case where *M* fails the test, even if it is indispensable to the explanation, it is only contributing descriptive information. If it passes the test, then *M* is indispensable to the explanation and also playing a genuine explanatory role in it.

This account provides a criterion for determining when the *explanans* carries non-descriptive information regarding the *explanandum*, and how this information is used to derive the *explanandum* and thus explain it. Baron (2017, 709) uses this resource to argue that mathematics does more than merely describe. So, if we apply this to the cicada case, the number-theoretic theorem provides information about why 13 and 17-years periods minimize intersection. The mathematics contains information about the physical features, so it seems that this account explains how mathematics contributes to explanation in a theoretical framework that looks to be compatible with some kind of (light-weighted, for Baron) platonism.

4.3. Ontological implications in Baron's account

This modification of the DN model, the Mathematical-Deductive theory of explanation, has the aim of characterizing mathematical explanation regarding its genuine explanatory role. Despite Baron's intention to get a lightweight platonist picture of mathematics, the ontological picture that results from this theory is, however, not clear. According to the theory, the relation

---

[39] Bear in mind that Baron places a relevance constraint, which affects the range of conclusions of A and A*.

between mathematical and physical facts is a relation of relevant logical consequence, which is to be understood in terms of the notion of information containment. As Baron recognizes (2017), there are two possible readings as to the level of ontological commitment that the theory carries.

The first one is that we should be committed to mathematical objects because it carries non-descriptive information about features of the world. This sounds compatible with mind-independent platonism, or even a stronger kind of platonism stating that how things are in the empirical world depends, in a way, on how things are in the mathematical realm (that is, the thesis of heavy-duty platonism).

The second possible reading is that we should be committed to mathematics because physical entities have mathematical properties. This appears to be a much weaker criterion for ontological commitment, and it easily fits into a moderate platonist perspective, which Field would characterize by stating that "relations between physical things and numbers are conventional relations that are derivative from more basic relations that hold among physical things alone" (Field 1989, 186–93).

The heavy-duty platonism is not Baron's preferred ontological choice, for he is going after some sort of light-weighted platonism. The second reading, on the other hand, is more in tune with Baron's purposes. With this perspective, the criterion would be that we should believe in the existence of mathematical entities because physical entities have mathematical properties, that is, mathematical facts carry non-descriptive information about facts of the world.

There are some concerns with this approach to the ontological question. Firstly, the argument based on the idea that if mathematical structures had been different then the empirical facts would have been different, so heavy-duty platonism is correct, does not hold. First, this idea is highly counterintuitive. Second, it makes more sense to think that it is the other way around, that is, if empirical facts had been different, then the mathematics used to describe them and make inferences about them would have been different, especially if we take mathematics as a product of human reasoning.

In the cicada case, mathematics allows the inference that certain life-cycle lengths (13 and 17) are the optimal solution in order for the species to survive. It is doubtful that it follows from here that there is a fundamental metaphysical relation involved or, at least, the resource of accounting for non-descriptive uses of mathematics is hardly the way to decide the question.

Besides, if Baron's analysis of substantive mathematical involvement in explanation is compatible with two different ontological positions, both heavy-duty platonism and some kind of moderate platonism, it is ultimately compatible with a nominalist approach. That is, this picture of mathematical explanation does not seem to settle the ontological question. The approach provides quite interesting elements of analysis, such as the incorporation of the relevance feature into the account, but it does not define a particular ontological picture of mathematical objects and leaves the door open to multiple possible ontological pictures (that could be compatible with the account).

5. How about a pluralist perspective? A sketch of Lange’s approach

As Reutlinger (2017) points out, Lange’s (2013c) account of explanation by constraint can be thought of as a pluralist approach, that is, an account according to which causal and non-causal explanations are covered by two –or more – theories of explanation, since there is no single theory that can capture all kinds of explanations. Therefore, we need to work with two or more theories to adequately account for causal and non-causal explanations. Lange’s work on philosophy of mathematics is vast, so here we only attempt to provide a sketch of his pluralist approach to mathematical explanation.

Lange (2013c) refers to Salmon’s distinction between “ontic” (causal) and “modal” theories of scientific explanation. The key feature about mathematical explanation is that it works in a modal way. Lange adopts a modal account, according to which there are non-causal explanations (mathematical explanations, among them) that operate by showing constraints in the *explanandum* phenomenon, that is, they show *why* the *explanandum had to occur*.

He exemplifies his approach by pointing out that “these explanations explain not by describing the world's causal structure, but roughly by revealing that the *explanandum* is more necessary than ordinary causal laws are. The Königsberg bridges as so arranged were never crossed because they cannot be crossed. Mother's strawberries were not distributed evenly among her children because they cannot be” (Lange 2013c, 491).

Therefore, mathematics’ role in this kind of explanations is that of putting constraints or “modal limits” to the *explanandum* and specifying the modal conditions behind them.

Moreover, Lange characterizes this necessity as being “stronger than causal necessity, setting distinctively mathematical explanations apart from ordinary scientific explanations” (Lange 2013c, 491). In this sense, mathematical modal facts are modally stronger than causal laws. This also serves as a criterion to distinguish between causal and non-causal explanations, since the modality involved is different and presents different strengths.

According to Reutlinger, Lange provides a pluralist account in that he argues that the modal conception applies to distinctive mathematical explanation in science, whereas the ontic (causal) conception does not apply (see Lange 2013c, 509-510).

In this section, the aim is merely to provide an example of a pluralist approach to non-causal explanation. However, Lange’s approach to mathematical explanation will reappear and obtain much more attention in the context of intra-mathematical explanation. There, some aspects of the approach will be useful to analyse mathematical explanations of mathematical phenomena.

6. Some potential problems within the CTE

Given that one of the main aims of the present work is to evaluate whether the CTE can provide a suitable account of mathematical explanation, it becomes necessary to address some of the potential problems it can present and whether they can be easily solvable. These have to do fundamentally with the application of counterfactuals, the veridicality condition and issues related to asymmetry.

6.1. The use of counterfactuals

As pointed out, one of the theories discarded by Baron is the counterfactual theory, so it could be thought that it is precisely the same theory presented by Knowles and Saatsi. However, these theories cannot be treated as just one account of mathematical explanation. The fact is that there is a relevant difference in the analysis and use of counterfactuals between Baron's proposal and that of Knowles and Saatsi's.

The main idea is that, while Knowles and Saatsi analyse counterfactual dependencies within empirical facts, Baron formulates counterfactual dependencies that are obtained between mathematical and non-mathematical facts.

a) Baron's counterfactuals

As for Baron's account (2016, 2017), the substantial role of mathematics in explanations has as an essential feature that mathematics plays a non-descriptive role. These non-descriptive uses of mathematics in explanations appear in the cicada case or the Königsberg bridges. On the other hand, as descriptive uses of mathematics we have, for instance, the case in which we use mathematics to measure. Baron employs the example of when we use mathematics to determine the time of arrival of a train given the departure station, the speed of the train, and the distance to the arrival station.

We have seen that in Baron's account we will find counterfactuals of the form "had the mathematics been different, physical things would have been different": "Suppose 13 had, in addition to 13 and 1, the numbers 2 and 6 as factors" (Baron 2016, 468). This counterfactual is what we can name "counter-impossible" (Williamson 2017) or "counterpossible", since the antecedents are worlds in which, for instance, the number 17 is not prime. The explanatory relation between the mathematical and empirical facts is summarized into counterfactuals such as "(CF1) If $13/2 = 6$, North American magicicada would not have had 13 year lifecycles" (Baron 2016, 468).

Baron contends (Baron 2016, 468-474) that the sort of dependencies between mathematical and physical facts that obtain are the same in descriptive and non-descriptive uses of mathematics, so the counterfactual account is not appropriate to articulate the peculiarity of non-descriptive uses, since there would be no difference between the case of the duration of the train journey and the case of cicada's life cycles. The problem with this account, according to Baron's analysis, is that it fails to satisfy the genuineness constraint, so the verdict is that the counterfactual account does not allow for the distinction between descriptive and non-descriptive uses of mathematics (which was one of his requirements).

b) Knowles and Saatsi’s counterfactuals

In Knowles and Saatsi’s account, the counterfactuals always deal with “changes” in the physical world, so instead of positing worlds in which the mathematics is different (non-prime 13 and 17, for instance), they posit worlds in which cicadas do not reproduce in 13 and 17-year cycles.

The authors formulate the generality and the two counterfactuals we see below in the context of a toy example, of the sort of the one analysed when the issue of explanatory virtues was presented. The toy example consists of the aim to divide 23 philosophical ideas into 3 sections (of a book, for example), so the initial statement is the following:

(A) There is no integer n such that $23/3 = n$. (Knowles and Saatsi 2019, 7)

With this starting point, they build the following counterfactual claim for scope generality, in order to show that the CTE is suitable to account for the generality that the mathematical explanation provides.

> **Scope generality example**: “(A’) If there is exactly one idea, then the philosopher will not be able to divide it evenly among three sections **and**$_{(2)}$ if there are exactly two ideas, then the philosopher will not be able to divide them evenly among three sections... **and**$_{(3 \times 10^{50})}$ if there are exactly 3×10^{50} ideas, then the philosopher will be able to divide them evenly among three sections” (Knowles and Saatsi 2019, 13).

This example shows how the mathematical explanation provides us with the scope for all relevant cases (numbers of ideas and collections). As for the use of counterfactuals, as we can see, Knowles and Saatsi analyse counterfactual dependencies between empirical facts, while Baron formulated them between mathematical facts and non-mathematical facts.

From here, there are two different –and opposed – conclusions: (1) Baron is not using the adequate approach to counterfactuals (so Knowles and Saatsi’s approach would remain unchallenged) or, alternatively, (2) Baron is right, which means that the counterfactual theory is not neutral since it does not allow the possibility of distinguishing between descriptive and non-descriptive uses of mathematics in an ontologically committing way, given that the counterfactual approach would not satisfy the genuineness constraint.

If the first is correct, then Knowles and Saatsi’s account still has a chance to prove its success, and the relevant counterfactual statements to examine would be those that apply mathematics (as it is) to the whole spectrum of possibilities in the physical world. If the second possible conclusion is correct, then the counterfactual theory of explanation would not be an adequate framework for examining the mathematical case. This being the case, it could be that Knowles and Saatsi have offered a biased theory of explanation by considering only those counterfactuals that render their theory ontologically neutral and unproblematic, and thus this might prejudge on the possibility of providing a demarcation between descriptive and non-descriptive uses of mathematics (which could potentially carry some ontological commitments they do not want to accept). Still, Knowles and Saatsi would plausibly argue that precisely Baron’s failure in accounting for extra-mathematical explanations within the counterfactual framework shows that the CTE should be applied in their fashion.

However, Baron himself admits that this situation could mean that the substantial role of mathematics cannot be accounted for in terms of their providing non-descriptive information that is ontologically committing (Baron 2016, 474). This would undermine Baron's objective itself, which is to formulate counterfactuals as determining explainers ("difference-makers"), thus carrying ontological commitment. Therefore, one more possibility is available, which is placing the *problem* in Baron's purpose itself.

There is a sense in which Knowles and Saatsi seem to be providing a more plausible picture of counterfactuals for extra-mathematical explanation, in that it can seem odd that we need to formulate counterfactuals "changing" mathematical facts in order to apply mathematics to the empirical phenomena to be explained. It seems more reasonable for us to keep mathematics *as it is* and focus on providing counterfactuals directed to what we want to explain, that is, the empirical phenomena. Since we are dealing with external mathematical explanations, that is, the application of mathematics to physical facts, then the range of variables should be in the physical side of the explanation, not in the mathematical apparatus we use to explain. Moreover, if the relevant counterfactuals just *vary* empirical information and leave the mathematical apparatus *unchanged*, then this is an indicative that it is not the mathematical entities the ones doing the explanatory work.

Baron's application of the CTE to the mathematical case does not seem to be the most plausible one. If any change in mathematics were required, then scientists and mathematicians would develop or find a mathematical tool that adequately fits into their goal. The idea that scientists reformulate well-established mathematical theories as Baron seems to be suggesting is highly implausible, that is, scientist do not apply mathematics to explanations by using counterpossible mathematics.

Given this situation, Baron's worries concerning mathematical counterfactuals should not pose a problem, at least when we are talking about extra-mathematical explanation. Of course, the situation changes when we address mathematical explanations of mathematical phenomena, since in this case the counterfactuals need to be *internal*, so the counterfactuals with impossible antecedent are more problematic. This will be discussed in Chapter 5.

6.2. The veridicality problem

As seen before, the CTE, at least in its original and more formal presentation, requires the explanation (consisting of nomic generalizations, statements about initial or boundary conditions and further auxiliary assumptions) to be true or, at least, approximately true in order to satisfy the veridicality condition we mentioned earlier. However, we can hardly say that this is the case for the counterfactuals we use to explain as understood by the CTE. That is, if we take mathematics to be not literally true, it is very unlikely for these counterfactuals to satisfy the condition. So, can we modify or extend the CTE in order to avoid the veridicality problem?

The worry that mathematical explanations do not satisfy the veridicality problem comes from there being many scientific explanations involving idealized (auxiliary) assumptions. This problem has already been detected by Reutlinger (2017). Interestingly enough, Reutlinger himself may have provided some suitable elements in order to find a way out of it. With this, I am referring to a paper titled "Understanding (with) Toy Models" (Reutlinger, Hangleiter and Hartmann 2018) which analyses the explanatoriness of toy models. Here, Reultinger et al. point at the distinction between *how-actually* and *how-possibly* explanations, a quite promising distinction as a way to solve the veridicality problem for mathematical counterfactuals in the context of CTE.

A toy model is a highly idealized and extremely simple model. Researchers construct such models in order for the experts in a particular field of inquiry to cognitively grasp with ease. Their success in providing understanding makes them a very common explanatory resource. Toy models are found across the natural and social sciences. As Reutlinger (2018, 1070) points out, paradigmatic examples "include the Ising model in physics, the Lotka–Volterra model in population ecology, and the Schelling model in the social sciences".

Toy models are characterized taking into account three essential features (Reutlinger 2018, 1072):

1. Models of this type are strongly idealized in that they often include both Aristotelian and Galilean idealizations.
2. Such models are extremely simple in that they represent a small number of causal factors (or, more generally, of explanatory factors) responsible for the target phenomenon.
3. These models refer to a target phenomenon.

Going further with the characterization of toy models, we should establish a distinction between embedded and autonomous toy models. Embedded toy models are models of a well-confirmed framework theory, and they are simple and idealized models of phenomena (see Reutlinger 2018, 1073), whereas autonomous toy models, while sharing the simple and idealized character with their embedded ones, are not models of a well-confirmed framework theory.

As suggested above, the problem of veridicality can potentially be solved by using the following distinction of modalities of explanation and understanding: how-actually explanations and how-possibly explanations (Hempel 1965).

- *How-actually* explanations possess an *explanans* which satisfies the veridicality condition, that is, that consists of actually true (or approximately true) statements.
- The *explanans* of *how-possibly* explanation refer to merely possible explanatory factors (for instance, to possible causes and mechanisms bringing about the *explanandum* phenomenon, if the explanation is causal).

The distinction between how-possibly and how-actually explanations can be incorporated in different standard accounts of explanation. For instance, it can be accommodated into the covering-law account and various causal accounts of explanations, such as Woodward's seminal theory of causal explanation. Of course, it could be included into the CTE as well.

The distinction is crucial for the case of toy models, a context in which Reutlinger (2018) argues that not all autonomous toy models provide how-actually understanding, since the interpretation is not compatible with the veridicality condition. These cases are better understood as providing how-possibly explanations and how-possibly understanding.

Hopefully, the resource of how-possibly understanding will also be suitable to be applied to the counterfactual theory of explanation in an attempt to solve its problem related to the veridicality condition of the *explanans.* As Reutlinger contends that some autonomous models are best interpreted as providing how-possibly understanding, we could argue that the same happens in the case of mathematical counterfactuals.

The solution based on the notion of *how-possibly* understanding has many advantages: first, it involves modal notions, which is in correspondence and suits well Knowles and Saatsi's account, in the sense that all their approach is based on counterfactuals' providing modal information about mathematical facts and facts of the world. Second, it is compatible with the idea that what is going on is that, had the empirical world been different, the mathematics we apply to it would have been different too. The resource of *how-possible* information might have to do with the structures that could have done the job of helping us explain empirical phenomena, so we could drop the veridicality condition, since it would not apply to the mathematical part of the explanations.

A different solution to the veridicality problem comes from Bueno and Colyvan's (2011) inferential account of applied mathematics, which, the authors argue, is compatible with both realist and anti-realist approaches to mathematical ontology. The difference will lie in the way in which each view interprets the success in obtaining inferences (see Bueno and Colyvan 2011, 366-367).

According to the inferential account, the main role of applied mathematics is inferential, that is, "by embedding certain features of the empirical world into a mathematical structure, it is possible to obtain inferences that would otherwise be extraordinarily hard (if not impossible) to obtain" (Bueno and Colyvan 2011, 351). Still, there are other roles for applied mathematics, such as "unifying disparate scientific theories, helping to make novel predictions (from suitable interpreted mathematical structures), providing explanations of empirical phenomena (from certain interpretations of the mathematical formalism)" (*ibidem*). These secondary roles are linked to the ability to establish inferential relations between empirical phenomena and mathematical structures or among mathematical structures themselves in the case of internal mathematical explanation.

The key aspect of this account, at least for the purpose of this work, is the role of choice and context dependence: "if, for example, we wish to determine the combined mass of a number of objects, we should use an interpretation mapping that assigns masses –not space-time locations, not lengths, or anything else –to each object. Similarly, at the interpretation stage we need to make a decision about how to interpret the sum obtained, because, in the uninterpreted mathematics, the sum is just a real number" (Bueno and Colyvan 2011, 355).

Bueno and Colyvan point out that in cases involving idealizations, there is no full mapping between the empirical set up and the mathematical structures (see 2011, 357). For these cases, there are partial mappings between the empirical and mathematical structures. The authors explain this partial mapping in terms of the partial structures approach, making use of the notions of partial structure and partial relation.

This approach would be useful to solve our veridicality problem in that it does not treat the question as an issue about truth or true explanations, the terminology is always of adequacy and choice of suitable structures. In doing so, it appears that there is no need for a veridicality condition at all. Of course, even if the inferential account can help solving the veridicality problem, the relevant aspects of it would need to be added to the counterfactual theory of explanation, and we would probably end up obtaining a new theory of explanation that combines the best of both. For instance, this contribution of Bueno and Colyvan's approach, applied to extra-mathematical explanations, could look like the following constraint: "This application of mathematics to empirical science involves a monomorphism ϕ from W to M", where W refers to the empirical world and M to the mathematical apparatus. This would incorporate the idea that the empirical *explanandum* (partially) instantiates the mathematical structure. A constraint of this sort would substitute the veridicality constraint and, thus, eliminate the veridicality problem for non-causal mathematical explanations[40].

---

[40] Suggested by Concha Martínez-Vidal.

From this analysis, it seems that the CTE, in Knowles and Saatsi's version, is faced with a dilemma between two available options:

1) The first option is to reformulate the counterfactual theory in order to weaken the veridicality constraint, incorporating some of Reutlinger's ideas on the distinction between *how-actually* and *how-possibly* information. The veridicality constraint would only apply to *how-actually* information; while *how-possibly* explanations (mathematical explanations would be of this sort) remain free of this condition. The modal component of this alternative should not be a problem for their account, since Knowles and Saatsi already introduce some crucial modal concepts in it.
2) The second option is to drop the veridicality constraint for non-causal explanations on the grounds that there is no need to work with true mathematical statements in mathematical explanations; it suffices if they are cognitively salient. This would be done by incorporating a constraint concerning the application of the mathematical structure to the empirical phenomena in the terms described above.

Both options look acceptable in order to avoid the veridicality problem. Still, the first one is more grounded and provides a justification as to why mathematical explanations are *free* of the veridicality constraint. Since mathematics provides how-possibly information and explanations, it is quite reasonable to regard this information as how the mathematical structure can help us explain facts of the world without it being literally true. Still, the second option seems to be supported by the intuition that an explanatory tool – mathematics – does not need to be true in order to have useful applications in science and ordinary explanations, as long as we chose the relevant structures in each case.

6.3. Flags and poles (the asymmetry problem)

The tension between the descriptive and the explanatory roles of mathematics has already made itself evident in this work so far. This problem is analogous to the asymmetry problem related to causality in scientific explanation[41].

The issue can be presented in the following way:

A vertical flagpole of a certain height under the sun casts a shadow of a certain length. Given either the height of the flagpole or its shadow, we can calculate the other one.

It seems quite uncontroversial that the deduction of the length of the shadow from the height of the flagpole is an explanation of the length of the shadow. However, the deduction in the other direction can hardly be considered an explanation, that is, it is unlikely that the height of the flagpole can be legitimately explained by the length of its shadow. The height explains the length and not the other way around, even though we can derive the height from the length and the length from the height. In the non-explanatory case, the equation can be considered a representation of the phenomenon, but it does not explain it.

This poses a question that Knowles and Saatsi should address within the CTE in order for their proposal to account for the crucial distinction between these two different roles of

---

[41] Though not presented here, the issue of asymmetry applied to mathematical explanation is also present in Lange's (2013c, 2017, 2018) theory of explanation by constraint. It was criticised by Craver and Povich (2017) precisely concerning asymmetry. These authors argue that Lange's approach does not account for the directionality of the explanatory relation in mathematical explanations. This might be something to bear in mind when Lange's approach is presented (Section 5 of the present chapter, Section 3 of Chapter 5).

mathematics in explanations of empirical phenomena: the use of mathematics to simply measure, track and describe phenomena, and the other more problematic use of mathematics to genuinely explain phenomena (such as what happens in the cicada case).

As far as I am concerned, it is likely that, in the application of the CTE to particular cases, the asymmetry issue will not present serious problems, given that the identification of the relevant counterfactuals will make clear the directionality of the explanation, thus excluding cases such as the flag and the pole. Here, the explanatory counterfactual will be clearly identified as the one going from length to shadow.

7. The CTE and the ontological question

There is no doubt that the philosophical study of mathematical explanation is nowadays highly influenced by the ontological debate concerning the existence of mathematical objects and abstract objects in general. Indeed, all the accounts of mathematical explanation that were addressed here had the ontological question in the horizon. Besides, the EIA debate is clear proof of the relevance of the ontological question. This debate can be interpreted as a question of how we should conceive of mathematical explanations, the role of mathematics in those explanations and the ontological consequences (if any) that arise from these role mathematics plays.

As we had the chance to observe, the positions vary from those that argue for a platonist ontological picture (Baker) to weaker platonist positions such as Baron's light-weighted platonism or nominalist positions. Here, Baker argues that mathematics' substantial involvement (in form of a non-nominalizable topic generality and explanatory depth) implies that the correct ontological perspective is that of accepting that mathematical entities exist. In contrast, Knowles and Saatsi attribute cognitive salience to this substantial role of mathematics, but this is not ontologically committing by itself. In addition, they provide a well-established theory that accounts for mathematical explanation and do not limit themselves to highlighting some explanatory virtues (as Baker seems to do). Baron also provides a well-established theory of explanation that can be interpreted as establishing certain relations between mathematical and physical facts, which leads to multiple metaphysical readings.

Baker has a point in acknowledging that mathematics does more than just representing phenomena in our explanations, since there are some compelling examples, such as the cicada case, the Königsberg bridges and the honeybee comb, that seem to highlight a deeper involvement of mathematics in explanations of empirical phenomena. Therefore, Baker is right in that we need to account for this substantial role of mathematics.

However, he probably jumps to the conclusion that heavy-duty platonism is right too quickly. Arguing for such a radical ontological position is probably too big of a leap considering the arguments he proposes to support that thesis. The fact that mathematics sometimes provides a high level of topic generality and renders explanations deeper (the meaning of which is not even clear) is probably not enough to argue for the existence of mind-independent abstract entities or the idea that how concrete entities are fundamentally depends on how mathematical entities are (which is the thesis of heavy-duty platonism).

It looks like, if there are no other good arguments to support the heavy-duty platonist position in the debate, the approach has no solid grounds and the EIA does not provide justification for a hard ontology in mathematics. If this is case, then platonism will ultimately rely on intuitions[42].

The CTE gets its place in the debate as an attempt to account for mathematical explanation without carrying an ontological commitment to the existence of mathematical objects. The account, as pointed out above, is intended to be ontologically neutral, that is, compatible with nominalism, as well as other ontological pictures. Accounting for mathematical explanations in

---

[42] James Brown (2012), for instance, argues for platonism with an argument based on our intuitions as the way our minds grasp mathematical concepts (and, in a way, get access to mathematical entities).

terms of counterfactuals can seem quite intuitive, since we often reason in terms of "what-if" questions

However, given that Knowles and Saatsi (2019, 13) argue that mathematics is not indispensable for achieving the desirable level of generality, the success of their approach could need nominalizations of all extra-mathematical explanations. This would make their argument somewhat weaker, since they would be committed to the need to provide formulations for the different cases with sufficient scope generality and excluding any mathematical statement, that is, versions like NES* for every single mathematical explanation we use in accounting for empirical phenomena. So, their claim that mathematical generalizations just need to be cognitively salient, that is, not ontologically committing or even true, could be not enough for their argument to get through. Moreover, it is highly doubtful (or at least it is in question) that we can actually provide reformulations for every single explanation in which mathematics appears playing a substantive role.

As pointed out before, Baron's analysis of the genuine role of mathematics in explanations is compatible with two different ontological readings: heavy-duty platonism and moderate platonism.

This is not unimportant. The fact that we can derive two different criteria of ontological commitment from Baron's account is problematic for his purposes, since he is clearly going for a particular ontological picture, that is, *lightweight platonism*, and it is precisely this intuition what drives his research and his proposal of the Deductive-Mathematical theory. Instead, he ended up constructing a theory that can also accommodate a heavy-duty platonist approach, somewhat far from his purposes. Moreover, I would argue, the very same theory can also accommodate a nominalist perspective, since it appears that the main difference between a deflationary platonism and nominalism is just a matter of interpretation, as has been seen in Chapter 1 when we addressed Leng's (2020) discussion on Maddy's approach.

However, this could mean good news for those looking for a well-established theory of mathematical explanation that is ontologically neutral. Baron's attempt to characterize extra-mathematical explanations in a lightweight platonism succeeds, but only if we admit that other (even *any other*) ontological pictures are possible and compatible with the account. Besides, it could be one more sign that perhaps the ontological question is not solvable via the analysis of mathematical explanation, which would pose serious problems for the EIA project itself.

This leads to the idea that relying on an independent theory of explanation does not suffice to settle the question and overcome the impasse of the debate. That is, it does not decide the debate on mathematical ontology by itself, since it seems that the idea that non-descriptive uses of mathematics in empirical explanations, which allow for information to go from the *explanans* to the *explanandum*, does not settle the metaphysical issue.

This is one of the main conclusions that we can draw from Baron's journey in providing an account that relies on an independent theory of explanation, and also from the CTE. Ultimately, no theory or account of mathematical explanation seems to settle the case as to which metaphysical theory is the correct one for the case of mathematics.

Since it is not necessary to hold a strong platonist view in order to account for the explanatory role of mathematics in science and ordinary explanations, there are reasons to reject it on the grounds that it is preferable to argue for a lighter metaphysics whenever possible. However, it appears from the analysis of the accounts presented by Baron, on one side, and Knowles and Saatsi on the other, that the main difference between a deflationary platonism and nominalism is ultimately just a matter of interpretation.

Also, as pointed out when we analysed Maddy's concern with the IA in Chapter 1, it is possible to believe a theory while remaining agnostic (or instrumentalist) about the existence of the objects to which the theory refers. In fact, the use of fictional idealizations this is a very common attitude in mathematical and scientific practice. This means that it makes sense to regard mathematical objects as some kind of idealizations, that is, we can believe in a theory and at the same time bear in mind that some of the objects to which the theory refers are just ideal. Numbers and fractions would share that feature with average families and infinite wires, and both thin realism (also referred to as "light platonism") and arealism tell the same story about the objective facts that underlie mathematical practice (see Maddy 2011, 112).

This is compatible with both a nominalist and an agnostic view (see Yablo 2002 or Plebani 2015). In this sense, Leng (see 2020, 127) is right when she points out (with Maddy) that if we take the predicate "exists" as it appears in natural science, there is no deep answer to the question of whether we should argue for the existence of those objects or they are just fictions. The difference between both approaches would be about whether mathematics needs to be considered true or not. As for the rest, both share the assumption that the goodness of mathematics lies in the objective depth of its concepts and theories, so our empirical evidence cannot decide between some sort of light platonism or a nominalist view.

As pointed out, we are committed only to the physical instantiation of mathematical structures and not to the system of abstract mathematical objects (see Leng 2012, 993). With this, comes the idea that the indispensability argument does not refute a nominalist approach, in that nominalists can account for the success of science, and also for the success of mathematics by arguing that the acceptability of mathematical statements comes from their following from accepted axioms.

The conclusion is in line with the idea also developed by Knowles and Saatsi (2019) in the context of the CTE, that there is no reason to take mathematical explanations in as involving mathematical objects in a strong platonic sense. The choice between a thin realism and an arealist approach will be a matter of interpretation and will ultimately depend on our intuitions.

8. Some conclusions

One of the purposes of this work is to examine the possibility of obtaining a **monist account** of mathematical explanation, that is, a single account that covers both causal and non-causal explanations. For this, the CTE was shown to be a good candidate.

Still, in order for the CTE to succeed at accounting for all kinds of mathematical explanation, two things must happen.

First, the counterfactual account should provide an answer to all the issues that have been raised against it, namely, the problem of addressing the role of counterfactuals in a way that accounts for mathematics' role in explanations of empirical phenomena (and decide the issue whether Knowles and Saatsi or Baron are right in their use and evaluation of counterfactuals), the veridicality problem and a better understanding of the ontological implications of the theory –that is, a deeper analysis of the question whether the CTE is ontologically neutral as Knowles and Saatsi want to argue, or if it is begging the question against the platonist. An sketch of how this could be done has been provided in the present chapter.

Second, in order to make sure the CTE is suitable as an account of all explanations in mathematics, we need to see how it applies to internal mathematical explanations, that is, mathematical explanations of mathematical phenomena. Knowles and Saatsi have only provided the first part of the analysis – the analysis of extra-mathematical explanation.

There is some hope, since the theory appears to be a work in progress, and there is probably much more to be said about the account, in particular, with respect to the modal aspects it introduces. We shall see in Chapter 5 how this framework could deal with internal mathematical explanation

Regarding the ontological debate, as we have seen, Baker (2017) and Knowles & Saatsi (2019) agree that the debate is at an impasse and explicitly aim their proposals to overcome this situation. More in particular, Knowles and Saatsi point out that, at the moment, neither side has provided a detailed, satisfactory account of explanatory generality, and how mathematics contributes to it.

With the idea of overcoming the impasse, Baker reformulates and strengthens his proposal, and by doing so he appeals to explanatory virtues: topic generality and scope generality. He also points at explanatory depth as one unique explanatory virtue that is achieved by mathematics.

From the other side of the debate, Knowles and Saatsi intend to surpass Baker's move and subsequent proposal by characterizing substantive mathematical involvement and its ontological commitments from the viewpoint of a well and independently established theory of explanation.

Baron (2016; 2017) rejects the counterfactual theory and defines substantive mathematical involvement in terms of a modified version of the nomological account of explanation. He aims at characterizing explanation in a light-weight platonism. However, as we have analysed, Baron's rejection of the CTE might be influenced for his choice of counterfactuals, that is, the fact that formulating counterpossibles for extra-mathematical explanation might not be the right methodology.

It looks like the strategy of relying on independently established theories of explanation does not settle the issue. The choice of one or another theory of explanation, or the way in which the theory is interpreted, is inspired by each author's preferred metaphysical view on mathematical entities.

Moreover, Baker's new explanatory version of the indispensability argument is not more convincing than the previous versions were. Already back in 2003, Baker argued that even if mathematics is indispensable, then "the status of no-difference is unclear, perhaps irredeemably so" (Baker 2003, 254). The explanatory version of the indispensability argument does not seem to take the platonist further than the classical version that did not rely on explanation. It is undeniable that MES_{GEN} is a quite impressive example of how mathematics has the power to unify phenomena of different domains by using the same explanatory pattern (the mathematical core about unit cycles that was common to both the cicadas and the gear ratios). However, it is yet to show how this explanatory feature of mathematics carries an ontological commitment to the entities that are involved, given that it can be interpreted in a light platonist or even nominalist fashion. This is the real challenge faced by positions such as Baker's.

For now, it is doubtful that the ontological question regarding mathematical objects can be successfully answered in the context of the EIA debate, since the choice of theories of explanation, or the way in which the theories are interpreted, appears to be inspired by each author's metaphysical intuition on mathematical entities. It also appears that the choice and interpretation also depend on those intuitions, so which the preferred theory is, is not independent from the account of explanation the authors accept.

As I mentioned, it is highly plausible that mathematics is actually providing us with something more than just representation of empirical phenomena in those distinct mathematical explanations, so the need to account for this feature is still there. The study of this debate and the different positions helps move the debate forward, so with this we get a better understanding of how mathematics explains –and why we use it to explain, independently from the ontological question.

As for the debate itself and the question of *who is likely to win*, what is clear is that no one has been able to settle the case yet. As tentative conclusions, we can point out that platonist strategies are not conclusive with regard to heavy-duty platonism, since there are still ways to avoid the radical platonist conclusion and provide different approaches to the relevant cases (such as the cicada case) and other ways to address mathematical explanation, as was shown. Given that Field's program is not likely to succeed; our attention turns to nominalist (and fictionalist) strategies, as well as moderate platonism. Interestingly, the accounts that were examined here cannot establish a clear difference (and preference) between both views, so it seems that they might not be differentiable by providing an account of mathematical explanation in the context of the EIA debate. If the question is between nominalism and moderate platonism and we have no resources to justify our choice of one or another, perhaps, on the pessimist side, the problem is that the task cannot be done. For now, our evidence shows that the EIA debate did not succeed at deciding the ontological question, and it is an indicator that it may never do so.

There is a persistent temptation to just drop the debate for the inconclusiveness it has shown (and the few perspectives of bringing a real breakthrough) and conclude that the EIA debate is not where the ontological question regarding mathematics will be answered. On the bright side, it does illuminate mathematics' role in science and how we use it to explain empirical phenomena.

This may lead us to adopt an agnostic position as to whether mathematical entities exist or not. Mathematics' usefulness is not in question, and our use of it does not vary depending on our metaphysical beliefs. In my opinion, after having addressed the debate on the EIA, it does not seem to be an extremely relevant and urgent question anymore. Ultimately, it would be a matter of intuitions –sometimes even mixed intuitions – where dialogues like the following would continue to occur:

- So, do you think numbers exist?
- *Of course they exist! They are everywhere! Look, here are* two *apples.*
- Yes, right. But those are apples, physical objects. I can’t see the number two in there.
- *Indeed, numbers do not exist by themselves.*
- So, are you saying numbers do not exist?
- *Of course numbers do not exist! That would be silly!*
- Do you realize you’ve just given me two opposite answers to the same question?
- *Yes… I guess it’s a difficult question, but…* ***why do you even care?***

Chapter 3: Intra-mathematical explanation: the role of mathematical proofs

1. Introduction

As we have seen, the first part of the present work is fundamentally concerned with applied mathematics and what is known in the literature as “external mathematical explanations”, i.e., mathematical explanations of empirical phenomena. Chapters 1 and 2 were dedicated to the case of external mathematical explanations, that is, the application of mathematics to explanations of empirical phenomena and the ontological implications it carries for mathematical objects. However, as mentioned, there is another dimension of mathematical explanation that deserves attention: internal mathematical explanation, that is, the use of mathematics to explain other mathematical results. This second part of the thesis addresses intra-mathematical explanation from a perspective in which mathematical proof has a salient role. These two perspectives provide unity to the analysis of mathematical explanation as a whole.

The debate on internal mathematical explanations in the literature seems to be dominated by an analysis of the role of proofs and how those can be explanatory. When we address the question of mathematical explanation and, in particular, internal mathematical explanation, it is unavoidable to approach the issue of mathematical proof, since it constitutes – or at least appears to constitute – the core concept of intra-mathematical explanation. This does not necessarily mean that only mathematical proofs constitute internal mathematical explanations. However, any account of mathematical explanation of mathematical phenomena must provide resources regarding how to conceive of mathematical proof in this context.

Indeed, in what follows, we will have the chance to point to some mathematical explanations that do not constitute proofs in the strict sense, and some mathematical proofs that fail to be explanatory, thus failing to be real mathematical explanations.

The plan

With this in mind, this chapter has as its main aim the clarification of the issue of mathematical proof (Section 2) by focusing on the following questions: First, (2.1) we will address the distinction between formal and informal proofs and introduce the debate around two main ways to conceive of mathematical proofs. Then, (2.2) some clues will be provided about the relation between proof and explanation, where one of the main issues will be the distinction between two senses of explanation that will be present throughout the whole discussion of intra-mathematical explanation. Third, (2.3) we will see a classification of different kinds of mathematical proofs.

After that, as related questions, we will examine the notion of explanation and its two possible meanings (Section 3), whether there is consensus in evaluating proofs (Section 4), what researches on educational contexts have to say on these matters (Section 5) and, finally, some clues on the aesthetic implications of mathematical proof (Section 6).

2. What is mathematical proof?

2.1. Formal and informal proofs

As often happens in philosophy with scientific and mathematical concepts, there is some controversy as to what constitutes a proof[43]. In this context, the distinction between formal and informal proof can both introduce more confusion and shed some light as to what constitutes a proof and the different meanings 'proof' can have. Let us briefly review some ideas in the literature concerning this debate in order to clarify the concept.

In "Prove – once more and again" (1997), Reuben Hersh points at the existence of two meanings of 'proof', which correspond to two tendencies in philosophy of mathematics. According to Hersh, one meaning is related to an account of what mathematics is in principle –the more formalist account of mathematical proof, whereas the other one is in correspondence with mathematical practice (see Hersh 1997, 153).

The first meaning is what we may call *formal proof* in a broad traditional sense of mathematical philosophy and philosophy of logic (from Aristotle to Frege and Gödel). In this sense, proof is "a transformation of certain symbol sequences (formal sentences) according to certain rules of logic (*modus ponens*, etc.)" (see Hersh 1997, 153). In this conceptualization, a proof is a sequence of sentences that can be structured as a logical deduction.

The other meaning of proof is *informal* and imprecise. According to this meaning, a mathematical proof is what mathematicians use to make other mathematicians believe in the truth of a theorem, an "argument which convinces the qualified, sceptical expert" (Hersh 1997, 153). It is the most common kind of *proof* if we go to mathematical practice, but it is hard to specify what it is with precision.

Hersh points out that there is an *official* standard viewpoint on proof, according to which any correct "practical proof" can be filled in to be a correct "theoretical proof" (see Hersh 1997, 154). That means that it should be possible for every *practical proof* to be translated into a *correct theoretical* one. This can be a problem, says Hersh, because not every informal proof can actually be directly translated into a formal one: "Take a mathematically accepted proof, and undertake to fill it in, to turn it into a formal proof. If you encounter no obstacle, very well! What if you do encounter an obstacle? That is, a mathematically correct step which you can't break down into successive modus ponens? Then you've discovered an implicit assumption, a hidden lemma! Join it to the hypotheses of the theorem and go merrily forward" (Hersh 1997, 154). Despite the sarcasm, this, of course, does not seem to be a satisfactory response to the problem.

We see that ordinary mathematical proofs are incomplete by the standards of normal logic. Hersh goes further to state that the formal concept of proof is irrelevant to the mathematician (see Hersh 1997, 155). The way mathematicians test their work is by refereeing (or warranting), and this should be equivalent to proof in the practical sense. His conclusion is that we can and will eventually change our notion of a rigorous proof, in order to accommodate this feature:

---

[43] This idea will appear again in Chapter 4.

“We'll allow machine computation, numerical evidence, probabilistic algorithms, if we find them advantageous (Hersh 1997, 162).

Granting Hersh can be right when he claims that there is always a conventional element in what we accept as a proof, it might be worth pointing out that the criterion is not arbitrary either, since it has to do with the sort of justification deductive proof provides, in contrast to other non-deductive methodologies. Maybe we can settle deductive logic as the methodology of proof, since anything that falls out of deductive logic seems to go beyond proof itself.

We should note that the idea of axiomatically systematizing proofs came quite recently with Frege’s aim to account for the foundations of mathematics from an epistemological viewpoint. In contrast, in ordinary mathematics we find not only gaps, but also skipped steps and the introduction of all kinds of resources in order to get the task done, that is, any path available can be suitable in order to obtain a proof or explain a result in a particular context. The conception of mathematical proof according to which the set of *resources* available is limited to the axioms and the rules of inference of the formal system is challenged by mathematical practice, since mathematicians do not write formal proofs in the strict sense when doing ordinary mathematics. This means that we might have to loosen our conception of proof in order to account for *real* proofs or just bear in mind that the formal concept of proof is more an idealization than a requirement for proofs in ordinary mathematical practice. In fact, it is highly likely that formalizing a proof has more to do with justification, whereas the actual obtaining of the proof is, most of the times, informal, as has been pointed out.

Interestingly, it looks like that this informal concept of proof is more closely related to explanation and understanding, whereas the formal more traditional meaning of proof refers to the mechanism we use to show *that* a theorem is true, independently of the question of *why* it is true. Perhaps this distinction can highlight that mathematical proofs have a role directly related to understanding, that is, an explanatory function. It might be that, when we go for a more informal proof, what we have in mind is the need for a proof that *explains* the phenomenon in question, in the sense of providing understanding when the formal proof cannot achieve that goal. The distinction between justification and explanation can be illuminating in this context. Put in this fashion, some proofs provide mere *justification* of a certain mathematical result, whereas sometimes we are not satisfied with such a proof and we go for a more *explanatory* proof or another kind of explanation (such as a diagram, an example, etc.).

In line with these comments, we can find Rav’s view that proofs exceed their purely logical-deductive function and have a relevant role in generating mathematical knowledge and understanding (see Rav 1999, 6). His thesis is that sometimes the search for a proof leads to other discoveries. For instance, Hilbert’s attempt to prove the Continuum Hypothesis led to Gödel’s inner model of constructible sets.

Besides, Rav distinguishes between what he calls a “proof” and “derivations”. According to this distinction, a *proof* is “a conceptual proof of customary mathematical discourse, having an irreducible semantic content” (Rav 1999, 11), whereas a *derivation* in a formalised theory **T** is “a finite sequence of formulas in the language of **T**, each member of which is either a logical axiom, or an axiom of **T**, or is the result of applying one of the finitely many explicitly stated rules of inference to previous formulas in the sequence.” (Rav 1999, 11). Therefore, there is a significant difference between proofs and derivations: an aborted proof with innovative ideas can lead to new discoveries, whereas an incorrect derivation leads nowhere. In this account, it is proof, in this sense, that constitutes knowledge in mathematics, which is why we need interpretation when presented with a formalized proof, so that we go beyond the underlying logic and get knowledge or understanding.

Rav’s diagnostic concerning mathematical practice and informal proofs resembles that of Hersh’s. We have remarkable mathematical theories of formal logic, but inadequate logical theories of informal (or practical) mathematics: “The ‘standard view’ of proof does not fit mathematical practice, nor is it capable of explaining the source of mathematical knowledge and the dynamics of its growth” (Rav 1999, 15). To support this idea, Rav points at examples of unaxiomatized theories (see Rav 1999, 14-15), such as matrix theory, graph theory and combinatorics, probability theory, number theory (as the term is understood by the general mathematical community, not axiomatized Peano’s arithmetic), and group theory.

The main conclusion is that proofs have an epistemic function since, rather than the statement-form of theorems, proofs are the bearers of mathematical knowledge, where theorems function as the labels or summaries for proofs: “The whole arsenal of mathematical methodologies, concepts, strategies and techniques for solving problems, the establishment of interconnections between theories, the systematisation of results —the entire mathematical know-how is embedded in proofs” (Rav 1999, 20). In addition, the social feature of mathematics is highlighted: “The social process of reciprocal crosschecks seems to be the only way to weed out errors and guarantee the overall coherence and stability of mathematical knowledge” (Rav 1999, 36). This social perspective does not make mathematics less objective. Rather, it enhances its reliability, since it includes constant checks and balances.

Still, without undermining the epistemic and social role of mathematical proof, we should bear in mind that the function of proof is not only epistemic. Since knowledge is factive (that is, if a proposition is known, then it is true), proofs are a means of getting mathematical knowledge and establish mathematical facts. This idea will be relevant later on when two senses (*strong* and *weak* or epistemic) of explanation are distinguished.

In contrast to these views, Azzouni (2009) argues that mathematical ordinary proofs already conform to what a “formalist” view would call a proof.

Azzouni joins the distinction between informal and formal proofs: “(a) Understanding an ordinary proof seems to proceed by the grasping of mathematical concepts, grasping the nature of the objects those concepts are about, and recognizing what follows from these—not by the recognized application of (explicitly formulated) rules. (b) Related to this, ordinary mathematical proof seems to be topic-specific, its transitions governed by insights into the subject-matter of the proofs” (Azzouni 2009, 11). Formalized proofs, in his view, have become the norms of mathematical practice. Therefore, if the implications (of assumptions to conclusion) of an informal proof cannot be replicated by a formal analogue, then the status of that informal proof as a successful proof will be rejected, regardless of the felt convictions of mathematicians (see Azzouni 2009, 14).

Therefore, Azzouni bases the thesis that formal norms are already accord with mathematical practice on the fact that their implementation did not cause a crisis in mathematics. “The striking point about the *Principia* program is this: implementing it didn’t expose a crisis in ordinary mathematical practice (i.e., the revelation of numerous informal mathematical proofs failing to actually show what they were purported to show). The recent computerization of formal analogues of informal mathematical proofs, and the checking of such reveals exactly the same thing: there has been no widespread rejection of mathematical results. Ordinary mathematical practice already *is* in accord with formalization norms, and apparently has been for its very long history” (Azzouni 2009, 15).

In conclusion, if Azzouni is right, there is no need whatsoever to reformulate our concept of proof, since ordinary mathematical proofs already are in accord with our formal or theoretical definition of what a mathematical proof is. However, this might strike us as impossible, given the multiple and various ways in which mathematicians introduce foreign resources and break the *formal law* of mathematical proof.

The formal concept of proof is addressed by Lakatos in “What does a Mathematical Proof Prove?” (1998). Here, a (formal) proof is a finite sequence of formulae of some given system, where each formula of the sequence is either an axiom of the system or a formula derived by a rule of the system from some of the preceding formulae. The particularity of this mechanism is that it allows us to decide of any given alleged formal proof if it really is a proof or not (see Lakatos 1998, 155).

Nevertheless, Lakatos points out that this formal concept of proof is not the one operating in mathematical practice, since mathematicians accept proofs that are convincing but do not prove the theorem in a logical sense. These proofs have no postulates, well-defined underlying logic, or way to formalize this reasoning: “What we were doing was *intuitively showing that the theorem was true. This is a very common way of establishing facts*¸ as mathematicians now say. The Greeks called this process *deikmyne* and I shall call it *thought experiment*” (Lakatos 1998, 157, emphasis in original).

In an informal theory there are unlimited possibilities for introducing new terms, hidden axioms, hidden rules in the form of new so-called ‘obvious’ insights, whereas in a formalized theory imagination is tied down to a recursive set of axioms and some rules. The crucial consequence of the contrast is that we might not be able to give a definition of (informal) proof which would allow us to decide at least *practically*, in most cases, if our informal mathematical proof is really a proof or not. That is, there is no method of verification (see Lakatos 1998, 157-160).

Moreover, Lakatos argues, the view that considers that “informal proof” is a proof with gaps is incorrect. He classifies mathematical proofs into three types: pre-formal, formal and post-formal. Pre-formal and post-formal proofs prove clear and empirical or vague and ‘quasi-empirical’ things, which, Lakatos argues, is the real though rather evasive subject of mathematics. This sort of proof is always liable to some uncertainty due to unthought-of possibilities. In contrast, formal proofs are reliable, but it is not absolutely certain what they are reliable about (see Lakatos 1998, 161).

In his paper “How to think about Informal Proof” (2012), Larvor addresses the tasks of the philosopher of mathematical practice in order to account for mathematical proof. He agrees that mathematical arguments suffer some philosophically important loss or distortion in the abstraction from ‘real’ mathematical proof to formal derivation. In this sense, the philosophy of mathematical practice has to specify what is philosophically important about mathematical practice, as represented by ‘real’ mathematical proofs, that is absent from derivations in formal logic (see Larvor 2012, 716-717). With this in mind, he distinguishes between formal and informal proofs in the following way:

Formal proofs: (a) are expressed in a general logical language, the well-formed formulae of which are explicitly defined (usually by recursion) and (b) consist of successive applications of explicitly specified rules of logical inference (in some systems some of these may be expressed as logical axioms).

Informal proofs fail to satisfy (a) and/or (b).

The standard of formal proofs in this approach is quite high. Indeed, Larvor argues that almost all mathematical proofs are informal by this standard, including the proofs published in research mathematics journals. The task for philosophers of mathematical practice is to identify and characterise those informal arguments that cannot be converted to satisfy (a) and (b), which

Larvor calls “essentially informal” arguments: “It is true that proofs written by experts for other experts are highly compressed. They omit steps that a ‘properly instructed’ expert reader can be expected to reconstruct. In this sense, published expert-level proofs *are* argument-recipes. However, they are *not* recipes for creating derivations in the proof-theoretic sense. To see this, observe that in filling in the gaps, an expert reader does not normally do the things required to translate a proof-used-in-earnest into a derivation satisfying (a) and (b)” (Larvor 2012, 725-726). In practice, mathematicians do not define a formal language and set of inference rules.

Informal arguments present the feature of having a strong dependence on their content, to the point that their validity or invalidity does not depend on their logical form alone, but also on their content. Besides, some context-dependence is present when Larvor argues that, for example, an appeal to authority may be a good argument dependent on the expert that it is appealed to and on what particular question (see Larvor 2012, 720). The view includes a broader variety of concrete objects with a relevant action in inferences, such as diagrams, models, expressions in special notations, experimental set-ups, etc. It allows us to see how the subject matter of informal arguments contributes to inferences, showing that inference is, in a way, action, instead of the view that postulates two highly abstract categories, the form of an argument and its content. This could help answering the question about which activities count as mathematical practice and which do not. Larvor (see 2012, 723) points out that the cost would be that we have to abandon the hope of establishing a general test for validity, as that would be impossible.

To sum up, this journey through the literature about the concept of proof has shown that mathematical practice goes beyond formal proof or even beyond immediately formalizable proof, that is, there is more to proof than (strict) formal proof and there is more to mathematical explanation than proof. This fact is in correspondence with the idea that mathematicians, as problem-solvers and explainers, use all available resources in order to find solutions to the problems they study. The authors arguing for a concept of proof that accounts for a more informal idea of mathematical proof have a point in that they seem to be more in correspondence with actual mathematical practice. This is not to say that the attempts to formalize mathematical proof and provide a more restrict and rigorous definition of mathematical proof does not have some advantages for some purposes. For instance, formalizing proofs can help us understand what is going on behind the process of getting a result, and it may help eliminate fuzzy and imprecise aspects in a particular mathematical proof. Besides, it is useful when studying how both mathematics and human reasoning work.

However, as has been shown, *real* proofs are rarely classical formal proofs consisting of premises, a perfectly logical chain of reasoning and a conclusion. Sometimes, there is a need to account for other elements in a mathematical proof, such as assumptions taken from other areas of study, diagrammatic reasoning, examples, illustrations of all sorts, and so on. Therefore, if we are to get a more complete grasp of what mathematical proof is all about, we must take proof in its broader sense by acknowledging all those informal – or *less formal* – proofs, and work with that broader concept of proof when trying to account for mathematical explanation.

The debate about formal and informal proof may be orthogonal to the construction of a theory of mathematical explanation, but having a clearer idea of mathematical proof is clearly relevant to the question of how to understand of internal mathematical explanation. This is due to the fact that there is undoubtedly some kind of connection between proof and understanding, so our concept of proof becomes relevant in this analysis. It looks like more informal or simply less strict concepts of mathematical proof not only fit mathematical practice better, but they are also more useful in accounting for mathematical explanation, since proofs are one of the preferred ways to explain mathematical facts. There is also a connection between the interpretation of proofs and explanation. For instance, if we are presented with a diagram with no interpretation, it is highly doubtful that the diagram constitutes a proof at all. The interpretation of any proof, as well as its subject matter, is relevant if we are searching for the explanatory role it might present.

Apparently, informal proofs are also linked to explanation in that when we need a better explanation of a particular mathematical fact, we quite often go for a more informal proof. Whether we should change our concept of mathematical proof or not is, of course, a debate that goes beyond the scope of this work. However, the distinction is useful in our characterization of mathematical proofs as (at least, potential) explanations.

2.2. Proof and truth

One of the main concerns regarding the issue of the concept of proof might be precisely how this concept determines our concept of mathematical truth, in an attempt to clarify ontological and epistemological aspects of mathematical objects. In this context, there are some ideas that might be worth pointing out.

Firstly, Hersh (1997) indicates two different perspectives regarding the purpose of proof, and the connexion between mathematical proof and mathematical truth. The first one is what he calls “absolutist”: “If mathematics is a system of absolute truths, an immaterial, indestructible aspect of Eternity, independent of human construction or knowledge—then mathematical proofs are external and eternal. They're to admire, hopefully to understand. Not to play with or to break apart. The absolutist teacher wants to tell only what he intends to prove (or order the students to prove). He'll try for the shortest proof or the most general one. Whether the proof is explanatory won't worry him, because the purpose of proof isn't explanation. The purpose is certification: admission into the catalogue of absolute truths” (Hersh 1997, 163). Here, explanation and understanding are not relevant when obtaining and communicating a proof.

The second perspective, the one he endorses, is the “humanist”, according to which mathematics is a tool: “For the humanist, the purpose of proof, as of all teaching, is understanding. Whether to give a proof as is, elaborate it, or abbreviate it, depends on what he thinks will increase the student's understanding of concepts, methods, applications” (Hersh 1997, 163). In this view, truth is a secondary matter, in the sense that the main goal of mathematical proof is understanding. This means that the central role of mathematics is to aid our understanding of concepts and applications, where, sometimes the most illuminating proofs are not the most general, the most formal or the shortest.

Resnik, however, sees “Proof as a Source of Truth”, as is the title of his paper (1998): “Taken literally and seriously, mathematics affirms truths about numbers, functions, sets, spaces and other entities, which are as real as rocks and yet inhabit neither space-time nor our minds. There are good reasons, which I shall not consider here, for philosophers of mathematics to take mathematics seriously and literally” (Resnik 1998, 318). His view is *immanent realism* about mathematical objects, which is the thesis that mathematical objects are abstract entities existing independently of our holding statements about them to be true. Immanent realism affirms that mathematical reality transcends our own existence, beliefs and experience. It derives its title from its immanent conception of truth, which applies only to sentences within its own language, whereas transcendent versions of realism employ conceptions of truth that transcend their home language through applying to sentences in a variety of languages (see Resnik 1998 319).

In addition, Resnik adopts a Tarskian definition of truth, so the truth/proof problem amounts to the problem of explaining why 'p' is provable only if *p* (see Resnik 1998, 320). This view carries a stronger concept of proof, as well as a more salient attribution of importance to the obtainment of truth as the (main) aim of proof.

Then, how does a proof show us anything about mathematical objects? Resnik’s answer looks simple: “Proofs make claims about mathematical objects. *To be able to understand a proof we must be able to understand claims about mathematical objects. If we are prepared to do that then we will be prepared to learn truths about mathematical objects from proofs*” (Resnik 1998, 325, original italics). However simple, this does not seem to be a satisfactory solution. In this view, we have access to mathematical objects from the information contained in our theories. Mathematical objects are some sort of combinations of the properties and relations to other entities that we attribute to them in our theories. Obtaining information about those properties and relations amounts to obtaining information about the abstract entities themselves.

Azzouni (2004), however, holds that in the mathematical context, co-referentiality has an entirely logical role that does not genuinely involve objects: when the terms are taken to co-refer, what is allowed is substitutivity *salva veritate* in all extensional contexts. Here, the new objects brought into the picture (and invented by the mathematician) are “ghostly reflections” of the new terminology that has been linked to the old terminology in certain mathematically valuable ways, so that we can refer to relations that hold among the objects under study. Moreover, these two observations entirely exhaust the so-called 'ontological content' of mathematical reference (see Azzouni 2004, 97).

Azzouni’s view in this paper can be summarized into two main ideas. The first one is that it is sensible for us to reject the apparent commitment to mathematical objects. The second one is that it also makes sense to dissolve away the idea that we continually pull out new content from the infinitely deep concepts that we have acquired, as opposed to the idea that we just continually augment our concepts by consistently augmenting the algorithmic systems from which such concepts arise (see Azzouni 2004, 98-99).

Azzouni suggests that objectivity is provided by some sort of mechanical procedure that is recognizable by other mathematicians (see Azzouni 2004, 103). This procedure can be more or less formal, from the formal rules of derivation in logic to informal procedures that, nevertheless, are included as standard practice among mathematicians. In light of this (necessarily brief) analysis, it may well be that our formal concept of proof is more of an idealization than a real standard to be met in most cases. Moreover, objectivity should probably be thought of as a procedure in which mathematicians perform various sorts of proof checking and proof recognizing, rendering a view of mathematics as a social activity that might resemble scientific practice in more ways than could be initially expected. This view is also supported by the concept of proof being somewhat participant-dependent. As I suggested earlier, the more informal concept of proof seems to fit actual mathematical practice better, which includes more resources than those of establishing some conclusion from the axioms and mathematical deduction. A broader concept of mathematical proof can perhaps do a better job in acknowledging less formal ways of getting a proof. Besides, it gives us an idea that the mere search for the *truth* is not the only aim of mathematical proof and, at least, the relation between proof and truth is not as straightforward as it might seem to be.

Of course, this does not undermine the value of the formal concept of proof, which is useful to help us understand the process of mathematical reasoning and mathematics as a formal system. However, understanding proof in a broader sense and having an open mind about this central concept can also facilitate the work of drawing conclusions about mathematical explanation, as this chapter will show.

Finally, the contrast between formal and informal proofs might be a question of degree. The issue of where we can draw the line or establish norms between what is a formal and an informal proof, or even between proofs and non-proofs is an interesting debate that could be pursued, where perhaps some of these norms are already in practice, so it is clear that much more work needs to be done on this subject.

2.3. Kinds of proofs

Finally, let us briefly review some kinds of proofs we can find in mathematical practice. This classification is not exhaustive, and it is related to how a proof is obtained.

For instance, there is **direct proof** (that is, a proof consisting of a combination of already existing theorems and lemmas with no further assumptions), **trivial proof**, **proof by contrapositive and contradiction** (also known as *reduction ad absurdum*), **mathematical induction**, and so on. The completeness proof in Gödel's version, for example, would be a **constructive proof**.

A special case for our purposes is **proof by exhaustion**. The reason it is relevant here is that this is the kind of proof we find in the Four-Colour Theorem case[44]. A proof by exhaustion, as the name suggests, is performed by exhausting all possible cases. The theorem is dissected into a finite number of cases, and then each and every one of them is checked separately. If all of them pass the test, then the theorem is true. It is not necessarily an atypical kind of proof, but it may not be as simple or elegant as a deductive proof. Moreover, the contrast between our understanding of a proof by exhaustion such as the Four-Colour Theorem proof and other more illuminating kinds of proof can be salient. Whereas with some proofs, as the Completeness Proof or even proofs of the Pythagoras Theorem we get to grasp *why* the theorem is true, in the case of the Four-Colour Theorem Proof we may believe in the correctness of the process and the result, but the proof leaves us with no real sense of knowing why the theorem holds or even without a good understanding of the proof. This may be due to our ability to follow some proofs and not others. We can hardly say that the Four-Colour Theorem proof is easy to follow, or even that it is possible to do so, given that a good part of the proof itself is performed by *non-human agents*. The key feature of the proof, as will be seen in Chapter 4, is that for the task of exhausting all the cases our human capacities are not enough, so we need to perform that task with the help of a computer, which has led to a good deal of controversy as to whether we can accept the proof or even questioning the concept of proof itself[45]..

Proofs by exhaustion constitute cases in which we know the truth of a theorem by putting all the particular cases together, which does not necessarily yield explanatory power. That is, a proof by exhaustion suffices to justify *that* a mathematical general statement is true, but it does not explain *why* this is the case, since the proof is just the sum of all the particular cases and not an explanation of the phenomenon in question. That is, even in the case of proofs by exhaustion where we can check all the cases, the way the proof is performed makes them explanatorily unsatisfying, as they do not seek or provide a reason why all the cases hold.

---

[44] See Chapters 4 and 5.

[45] See Chapter 4, Section 3.

To sum up, there are many ways to prove a mathematical result, which gives us an idea of how rich the concept of mathematical proof can be. Nowadays, many non-traditional methods and combinations of methods are used in mathematical proofs, which can make the concept of proof itself fuzzier and contribute to the thesis that an informal concept of proof suits better what goes on in mathematical reasoning through proofs. This yields many more ways to obtain a proof than those presented in this section, and makes the subject much more complex than it is when we address it from the constraints of strict formal methods.

3. Two meanings of ‘explanation’

A conclusion that can be drawn so far from the analysis of this chapter is that there is an unavoidable connection between mathematical proving and mathematical explanation. With the aim of getting deeper into the sense in which mathematical proofs can be explanatory, there is one key distinction that will help us go through the explanatory power of the cases presented in Chapter 4 and also to clarify the ways in which mathematical explanations function. I am referring to Brown’s distinction between a strong and a weak concept of explanation, which appears, though sometimes implicitly, in this work when addressing the explanatory power of certain proofs (or, more broadly, explanations).

In *Platonism, Naturalism and Mathematical Knowledge* (2012), James R. Brown takes a platonist approach to abstract objects, more in particular, mathematical objects, when analysing mathematical explanation. As a platonist, Brown adopts a view in which there are two separate realms: the concrete and the abstract realms. The concrete realm contains physical entities and processes, whereas the abstract realm contains abstract entities, among which we can find mathematical entities.

Within this framework, Brown distinguishes between two different notions of explanation, a strong sense and a weak sense of explanation. It is this distinction what will be of interest to us from Brown’s approach because, independently of the theses Brown defends, these two different ways of conceiving of explanation will be useful in our analysis of the concept of explanation in the case of mathematics.

1. In a *strong* sense, according to Brown’s terminology, there is a concept of explanation that makes reference to the *why*. In this sense of explanation, the *explanans* provides us with the *why* of the *explanandum*. Taking a very simple example from Brown (2012, 4), this is the sense of explanation we have in mind when we assert that the third Newton law explains the action-reaction forces.
2. The second sense of explanation, *weak* or *epistemic*, is closely related to understanding and justification. It is the type of explanation we have in mind in assertions such as “The teacher explained Newton’s laws”.

Brown even offers an illustrative example where both senses of explanation are shown: "The teacher explained how Newton's theory explains the tides" (2012, 15). Here, the first ‘explained’ is referring to the weak sense of the term, since the teacher is trying to get her students to understand Newton’s theory. The second ‘explains’ refers to the strong sense of the term, because the issue is why Newton’s theory explains the phenomenon of the tides.

Brown does not address the debate about the nature of explanation. However, he does argue for a representational view, according to which the mathematical realm represents the natural one. Mathematics hooks on the world by providing representations in the form of structurally similar models. In this view, mass is not a real number, it is *represented by* a real number. Therefore, the abstract realm cannot *explain* happenings in the physical realm, but it can provide tools needed for concept formation: models and analogies.

This distinction regarding the concept of explanation has, as Brown puts it, some consequences for mathematics. According to his view, mathematics is explanatory, but only in the second (*weaker*) sense. Brown’s main thesis is that abstract entities do not explain empirical phenomena and processes in a strong or substantive sense. However, they can provide tools to the constitution of concepts through models and analogies. Brown even goes beyond this idea to argue that in some cases this kind of understanding is the only one available, which mathematics is indispensable to the development of science and our understanding of the world (see Brown 2012, 1).

This distinction implies that, according to Brown's view, there is a clear contrast between mathematical theories and scientific theories regarding their explanatory power. Mathematics is not descriptive of what happens in the world, it rather provides models in form of analogies about how things could be[46].

This idea contrasts with the commonly accepted assumption that mathematics explains in a substantive (strong) sense, and this is precisely what makes it indispensable to scientific theories. Against this, Brown (see 2012, 4) argues that the fact that we use mathematics in science and ordinary life is not a sign that it is explanatory in the strong sense, that is, that they convey the *why* of phenomena. Newton used Latin to explain his laws, but Latin by itself does not play an explanatory role, and there is nothing in Latin that addresses the *why* of Newton's Laws (see Brown 2012, 4). However, it is a useful tool that helps us understand those laws, so its role would be more linked to the second (weak) meaning of 'explanation'. The case of mathematics would be analogous, it is a good auxiliary tool for our knowledge linked to understanding, rather than a source of substantive explanations, in his view.

Note that Brown does not argue that these two meanings of explanation exhaust the concept of explanation or that we cannot achieve a more complete analysis of the concept, but this, at the very least, illustrates that the predicate ”explain” can mean several things.

Lastly, the first (strong) sense of explanation has been referred to by other authors with terms such as the "substantive", "metaphysical" or even "objective" sense of explanation. The second (weak) sense of explanation is directly related to our understanding of an explanation, the way in which we get to know something, explanation in a *user-friendly* way. This points at the fact that, even though Brown makes the distinction explicit in his analysis of mathematical explanation, the contrast was already in the literature in more or less systematic ways. Therefore, in a way, this is just making explicit the distinction that *was already there*, which will be useful in our study of the explanatory power of (particularly) intra-mathematical explanations.

---

[46] Even though analysing Brown's ontological view on mathematics, it looks like this idea that mathematics provides models and analogies about how things could be resembles the distinction between *how-actually* and *how-possibly* information, where Brown here would be referring to a *how-possibly* kind of information.

4. Is there consensus in evaluating proofs?

It is true that when a proof is more explanatory than some alternative, mathematicians and philosophers usually have some sort of consensus on the matter. As just an example, this happened with the completeness proof, and the result is that we no longer use Gödel's proof of the theorem, since Henkin's version is a more unified proof, which provides more cognitive salience[47]. Nevertheless, there is some controversy as to which proofs are preferable, in the sense that there seems to be no consensus on the criteria that makes some proofs better than others, or even criteria for what constitutes a *good* proof.

There is an interesting research by Juan Pablo Mejia-Ramos and Matthew Inglis titled ""Explanatory" Talk in Mathematics Research Papers" (2013), which was performed by exploring the ways in which mathematicians talk about explanation in their research papers. Their thesis is that mathematicians do not frequently use this family of words (explanation, explanatory…) and that their use is considerably more prevalent in physics papers than in mathematics papers. Therefore, we find a contrast between how mathematicians approach the subject of explanation and how physicists do. Interestingly, questions around the explanatoriness of mathematical proofs are more frequent among philosophers of science and mathematics than it is among mathematicians themselves.

Besides, they argue, there is some evidence that mathematicians revise proofs in order to gain insights into how they can solve other problems. There is a pragmatic component to the study of their colleagues' proofs, that is, an interest in obtaining new research tools that is much more predominant than the aim of learning the reasons why particular results are true. The authors argue that in the case of physics the situation is just the opposite: physicists are more interested in discussing explanations in order to understand the reasons why of the phenomena they are studying, rather than just looking for new approaches for their research. This may have to do with the fact that physicists and mathematicians have different objectives in their research. Besides, it shows that there might be some variety of roles that mathematics plays in natural sciences and within mathematics, which goes beyond mathematics' mere ability to function as a tool in derivations or establish the truth of results.

This is related to Hanna's (1990) distinction between proofs that explain *why* a theorem is true, while others simply demonstrate *that* the theorem is true, stressing the idea that some proofs succeed at explaining a result more than others. Along with this idea, Hafner and Mancosu (2005) also make a distinction between two uses of intra-mathematical explanations: those that constitute instructions on how to employ a certain mathematical technique, and those that account for the mathematical facts themselves and *why* they are the case (see Hafner and Mancosu 2005, 217).

We can also find the distinction in Colyvan ("The Ins and Outs of Mathematical Explanation", 2018). According to the distinction, there are explanatory proofs, the ones that tell us *why* the theorem in question is true, and non-explanatory proofs, the ones that merely tell us that the theorem in question is true (see Colyvan 2018, 1). For instance, in the case of Königsberg Bridges, Euler's explanation is explanatory, whereas a mere combinatorial proof (that is, exhausting the possible combinations) is not.

---

[47] The completeness proofs are addressed in Chapters 4 and 5.

In search of what makes some mathematical proofs explanatory, Colyvan also points to the variety of mathematical proof, such as conditional proof, *reductio ad absurdum*, finite induction, transfinite induction, disjunctive syllogism, universal generalisation or proof by cases. He discards deductive explanation and relying on structure to account for the explanatory roles of mathematics and suggests going below the level of structure of the proof and the details of the proof as a more promising strategy (see Colyvan 2018, 4). Here, there are two different lines of thought, of which the first is that a proof is explanatory because it proceeds via the right kind of paths which connect results to other results in the same area of mathematics. The second line of thought is related to when a proof is seen as explanatory because it connects different areas of mathematics.

As we can see, once again, it is shown that one of the main roles of mathematical explanation is its power to unify different results, both within a particular area of knowledge or by connecting different areas of mathematics. There are two kinds of explanation working here, the local and the global power of unification. However, Colyvan does not provide major conclusions as to how mathematical explanations work and what is the explanatory role of mathematics.

In the same line of thought, there is a paper by several authors (Matthew Inglis, Juan Pablo Mejia-Ramos, Keith Weber, and Lara Alcocka, 2012) highlighting that very often mathematicians do not have fixed standards when evaluating the explanatory virtues of mathematical proofs. They performed a study in which 109 mathematicians were asked to judge the validity of a proof. From it, they extracted the following results. First, there is a substantial disagreement among mathematicians regarding even if the argument was a valid proof or not. Second, applied mathematicians were more likely than pure mathematicians to judge the argument valid. Third, the participants who considered the argument invalid were more confident in their judgements compared with those who judged it valid. Lastly, the participants who judged the argument valid tended to maintain their judgement even when they were presented with reasons for why the proof should be judged invalid.

Therefore, the evidence suggests that there is not a single standard of validity among contemporary mathematicians. This research can help support the suspicion that there is no agreement on the question of which explanations are better and which arguments constitute proofs. There is a chance that, independently from the actual competence of the mathematicians involved in assessing the proof, there is more fuzziness than wanted in the subject.

A similar idea to that of mathematicians having no clear standards of evaluation of proofs can be found in the paper by Colyvan et al. (2006), “Two Flavours of Mathematical Explanation”, on the apparent plurality in the analysis of mathematical proofs. Here, the authors develop the thesis that there are different explanatory virtues that come in degrees and, at least in principle, there is no clear standard for what constitutes a good mathematical proof, and some pluralism might make sense.

They do so by examining two different proofs of the Free Group Theorem, about which the mathematical community seems to be divided about which proof is the most explanatory[48].

The constructive proof shows how the universal property depends on features of its construction. This structural feature of the proof appears to fit the dependence-based model of explanation in the philosophy of science. In that case, assuming that a proof has explanatory value if it fits this model of explanation, then the constructive proof has a distinctive kind of explanatory value (see Colyvan et al. 2018, 242). This proof is constructed out of a set, and we are shown how the universal property definitive of free groups falls out of this construction.

---

[48] For the detailed presentation of both proofs, see Colyvan et al. 2018.

Colyvan et al. (2018, 243) believe that that this is similar to Steiner's idea that to explain the behaviour of an entity, one deduces the behaviour from the essence or nature of the entity and mathematical proofs exhibit this deductive structure. This proof is explanatory, since it satisfies the key requirements of the already existing models of explanation.

The abstract proof provides no information about the intrinsic structure of the component groups. It works with abstract relationships among groups to show how those relationships guarantee that (the subgroup of) this Cartesian product satisfies the universal property. This proof lacks explanatory value if we intend it to be modelled by a dependence account. However, it can present another kind of explanatory virtue based on the unificationist account of explanation that can be found in Friedman (1974) and Kitcher (1981). From this viewpoint, the proof shows that the event is part of a general pattern of events in the universe.

Colyvan et al. do not endorse a particular view of these two. Rather, they suspect that both provide a means of spelling out a source of explanatory value, and these sources need be neither necessary nor sufficient conditions for possessing explanatory power. They even find the proofs hard to compare due to their difference regarding their primary sources of explanation. So, the conclusion of the analysis is that both proofs –though different –show features that account for their explanatoriness. From this, they make cases for two plausible competing accounts of mathematical explanation and argue that there might be more than one kind of explanation at work in mathematics.

Indeed, the tendency in the literature seems to be to consider the possibility that many explanatory virtues have a role in explaining explanation, in showing which proofs are explanatory and which are not. It is perfectly plausible that these explanatory virtues come in degrees. We may even think that there can be some subjective aspects at work and even context dependence, as was shown when some aspects of the inferential account were incorporated into the counterfactual theory of explanation. If we take this road of focusing on the context and subject dependence of mathematical explanation, we might even end up doubting that there is an *objective* (understanding by 'objective', independent from context) sense of explanation.

In addition, it might be sensible to think that Steiner's and Kitcher's approaches capture two opposite but crucial features of mathematical explanation that any account should consider and accommodate[49].

The fact that these authors emphasize the connection between mathematical explanation and the dependence on the particular context or audience means that the concept of explanation at work here is the weaker meaning of explanation, connected to understanding, that is, a more epistemic sense of explanation. Nevertheless, it does not imply that the epistemic conception of explanation is the only one to bear in mind. Rather, what this analysis means is that one of the main roles of mathematics is that of illuminating certain aspects of (mathematical) phenomena by making them easier for us to cognitively grasp them. There is still room for a sense of explanation connected to some sense of objectivity, which covers a different, though related, explanatory role of mathematics.

It is clear, nevertheless, that there are several standards when evaluating proofs and multiple features to take into account. This suggests that our approach to mathematical explanation needs to be broad enough to accommodate the diverse features that make a proof explanatory, or even the attitude of keeping an open mind to the possibility that several frameworks or theories may work to account for the explanatory role of mathematics. This being the case, perhaps some kind of pluralism might make sense in approaches to mathematical explanations.

---

[49] A sketch of Kitcher's and Steiner's approaches will be displayed in Chapter 5.

5. Educational contexts

Although it may initially seem out of the scope of this work, studies targeting educational contexts can shed some light to the issue of mathematical explanation and how it works, by connecting it to particular cases of the acquisition and development of mathematical knowledge or mathematical thinking. In particular, three main issues can be taken up from the study of mathematics from an educational perspective. First, it can contribute to support the thesis that there are no fixed standards when evaluating proofs. Second, the teaching of mathematics can highlight the power of unification in mathematical explanation, by examining how unifying phenomena contribute to a better understanding of mathematical facts by students. Third, there is some evidence that understanding may have a more important role in mathematics than rigour itself, which contributes to the thesis that the concept of proof in *real* mathematical practice is not the strict formal one, but a more informal concept of proof where one of the main aims is to provide us with understanding. In this conception, various methods and resources are accepted, such as diagrams, computer aid, etc.

In order to support these ideas let us look at a few studies and analyses performed from an educational point of view.

A research study performed by Komatsu, Fujita, Jones and Sue (2018) found that unification enabled students to reach a more sophisticated level where they not only they verified that the statement —in this case, that the sum of the interior angles of a star polygon increases by 180° if the number of the vertices increases by one—was true, but also understood *why* the statement was true. The research is focused on the pedagogical sense of explanation, and concludes that introducing Kitcher’s explanatory unification into school mathematics (especially in the teaching of proofs) is worthwhile because one of the goals of mathematics teaching is for students to engage in mathematical activity authentically, that is, in the way practiced by the mathematical community. In addition, the synthetic power of unification is recognized as advanced thinking and it may offer learning opportunities for cultivating students’ mathematical and higher-order thinking. This shows that explanatory unification can be an effective activity for improving mathematical understanding among students.

Bear in mind that in Kitcher’s unification account, understanding is not just a matter of reducing the phenomena we do not comprehend, but of seeing connections and common patterns in what initially appeared to be different situations. In this sense, science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation repeatedly and, in demonstrating this, it teaches us how to reduce the number of types of facts we have to accept as ultimate (see Kitcher, 1989, 432).

Hanna (2010) explores the role of mathematical proofs in education and the usefulness of them in promoting understanding among students, and some other issues, such as the potential of applications of dynamic software such as heuristics, exploration and visualization, which make it a valuable tool in teaching mathematics, or the need for a view of the nature of mathematics and the relationship between deduction and experimentation in mathematics in the context of education. She argues that a proof is both convincing and legitimate when it leads to real mathematical understanding, so the component of understanding is even more important than the rigour component (2010, 6-7). Understanding, here, means helping us think more clearly and effectively about mathematics.

We have already discussed the fact that proofs are more than mere syntactic derivation. This is even more salient in mathematics education, where students are taught the standards of deductive reasoning but there is a side of the process where proofs are used to convey understanding (see Hanna 2010, 7).

As Hanna puts it, a proof can have several functions: verification (concerned with the truth of a statement), explanation (providing insight into why it is true), systematisation (the organisation of various results into a deductive system of axioms, major concepts and theorems), discovery (the discovery or invention of new results), communication (the transmission of mathematical knowledge), the construction of an empirical theory, the exploration of the meaning of a definition or the consequences of an assumption and the incorporation of a well-known fact into a new framework and thus viewing it from a fresh perspective (see Hanna 2010, 8).

This presentation of the several functions that a proof can have supports the thesis that there is more to proof than just mathematical derivation and the establishment of a mathematical fact. Besides, in mathematical proof, there is room for more than just logical resources. This includes the use of software with dynamic graphing capabilities in teaching geometry, which makes it easier to pose and test conjectures, so the use of computational resources is acceptable.

Hanna argues that the use of proof by exploration and analytic proof is not problematic, moreover, both are needed for good mathematical practice: "while exploring and proving are separate activities, they are complementary and reinforce each other. Not only are they both part of problem solving in general, they are both needed for success in mathematics in particular. Exploration leads to discovery, while proof is confirmation (Hanna 2010, 14). In fact, exploration has been an important part of mathematical practice, even before the appearance of computers on stage.

Following this analysis, it is clear that the explanatoriness of proofs should not be taken for granted, the fact that we have a proof of a particular mathematical result does not mean that we get understanding or that we were able to cognitively grasp its truth. Rather, at times, what we find is a procurement for more explanatory proofs or better explanations of results in teaching mathematics in order to convey understanding of the mathematical apparatus that goes beyond simply establishing the truth of certain results.

The evidence suggests that there is not a single standard of validity among contemporary mathematicians. This research can help support the suspicion that there is no agreement on the question of which explanations are better and which arguments constitute proofs. Again, perhaps there is more fuzziness than wanted in the subject. In addition, there is more to proof than logical deduction, and other more "experimental" resources can have a role. This goes in line with what will be discussed in Chapter 4 about the use of computers, where it is suggested that perhaps the use of computers is just an extension of our human cognitive resources, and other auxiliary tools may be used in a similar way.

6. Aesthetic concerns

References to the elegance or beauty of proofs goes back to ancient times, as this – certainly beautiful – fragment from Aristotle's *Metaphysics* demonstrates:

> Now since the good and the beautiful are different (for the former always implies conduct as its subject, while the beautiful is found also in motionless things), those who assert that the mathematical sciences say nothing of the beautiful or the good are in error. For these sciences say and prove a great deal about them; if they do not expressly mention them, but prove attributes which are their results or their definitions, it is not true to say that they tell us nothing about them. The chief forms of beauty are order and symmetry and definiteness, which the mathematical sciences demonstrate in a special degree. And since these (e.g. order and definiteness) are obviously causes of many things, evidently these sciences must treat this sort of causative principle also (i.e. the beautiful) as in some sense a cause. (*Metaphysics*, 1078a31 – 1078b5)

However, recently, there has been an increasing interest in the relation between explanatoriness and the aesthetic value of mathematical proofs. There is a tendency to attribute aesthetic virtues to some proofs we consider particularly elegant, simple or just beautiful, either for their presentation or the *process* going on in our minds when we follow a proof that is illuminating in a particularly *beautiful* way, almost as if they were works of art.

While the aesthetic side of mathematics could constitute a research project on its own, let us see some brief – and, necessarily, partial – remarks on the subject.

The aesthetic concerns with mathematical proofs have gained some interest recently, but this section will focus on some clues provided by Marcus Giaquinto (2016) in "Mathematical Proofs: The Beautiful and The Explanatory", where he links the aesthetic value of some proofs with their explanatoriness.

After acknowledging that some proofs are limited to establishing the truth of their conclusions, while others go beyond that by showing *why* their conclusions are true, Giaquinto works from the assumption that beauty and explanatoriness tend to go together. However, he ends up concluding that we have reason to doubt that explanatory proofs tend to be beautiful, and insufficient reason to believe or disbelieve that beautiful proofs tend to be explanatory. It is not obvious how abstract entities can be beautiful if we take this in a strict sense. However, the door is open to the idea that mathematical proofs can be susceptible to aesthetic evaluation and some of them correctly judged beautiful.

Giaquinto suggests as the notion of correctness for judgements of beauty that "a judgement of beauty about something of a particular genre is correct only if that thing has a propensity to give significant disinterested repeatable pleasure to connoisseurs of the genre from mentally apprehending it" (Giaquinto 2016, 55). From the analysis of two different proofs of the same result, he points at some properties that can be regarded as contributing aesthetically to a proof (see Giaquinto 2016, 61-62):

- **Clarity.** The fact that the central ideas of the proof and its structure are clear.
- **Brevity and simplicity of methods**. A balance between these two virtues is required: "We may feel that a proof of a given theorem is too long-winded for that theorem or that it uses a sledge-hammer to crack a nut. Perhaps it is better

to say: of two proofs of a given theorem, the one with the better balance of length and accessibility is preferred to the other, ceteris paribus" (Giaquinto 2016, 63).

- **Imaginativeness.** Presenting the proof in an imaginative way may cause an aesthetic reaction to it.
-

These properties show a direct link between beauty and explanatoriness in proofs. Regarding explanatoriness, Giaquinto's definition of an explanatory proof is quite sharp: "A proof explains its conclusion if and only if anyone who can properly understand the proof could come to know why its conclusion is true by following the proof" (see Giaquinto 2016, 64). He uses a quote from Michael Atiyah, which is relevant here as well, regarding the issue of explanatory versus non-explanatory proofs and mathematics: "I remember one theorem that I proved and yet I really could not see why it was true. It worried me for years . . . I kept worrying about it, and five or six years later I understood why it had to be true. Then I got an entirely different proof . . . Using quite different techniques, it was quite clear why it had to be true[50]". Indeed, this goes along with the idea that sometimes mathematicians look for better proofs when they think that a result, while established by a formal proof, is not well understood, or the proof in question is obscure and it does not illuminate the relevant aspects that lead to our understanding of the result in question.

As suspected[51], we can see from this brief presentation that the virtues that can make a proof beautiful collapse with those we pointed out earlier as *explanatory* virtues, that is, features contributing to the explanatoriness of a proof, either in the strong substantive sense or the one related to our understanding in subjective terms.

Giaquinto seems to agree with this idea that explanatoriness and judging that a proof is beautiful are often connected, since obtaining understanding from a proof "is pleasing because it satisfies a major intellectual desire. An explanatory proof can produce pleasure in people with a suitably developed mathematical mind by their direct intellectual grasp of the proof. Moreover, it is disinterested pleasure, independent of appetites or instrumental goals" (Giaquinto 2016, 66).

However, he rightly suggests that, at the moment, there is not enough evidence to conclude that beautiful proofs tend to be explanatory since, in order for that to be a justified conclusion, we would need to scrutinize a large sample of proofs. This deeper study is still to be performed. For now, it remains an interesting and potentially fruitful field of research, in which these and other features related to explanation and aesthetic values may be taken into account.

---

[50] Quote from Michael Atiyah: "Interview with Michael Atiyah," *Mathematical Intelligencer*, Volume 6 Issue 1, (1984), pages 9-19.

[51] This refers to a brief comment on the classification of explanatory virtues by Keas. See Chapter 1, Section 6.1.

7. Conclusions

This chapter aimed at providing a brief analysis of the concept of proof, since proofs are probably the main items of internal mathematical explanation. The result of that analysis was, in very few words, a challenge to the traditional strict concept of mathematical proof, according to which proofs are limited to logical deductions from axioms to conclusion. As we intended to show, the concept of proof at work in mathematical practice seems to differ from the formal concept of proof. This informal sense of proof has room for resources other than deduction with logical rules, and diagrammatic thinking, computer processing and so on can have a place in proving mathematical theorems.

One of the key conclusions of this chapter is that it is commonly accepted in the literature that there is a distinction between proofs that merely establish a mathematical fact and proofs that also explain *why* that mathematical fact holds. This distinction accounts for the fact that mathematicians do not always stop searching for proofs of a phenomenon once they construct the first one, which suggests that there is more to mathematics than just establishing mathematical truths.

In this chapter, we have also seen some preliminary features of explanatory proofs, such as the ability to unify different results within an area of mathematics, or even connecting different areas. However, it seems to be at least doubtful that we can get a clear view on how mathematical explanation works in order to construct a model that tackles it specific features.

Nevertheless, it has become clear that there is a relevant dependence of mathematical explanation on the particular context, which links mathematics with the weaker sense of explanation, the one connected with understanding. It leaves open the question of whether we can talk of a strong sense of explanation applied to mathematical proof, and mathematical explanation (both internal and external) more in general.

The role of mathematics in educational contexts and students' interaction with mathematical proof and the connections between explanatory proofs and aesthetical virtues contribute to a better understanding of the explanatory role of mathematical proofs.

To sum up, there is no doubt that mathematical explanations provide us with understanding, an explanatory role that can be thought of as what Brown characterizes as "weak sense" of explanation.

As we have seen, Brown takes this distinction to base his claim that mathematical entities do not explain (empirical phenomena, in his analysis) in a strong or substantive sense. Their role is to provide tools (models and analogies) to the construction of concepts. However, it could be that mathematical statements (or the connections those statements establish) are the ones doing the *strong* explanatory work, not the entities. Therefore, it is possible to argue for a strong explanatory role of mathematics without necessarily committing to the existence of the relevant entities. In order to argue for the thesis that mathematics also explains in the strong sense (or against that thesis) more work needs to be done.

For now, let us take the notions analysed in this chapter, which will be useful in Chapter 4 and, more specifically, Chapter 5. There, we will discuss the issue of a unified theory of mathematical explanation in light of the cases that are presented in Chapter 4.

CHAPTER 4: Internal mathematical explanation: some case studies

1. Proofs and explanatoriness: an introduction

As has been seen, in philosophy of mathematics there is a shared –sometimes implicit – idea that not all proofs are explanatory, that is, that proofs vary with respect to their explanatory value, and there is even doubt and different views as to what constitutes a proof and what does not, giving us the idea that even the concept of proof itself is debatable.

What is perhaps unquestionable is the fact that if we are to examine explanation in mathematics, the whole theorization around proofs has a lot to say on the matter, given that some of the key issues with internal mathematical explanation are the concepts of proof and explanatory proof.

Of course, it is not claimed here that the only means by which we can explain mathematical facts is formal mathematical proofs. There might be other kind of mathematical explanations that, though they do not constitute proofs in an strict formal sense, are nevertheless explanations of some other kind. All these issues will receive some attention in the following analysis.

With this in mind, the present chapter provides a presentation of the case study that will later on (Chapter 5) be used to test the CTE and clarify several issues of intra-mathematical explanation in general.

The plan

This chapter's purpose is just to shed some light onto the subject of which are the criteria for a proof to be explanatory and to analyse some aspects of the educational and pedagogical sides of mathematics and mathematical proof. Moreover, the issue of what constitutes a proof will also be addressed as a secondary aim.

In order to complete this task, we will analyse some illustrative examples of explanatory and non-explanatory proofs, and even cases of "proofs" where there is some doubt as to whether they constitute a proof or not.

With this in mind, this chapter constitutes a presentation of three particular case studies: The Completeness Proof (Section 2), the Four-Colour Theorem proof (Section 3) and the visual representation of the triangular numbers formula (Section 4), plus the issue of whether proofs by mathematical induction can be considered explanatory or not (section 5). These cases will later be used in Chapter 5, where the discussion on intra-mathematical explanation will take place with much more detail.

2. The Completeness Proof: from Gödel to Henkin

2.1 The relevance of the proof

The first case that will be taken into account is the Completeness Theorem, which is, undoubtedly, one of the most relevant results in mathematics. We will go through the proof in its two main versions, namely, the original from Gödel's (1929) work, and the most studied one by Henkin (1949). It is undoubtedly one of the most relevant results in mathematics.

Some literature on this case can be found. Mostly, it is concerned with the contrast between Gödel's and Henkin's proofs, with the aim of highlighting that the latter is more explanatory and provides a more unified proof that can be adapted and applied to other areas of mathematics.

Let us go through a review of the explanatoriness of proofs of the completeness theorem as it has been addressed in the literature, in order to get a clearer view on what we mean when we state that a proof is explanatory, or more so than another one.

John Baldwin (2018) wrote a paper addressing both proofs, arguing on the one hand that they are significantly different, and on the other that not only is Henkin's proof considered more explanatory, but also it allows us to analyse the role of unification in mathematics since it led to the reformulation of the entire discipline of arithmetic with modern model theory. In this vain, Baldwin addresses two main issues: (i) he offers a case study of the explanatory value of a particular milestone proof and (ii) he examines how Henkin's proof began the recasting of an entire discipline.

Firstly, he states both theorems in order to show the contrast:

Theorem 1, Gödel's formulation: *Every valid formula expressible in the restricted functional calculus can be derived from the axioms by a finite sequence of formal inferences.*

Theorem 2, Henkin's formulation: *Let* S_0 *be a particular system determined by some definite choice of primitive symbols. If* Λ *is a set of formulas of* S_0 *in which no member has any occurrence of a free individual variable, and if* Λ *is consistent then* Λ *is simultaneously satisfiable in a domain of individuals having the same cardinal number as the set of primitive symbols of* S_0*.*

Gödel worked in a background theory of naïve set theory and studies a single system of logic with predicates of arbitrary order, which is essential to the proof. He uses a definition of truth for atomic formulas, which is extended by deductive rules to determine truth in a structure for arbitrary sentences.

Henkin worked in a background theory of naïve set theory and studies the first order logic of each vocabulary. The proof for each vocabulary adds only constant symbols. He has a uniform definition of truth in a structure for each vocabulary that has no dependence on the deductive rules of the logic. This view underlies modern model theory.

One of the main differences between the proofs is that Henkin's proof, in contrast to Gödel's version, allows numerous generalizations, so much so that Henkin's proof has become a motif in logic, and it was a crucial step towards the modern conception of vocabulary. He makes the modern convention of a fixed vocabulary completely explicit.

In a fundamental sense, Henkin's argument is explanatory because he has identified the key features connecting the hypothesis and the conclusion, modifying both the syntactic and the semantic component. Henkin himself points out that his proof (unlike Gödel's) generalizes easily to uncountable vocabularies. The first order theory of R-modules can be developed uniformly regardless of the cardinality of the ring R.

According to Baldwin, the explanatory value of a proof can only be evaluated in terms of the intended audience, thus recognizing a certain degree of context-dependence in the explanatoriness of proofs[52] and employing an epistemic notion of explanation. The importance of Henkin's proof comes from it being a major component in the turn from model theory as an attempt to understand mathematical reasoning to model theory as a tool in many areas of mathematics.

Manzano and Alonso (2014) also address this case in a paper focusing on the evolution of the notion of completeness in contemporary logic. The paper is titled "Completeness: From Gödel to Henkin" and it is partly focused on the subject that we are addressing in the present section: the contrast between both proofs and the implications of the fact that currently it is Henkin's proof that is the one we actually use (sometimes even wrongly attributed to Gödel). They address the advances that Henkin's proof has brought to mathematics and logic. The proof is regarded as showing what goes on in the completeness theorem and the idea behind the proof has been repeatedly used in obtaining results about other logical systems, which results in a gain of application for more domains.

In order to obtain a higher level of detail, let us separately review key aspects of both proofs.

2.2 Gödel's Completeness Proof

Gödel's completeness proof is found in his doctoral dissertation of 1929 and also rewritten and published as an article in 1930.

The main thesis, as we know, is that an argument is valid only if it is derivable. This establishes that the deductive system is rich enough to provide a deduction for every valid argument. Together with the soundness result, this establishes the well-known result that all and only valid arguments are derivable.

Gödel's original proof is not easy to read today, for it includes concepts and formalization that we no longer use and thus are usually not available for students with standard background in logic and arithmetic. This is, partly, due to our current preference for Henkin's proof.

In order to get some historical perspective[53] on the question of completeness, let us recall that it was stated for the first time by Hilbert and Ackermann in 1928, *Grundzüge der theoretischen Logik*. This work was familiar to Gödel. Hilbert and Ackermann pose the question as whether a certain axiom system for the first-order calculus is complete, that is, whether all correct logical formulas for each domain of individuals can be derived.

---

[52] As we shall see, context-dependence will have a predominant role when explanatoriness is addressed in Chapter 5.

[53] Part of the presentation of Gödel's proof is from:
Kennedy, Juliette, "Kurt Gödel", *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/goedel/>. (Accessed 4th February 2020.)

In this framework, the Completeness Theorem states the following:

> **The Completeness Theorem:** Every valid logical expression is provable, that is, every logical expression is either satisfiable or refutable.

It is notable that Gödel goes through an exhaustive definition of all the relevant basic concepts he later uses in the proof. For instance, he defines a ‘logical expression’ as a well-formed first-order formula without identity. We can say an expression is ‘refutable if its negation is provable, it is ‘satisfiable’ if it is true in some interpretation and ‘valid’ if it is true in every interpretation.

Gödel uses the same calculus that had been used by Hilbert and Ackermann, where expressions are presented in normal form, that is, all the quantifiers occur at the beginning of the expression. Another relevant notion is the degree of a formula. The degree of a formula is the number of alternating blocks of quantifiers (at the beginning of the formula, since they are expressed in normal form).

Gödel’s strategy is to show that if the completeness theorem holds for formulas of degree k, it must hold for formulas of degree $k + 1$. Then, the question of completeness is reduced to the question of completeness for formulas of degree 1. Therefore, the main issue is to show that any normal formula $(Q)\varphi$ of degree 1 is either satisfiable or refutable, where “(Q)” stands for a (non-empty) block of universal quantifiers followed by a (possibly empty) block of existential ones. From here, Gödel defines a well-ordering of all tuples of variables that arise from the need to satisfy φ as dictated by (Q).

Following Kennedy’s (2018) presentation, let us take the case where $(Q)\varphi$ is $\forall x_0 \exists x_1 \psi(x_0, x_1)$. We list the finite conjunctions of the quantifier-free formulas $\psi(x_n, x_{n+1})$ in increasing length. In any domain consisting of the values of the different x_n, in which each $\psi(x_n, x_{n+1})$ is true, the sentence $(Q)\varphi$ is true.

There is a lemma that claims the provability of the formula $(Q)\varphi \rightarrow (Q_k)\varphi_k$, for each k, where the quantifier free formula φ_k asserts the truth of ψ for all tuples up to the kth tuple of variables from (Q), and $(Q_k)\varphi_k$ is the existential closure of φ_k. As Kennedy points out, this lemma is crucial, for it is the main step that was missing from the earlier attempts at providing a completeness proof by Löwenheim and Skolem. Besides, the lemma provides the connection between syntax and semantics completely explicit.

Kennedy uses the example of the formula $\varphi = \forall x_0 \exists x_1 \psi(x_0, x_1)$, where $\psi(x_0, x_1)$ is quantifier-free, in order to show how a particular formula is found to be either satisfiable or its negation provable. For that, we make the following definitions:

φ_0 is the expression $\psi(x_0, x_1)$
φ_1 is the expression $\psi(x_0, x_1) \wedge \psi(x_1, x_2)$
…
φ_n is the expression $\psi(x_0, x_1) \wedge \ldots \wedge \psi(x_n, x_{n+1})$.

The lemma we mentioned above shows that for each n, we can derive $\exists x_0 \ldots \exists x_{n+1} \varphi_n$ from φ.

From here, there are two possible cases:

1. For some n, φ_n is not satisfiable. In this case, by the completeness theorem for propositional logic, $\neg\varphi_n$ is provable, and hence so is $\forall x_0,\ldots,x_{n+1}\neg\varphi_n$. Thus $\neg\exists x_0\ldots\exists x_{n+1}\varphi_n$ is provable and $\neg\varphi$ is provable, i.e., φ is refutable.
2. Each φ_n is satisfiable. In this case, there are only finitely many possible models with universe $\{x_0,\ldots, x_{n+1}\}$. They can be ordered as a tree by defining a model M to be below a model M' if M is a submodel of M'. The resulting tree is finitely branching but infinite. By König's Lemma, there is an infinite branch B, which Gödel explicitly constructs in the original proof. The union of the models on B forms a model M with universe $\{x_0, x_1,\ldots\}$. Since M satisfies each φ_n, the original formula φ holds in M, so φ is satisfiable.

It is worth noting that Gödel’s completeness proof also gives the Löwenheim-Skolem theorem. Moreover, Gödel extends this result to countably many formulas and first-order logic with identity, and proves the independence of the axioms. In his 1930 paper, Gödel includes the compactness theorem[54] and uses the theorem to derive a generalization of the completeness theorem. As a relevant consequence of the completeness theorem, which is that categoricity fails for Peano arithmetic and also for Zermelo-Fraenkel set theory. The existence of non-standard models of the first-order Peano axioms follows from completeness and compactness. Skolem (1933) constructs a non-standard model of True Arithmetic, which is an extension of Peano Arithmetic consistent of the set of arithmetic sentences true in the natural numbers.

2.3 Henkin’s Completeness Proof

Henkin’s completeness proofs[55] rest on the central theorem that every consistent set of formulae has a model. In order to build the model for the first-order case, first we have to extend the consistent set of sentences to a maximal consistent set of sentences, containing witnesses and the constructed model is the particular one described by the formulae in the maximal consistent set. These sets are characterized as being a very precise and detailed description of a model.

In type theory, the lambda operator can define a constructible hierarchy, combined with the description operator’s ability to provide “formal beings”. Henkin intended to mentally represent the functions in the hierarchy of types that can be named by lambda expressions.

The procedure is as follows: “he considered a hierarchy of types with a universe given by the set of natural numbers and a constant as the name for the zero and one for the successor function. In the universe of subsets of the universe of individuals, there will clearly be both objects with a name and without one, because with a denumerable universe of individuals, the set of subsets is non-denumerable, but the sets with a name are denumerable. He then embarked on the recursive process of finding an object in the hierarchy of types as a denotation for each formula and the difficulties appeared with the selector $\iota_{\alpha(0\alpha)}$, whose interpretation should be a choice function, although it seemed too hard to identify a particular one” (from Manzano and Alonso 2014, 24).

---

[54] The compactness theorem states that a countably infinite set of quantificational formulas is satisfiable if and only if every finite subset of those formulas is satisfiable.

[55] For an accessible presentation of Henkin’s Completeness result, see Manzano 1999.

Henkin's interest was directed to the semantic side of type theory, trying to develop the hierarchy of objects with a proper name. Here, he realized that in the elimination of repetitions, he made use of lambda conversion rules and, thus, he used both the semantics and calculus. Then, he realized that there was a relation between formulae and sets denoting truth-values, which meant that, in order to reduce the universe of objects named by propositions (that is, the truth-values) to only two, the set of axioms had to be expanded to a maximal consistent set, so that the equivalence classes induced by the relation $\vdash M^{\alpha} = N^{\alpha}$ would also be reduced to only two (see Manzano and Alonso 2014, 25).

So, "THEOREM VI: *If S is any formally consistent set of T sentences, then there is a denumerable general model $\mathcal{M}$ of T for which each sentence of S is true. (Each domain D_n of $\mathcal{M}$ is denumerable)*" (Henkin 1996, 141).

The set of valid formulae is so large because the concept of a standard structure is too restrictive. However, if we are willing to accept nonstandard structures for the theory of types, then the set of valid formulae is correspondingly reduced. In a non-standard model, the universe for monadic variables is a subset of the power set of the universe of individuals, and similarly for other higher-order variables (Manzano and Alonso 2014, 25).

Henkin's proof is based on the extension of the original set of sentences to a maximally consistent set. The model is also built on this extended set of sentences, but in first-order logic we need not build the type hierarchy, since only one universe is needed. The detailed description of the behaviour of the non-logical constants of the formal language provided by the maximal consistent sets solves the issue of their interpretation.

The fundamental of his completeness theory is condensed in the first two theorems.

Theorem 1 (see Henkin 1996, 134) states that any set *S* of *L*-sentences (well-formed formulas without free variables) that is fundamentally consistent in the deductive system of *L*, is satisfied by some denumerably infinite *L*-structure $\mathcal{M}$.

From this theorem, three corollaries are obtained[56]:

The first one asserts that the first-order pure functional calculus is complete, so every logically valid *wff* is provable by using the formal axioms and rules of inference in *L*. Henkin points out that this result collides with Gödel's (1929) completeness theorem, so Gödel's completeness theorem becomes a special case of Theorem 1, proved by Henkin. Therefore, the theorem expresses *strong completeness*, whereas the corollary expresses *weak completeness* (which is the one proven by Gödel).

The second corollary of the theorem corresponds to the Löwenheim-Skolem theorem. According to the theorem, if a set of closed *wffs* of *L* is satisfied in some *L*-structure, then it is satisfied in a denumerably infinite *L*-structure.

According to the third corollary, a set *S* of wffs of *L* is simultaneously satisfied in some *L*-structure $\mathcal{M}$ if and only if every finite subset S_1 of S is satisfied in some *L*-structure $\mathcal{M}_1$. This corollary expresses the property of compactness for first-order logic. As we know, this tool is very useful in model theory. Besides, as Henkin himself asserts, the strong completeness theorem for first-order logic can be extended to the case of first-order logic with identity.

Theorem 2 (see Henkin 1996, 136) states that if *L* is any consistent set of sentences from an applied first-order L calculus (which can contain identity), extended, then there is an *L*-

---

[56] See Henkin 1996, 134-135.

structure, $\mathcal{M}$, which has a domain of which the cardinality does not exceed the cardinal number of the set of all L-symbols, which satisfies every sentence of S.

This theorem also presents a corollary: A set S of sentences of a system L of the kind described in Theorem II is satisfied in some structure $\mathcal{M}$, if and only if every finite subset S_1 of S is satisfied in some structure $\mathcal{M}_1$. This corollary expresses the compactness for sets of first-order sentences of any cardinality. Hence, compactness is extended to infinite sets of first-order sentences.

To conclude, in a deeper analysis of how Henkin's completeness proof came to light, it is notable that, contrary to what one might think, the discovery of the proof of completeness for type theory came before the discovery of the proof for first-order logic. In fact, the discovery of the completeness proof for first-order logic is the result of a modification of the proof for type theory, which Henkin already had. However, as we know, the order of publishing is the other way around: "The Completeness of the First-Order Functional Calculus" (1949), and later "Completeness in the theory of types" (1950). In the former, we can find his version of the completeness proof for first-order logic.

2.4 Completeness and explanatory power

It is widely accepted that Henkin's proof has more explanatory power than Gödel's. Let us review some of the reasons for this idea, and for the fact that we currently use Henkin's proof when we study completeness and only a few even encounter Gödel's version.

Henkin himself argues that this new method for obtaining the proof presents two main advantages: (1) an important property of the formal systems associated with completeness can be generalized to systems which contain infinite and non-denumerable primitive symbols, which is interesting in abstract algebra; and (2) the proof suggests a new approach to the problem of completeness in functional second-order calculi (see Henkin 1949, 159).

These two advantages may not have been thought of as directly related to the explanatory power of the new proof. However, as we will see in Chapter 5, these advantages have important consequences for the explanatoriness of the proof, since the ability to unify different areas and to provide general explanations which highlight features that different systems have in common also contribute to the explanatory power of a proof in a relevant way.

Henkin's completeness proof is more than just a new proof; it is also a new method that has been proved to be very productive. Moreover, it has provided a new approach to the question of completeness in second-order logic, an approach that goes beyond that of Gödel's.

This does not make Gödel's proof less relevant in any sense. As we know, Gödel's proof is of great significance in that it opens a new field where semantic notions acquire a status they did not have before. Gödel establishes the difference between semantic and proof-theoretic notions, with respect to their metatheoretical behaviour (see Manzano and Alonso 2014, 8). In fact, Tarski later took advantage of this framework to provide mathematical definitions of semantic notions (model, class of admissible models in a language, semantic consequence…).

Gödel's result is strongly dependent on the mechanisms that Lowenheim and Skolem got a few years earlier. In these mechanisms, the language of only first order is used to construct realizations, which we now refer to as models. However, Gödel never used semantic notions that are now basic.

In Gödel we can also find a strong completeness result, which is obtained from compactness. Henkin proved this theorem in order to obtain his completeness theorem in the version we currently use.

With the mathematization of semantics introduced by Tarski and the generalization of the completeness proof that derived from Henkin's method, it was possible to export things to other domains of logic, leading to the recasting of the whole discipline.

The key feature of the proof is extending *L* to a new language L_1, by adding an infinite sequence of individual constants:

> "The proof of Theorem I shows how, starting with a consistent set *S* of formal sentences of a first-order language, one can obtain a model *M* satisfying *S* by using newly adjoined individual constants for the elements of *M*. Conversely, starting with a structure *M* for a first-order language, the set of all sentences true of *M* is a (maximal) consistent set; and if we adjoin individual constants to the language to serve as names for the elements of *M*, then for each true sentence of form ∃xFx there will be a constant *u* such that *Fu* is true of *M*. Observing how one can go back and forth between consistent sets *S* of sentences and models of **S,** led me to my next discoveries-Theorem II of my dissertation (extending Theorem I to non denumerable languages), its important Corollary (compactness), and my proof of Theorem III (representing Boolean algebras). (Henkin 1949: 152-153)

The main idea for constructing the model for first-order logic is to extend the consistent set of sentences to a maximally consistent set of sentences with witnesses. The constructed model is the one described by the formulas of the maximally consistent set.

There was a relevant realization process with the transformation of the concept of an axiomatic theory. This realization included that a given system of symbols and sentences can be given several interpretations, so that a single language could refer simultaneously to many domains. As Henkin himself points out[57]: "The realization that sentences proved in an axiomatic theory give information simultaneously for a great many domains has had a revolutionary effect on both pure and applied mathematics. As regards applications, it meant that by moving to a more abstract level one could achieve a great economy of effort, handling problems from diverse domains by means of a single investigation".

Manzano and Alonso (2014, 118) point out that the process came with a considerable degree of abstraction, this included "on the one hand, the realization that the nature of the objects that constitute the universe of a structure is irrelevant and, on the other, that what matters is the relationships that hold these objects together. In this branch of mathematics the shift from common single mathematical structures, such as elementary arithmetic, to the groups, rings and fields and from classes of structures to abstract algebra was made effective". This is in correspondence with the three senses of abstraction distinguished by Jansson and Saatsi (2017)[58]: the abstraction from the particular entities, the abstraction from mathematical laws that hold for those entities and the abstraction that is related to regularities about structures, which is directly linked to explanatoriness by the authors[59].

Nowadays, the completeness of a theory is understood as a syntactical property related to the logical calculus and its rules, that of being maximal or being full. Manzano and Alonso (see 2014, 19) argue that the notion of the completeness of a deductive calculus, as we understand it today, was born as a generalization of the concept of the completeness of a theory.

So, another contrast is related to the languages that are targeted in one proof and the other. Gödel solved the issue for first-order logic positively and negatively for any ω-consistent

---

[57] Henkin 1967, 24-25 (Quote taken from Manzano and Alonso 2014).

[58] See Chapter 2 of the present work.

[59] This parallelism was suggested by Concha Martínez Vidal.

recursively axiomatized logical system in which arithmetic could be embedded. The lambda calculus for the theory of types together with the usual semantics over a standard hierarchy of types is strong enough to express arithmetic, thus it had to be incomplete (see Manzano and Alonso 2014, 21). However, Henkin showed that if we accept other hierarchies of types that do not necessarily contain all the functions but contains the definable ones, it can be shown that all consequences of a set of hypotheses are provable in the calculus. That is, with this new semantics, the valid formulae are reduced to coincide with the formulae that the rules of calculus generate.

The completeness theorem for first-order logic, as it appeared in Gödel's thesis, is generalized by Henkin in three ways highlighted by Manzano and Alonso (2014, 27). First, a finite or denumerably infinite set of constants is admitted, thus going to an applied first-order functional calculus. Second, there is an extension to cover the case for first-order logic with identity. Third, the proof for the applied first-order functional calculus admits a non-denumerable set of individual and predicate symbols.

Henkin's new perspective also has some relevant consequences regarding the nature of the objects that constitute the domain of the model, the applications to mathematical theories and its non-constructive character, as Manzano and Alonso (2014) point out. As for the nature of the objects that constitute the domain of the model, we should note that Gödel thought of the Completeness Theorem as a statement that was specific to first-order logic.

Now we know that it is a type of theorem applicable to other domains. This was achieved when Henkin uses models, which have the feature that the nature of the objects that constitute their domain is purely formal, so that the first-order logic model no longer has the natural numbers as its domain. General models admit hierarchies of sets, and the domains of those hierarchies can be constructed and described in the formal language, which is related to the universe of constructible sets in the realm of set theory. This is one of the features of Henkin's proof which makes it adaptable to many other logics, and his method has been extended to other logics and has become a standard procedure nowadays.

Henkin's vision requires a high level of abstraction, since it is a non-constructive proof[60]. The non-constructive feature of the proof is assumed by using the analysis of the theory of types as the starting point. The proof shows a much higher level of abstraction than was usual at the time. Besides, Henkin's new proving method provided some crucial tools for the construction of semantics as we understand it nowadays, associated with non-constructive reasoning and independent from proof theory. According to Manzano and Alonso (2014, 8), there are three main contributions to this phenomenon: the completeness proof by Gödel (1930), Tarski's introduction of semantic notions in a mathematical form, and the completeness proof (together with its method) by Henkin.

As we have already mentioned, it is necessary to have a clear perception of the independence of derivability and semantic consequence, supported in its different metatheoretical behaviour, to reach the present conception of logic. This distinction could not be conceived without the decisive contribution by Tarski and Henkin.

Some of these points are also addressed by Peter Koepke, in a paper titled "Gödel's Completeness Theorem with Natural Language Formulas" (2007). In the paper, Koepke argues that there is a way in which Gödel's completeness theorem is essential to other results, such as Gödel's incompleteness theorem, which is often considered a much more important result for arithmetic (2007, 13-14).

---

[60] This is notable, since Alonzo Church, his thesis advisor, was representative of the constructivist tradition in mathematics.

Koepke also addresses the broad spectrum of applications of Henkin’s version of completeness due to its universal applicability to first-order theories. He even lists some of the areas of application of the completeness theorem: mathematical applications, mathematical correctness and proof checking, formalization in other areas, automatic theorem proving, artificial intelligence and non-classical logics (see Koepke 2007, 14). Below is a summary of these applications.

As for the **mathematical implications** of the completeness theorem, Koepke also points at the fact that the proofs have immediate consequences in first-order logic results such as the compactness theorem or the Löwenheim-Skolem theorems. These can be used to construct structures with particular properties, which has many algebraic and other applications (nonstandard analysis, for instance).

The completeness theorem also provides an absolute criterion for **mathematical correctness and proof checking**. A mathematical proof is correct if and only if it can be reformulated as a formal proof, which serves as a criterion for correctness even for our more common informal or semi-formal proofs.

As for **formalization in other areas**, the formal method used in the proof has provided some motivation for other sciences to formalize their statements and methods, which presents some correlation to the tendency to present data in a way that can be operated on by algorithms.

In this vein, the completeness theorem serves as a guarantee that, in principle, all universally valid statements can be **proven automatically**, by checking whether the enumeration of all possible proof texts contains a proof of the valid statement. Of course, this method is not used, due to the size of the set in which we would search for the proofs. However, nowadays there are practical automatic theorem provers for limited domains.

Another area of application for the completeness theorem is **artificial intelligence**. Here, in the beginning it was characterized by logical formalization.

Lastly, while first-order logic is considered to be the optimal logic for having a complete, decidable proof calculus, many **non-classical logics** have been created departing from first-order logic, so we can find modal logics, temporal logics and algorithmic logics.

Many of these applications have to do with the fact that Henkin’s theorem and its proof provide a high degree of unification of domains, which yields a proof that is applicable to cases other than the original one. Apart from this, but related to it, Henkin’s proof has also served as an inspiration for other areas of research, in what has to do with the formalization of proofs and, more in general, formulation in other domains and, moreover, the development of other disciplines, such as artificial intelligence, which depended on logical formalization at early stages.

3. Computers and proofs: The Four-Colour Theorem

3.1 The Four-Colour Theorem

In the paper "The Four-Color Problem and its Philosophical Significance" (1998), Tymoczko addresses the proof of the Four-Colour Theorem. The proof was first announced in 1976 and then published in 1977 by Appel, Haken, and Koch.

The Four-Colour Problem is characterized by Tymoczko in the following way: “The old four-color problem was whether every map on the plane of sphere can be coloured with no more than four colors in such a way that neighbouring regions are never coloured alike” (Tymoczko 1998, 246). The problem resisted numerous attempts by mathematicians for quite a long time. It had already been proved that five colours suffice to colour a map, but still no single map was ever found that required more than four colours. Following this lead, some mathematicians thought that four colours were not sufficient and worked on methods to produce a counterexample, of course, without success. Finally, Kenneth Appel and Wolfgang Haken, assisted by John Koch, published a proof that, in the end, four colours do suffice. This is how the four-colour problem led to the four-colour theorem (4CT) (see Tymoczko 1998, 246).

3.2 Proofs and computers

This proof would not be so relevant, and its implications would not be an issue were it not for the fact that the proof was (partially) performed by a computer, particularly the part consisting of combinatorial checking, in the context of a proof by *reduction ad absurdum*. For this proof, the number of combinations that need checking is so large that the task cannot possibly be performed by *human means* alone. This inevitably brings up the debate about computer proofs in mathematical practice.

In this debate, Tymoczko's claim is that acceptance of such computer proofs forces us to adopt a *quasi-empirical* account of mathematics: “[…] all such evidence is quasi-empirical and not at all like the a priori constructions to which foundationalists restrict mathematical proofs. It is not surprising that some foundationalists refuse to admit computer proofs to normal mathematics on the grounds that it would change the fundamental character of mathematics. According to this view, which Tymoczko rejects, mathematics can and does incorporate some empirical elements into its practice.

In contrast, Tymoczko’s view is in correspondence with a formalist or foundationalist context where proofs are expected to fit in *a priori* constructions in order to guarantee their conclusions. However, the method of checking proofs by verifying that each and every inference is correct does not seem to apply to computer proofs involving a huge number of steps or even inferences that humans cannot perform. As Tymoczko (see 1998, 243) points out, hard copies of this kind of proofs can be unobtainable due to the amount of computer time that would be required to print them out or even with the number of steps of those proofs they would be utterly useless to mathematicians.

Tymoczko goes even further and makes a stronger claim: the use of computers makes the 4CT proof the first one that is empirical. This means that the Four-Colour Theorem would be the first mathematical proposition to be known *a posteriori*, which raises the problem of the distinction between mathematics and the natural sciences (see Tymoczko 1998, 246). His thesis is that if we accept this proof, some of the following commonly accepted beliefs about mathematics would have to be modified:

> 1. “All mathematical theorems are known a priori.
> 2. Mathematics, as opposed to natural science, has no empirical content.
> 3. Mathematics, as opposed to natural science, relies only on proofs, whereas natural science makes use of experiments.
> 4. Mathematical theorems are certain to a degree that no theorem of natural science can match” (Tymoczko 1998, 250).

Tymoczko’s view is that the proof of the Four-Colour Theorem is convincing but not surveyable, in the sense that it is not verifiable step by step. It resembles an ordinary proof but with a lacuna where a key lemma is justified by *non-human work*, that is, by computer. Therefore, the belief in the truth of the theorem, Tymoczko argues, depends on two interrelated factors: the reliability of the machine and the reliability of the program. The reliability of the machine is ultimately a matter for engineering and physics to assess, whereas the task of evaluating programs is a topic of computer science (see Tymoczko 1998, 258).

Tymoczko concludes that accepting the 4CT forces us to modify our concept of proof, and that the 4CT would be an example of an *a posteriori* necessary truth and, a fortiori, a counterexample to the claim that all known necessary truths are known *a priori* (Tymoczko 1998, 261).

This view is not universal. Let us address some different approaches to the issues raised by the 4CT. Teller, in “Computer proof (1980), questions both main conclusions of Tymoczko’s analysis: (a) that we have to change our concept of proof if we accept the Four-Colour Theorem, and (b) that the use of computers in mathematics introduces empirical experiments into mathematics, forcing us to address the distinction of mathematics from the natural sciences.

It can be suggested that the use of a computer in the 4CT could be just an extension of our means of surveying and not a change in our concept of proof, and Teller believes that is the case: “the fact that I cannot follow a complex proof produced by a good mathematician does not show that such a mathematician's complex proof is a proof in a different sense of the word from a proof that I can follow. In the same way, the fact that no mathematician may be able to follow a proof produced by a computer does not show that such a computer-produced proof is a proof only in some new sense” (Teller 1980, 800). Teller has a point since, after all, Tymoczko provides us with no grounds for establishing that the fallibility of computers and mathematicians are relevantly different.

Moreover, Teller argues that there actually is a difference between experiments in natural sciences and in mathematics: an experiment in the natural sciences determines a spatiotemporal fact, which then may or may not be generalized beyond the local place and time to which it directly applies, whereas a correct mathematical proof establishes a kind of fact that is non-spatiotemporal (see Teller 1980: 801-802).

A slightly different view is held by Michael Detlefsen and Mark Luker, in a paper entitled "The Four-Color Theorem and Mathematical Proof" (1980). They agree with Tymoczko's claim that evidence of an empirical sort is utilized in the proof of the 4CT. However, they defend that this is not a novelty. The aim of their paper is to show that mathematical proofs involving calculation often depend upon empirical evidence (see Detlefsen and Luker 1980. 804). With that in mind, they present several examples of machine-assisted proofs which would be unsurveyable according to Tymoczko, and are antedating the proof of the 4CT. Then, they argue that "the only difference between the older and the newer appeals to "by computer" is in the nature of the computer. Traditionally, the "computer" was a human mathematician. These days, it may also be the likes of an IBM 370-160A. The essential epistemological quality of the appeal is, we shall argue, the same" (Detlefsen and Luker 1980, 805).

In their analysis, they distinguish four separate assumptions required for confidence in the result of a computation: (a) that the underlying algorithm to be used is mathematically sound, (b) that the particular program to be used is a correct implementation of this algorithm, (c) that the computing agent correctly executes the program and (d) that the reported result was actually obtained.

Parts (a) and (b) are subject to deductive proof, at least in principle. Parts (c) and (d), however, are not. The belief in their validity ultimately rests on empirical considerations, whether the calculation is performed by an IBM 370-160A or by a human mathematician (see Detlefsen and Luker 1980, 808).

The conclusion is that computers are fallible in a similar way that humans are. Human processing is subject to such things as fatigue, limited knowledge or memory, and to the psychological desire to force a particular result to "come out", among many others.

If we take into account the reliability of the proof (given that Tymoczko argues that a proof is a construction that can be looked over, reviewed, and verified by a rational agent), Detlefsen and Luker argue that the reliability of first-hand survey is just as incapable of proof as the reliability of a given computer: "survey is no guarantee of correctness, as many cases in the history of mathematics show. Kempe published a mistaken proof of the 4CT that remained uncriticized for eleven years. Cauchy, Lame, and Kummer all believed that they had proved Fermat's Last Theorem at one time. And Rademacher claimed in 1945 that he had solved the Riemann hypothesis" (Detlefsen and Luker 1980, 816).

Hersh (1997) agrees with the idea that mathematical proofs are a posteriori, whether they use computers or not:

> "But despite the illusions of idealist philosophy, old-fashioned person-made proofs are also a posteriori. They also depend on credence in the world of experience, the material world. We believe our scribbled notes don't change from hour to hour, that our thinking apparatus (brain) is reliable more often than not, that our books and journals are what they pretend to be. We sometimes believe a proof without checking every line, every reference, and every line of every reference. Why? We depend on the integrity and competence of certain human beings—our colleagues" (Hersh 1997, 153).

In the end, human and computer proving are not such different procedures. Moreover, Hersh concludes that the use of computers in pure mathematics has increased and will continue increasing: "In fields like chaos, dynamical systems, high-Reynolds-number fluid dynamics, we come to trend B. Here, as in number theory, the computer is an explorer, a scout. It goes much deeper into unknown regions than rigorous analysis can. In these fields we no longer think of computer-obtained knowledge as tentative—pending proper proof. We prove what we can, compute what we can" (Hersh 1997, 158).

As final remarks, I would like to point out that there is something interesting and attractive about Detlefsen and Luker's idea that computers do nothing significantly different than what humans do when they perform a proof, in the sense that it could be argued that, ultimately, computers just process data in a similar way that we humans do when proving a theorem, but at a much higher speed. The idea that computers are fallible in a similar way that humans are has some attractiveness. This could suggest that the work of the computer could simply be an extension of human powers. This point would undermine both the novelty of computer proving and the idea that we need to modify or adapt our concept of proof in order to account for these cases, thus undermining Tymoczko's theses.

In fact, even Tymoczko admits that the belief in the validity of the proof is reasonable: "In the case of 4CT, most mathematicians feel that the reliability is sufficiently high to warrant a qualified acceptance of the theorem. In the first place, the problem was reducible to computer-manageable complexity. There is a very clear idea of what the computer is supposed to be doing –we have a good understanding of reduction techniques. Moreover, there is a great deal of accumulated evidence for the reliability of computers in such operations, and the work of the original computers was checked by other computers. Finally, there is good reason to believe that the theorem could not be reached by any other means" (see Tymoczko 1998, 258).

From this discussion, it seems reasonable to conclude that the proof of the Four-Colour Theorem is a genuine proof in that it establishes the truth of the theorem. Another question is whether it constitutes an explanation or not. Even though there will be a deeper analysis of this issue in Chapter 5, let us anticipate the main thesis regarding this issue, which can be summarized by stating that, in terms of explanation, the Four-Colour Theorem proof can be considered an example of a non-explanatory proof. That is, although the proof establishes that the result is a theorem (hence, accomplishing its *mission* as a proof), there is a significant part of the procedure that is opaque to us, which causes the proof's failure to illuminate or provide the reasons why the theorem is true and, hence, we can colour any map with just four colours. This *opaque part* is the checking task performed by the computer that cannot be performed or even thoroughly checked by purely *human resources*. From the proof, we know that we can colour any map with four colours, but we remain not grasping *why* that is the case.

It is true that the task performed *by computer* is a merely combinatorial one, but we would still like to know exactly how that task was performed. Moreover, we would probably like to be able to reproduce such task. This suggests that, when proving and/or explaining a phenomenon, we do not just gather the data, we also aim at explaining them. Ultimately, the problem with the 4CT is that *we want to understand the proof, not just be told there is one.*

It is also notable that some think that there is reason to believe that the theorem could not be reached by any other means, which seems to point at the idea that it could be the case that some theorems lack a comprehensible explanation.

Perhaps, the fact that the computer does such a central work in the proof is what is behind its non-explanatoriness. Note that there are more proofs sharing this feature, such as the proof of Gödel's Incompleteness Theorems by Lawrence Paulson (2014)[61]. This is another example of a computer-assisted proof which lacks explanatory power.

For our purposes, we will conclude with the thesis that not all proofs are explanatory, the Four-Colour Theorem proof being a clear example of this case.

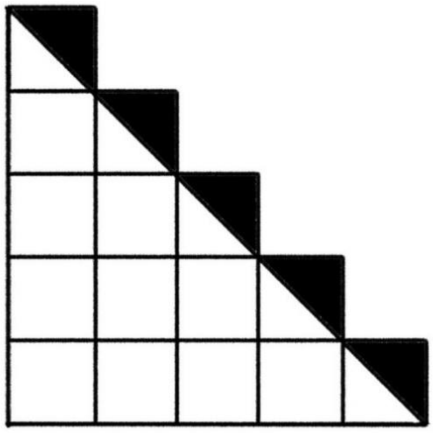
---

[61] Paulson, Lawrence. (2014), "A Machine-Assisted Proof of Gödel's Incompleteness Theorems for the Theory of Hereditarily Finite Sets", *The Review of Symbolic Logic*. 7. 484-498.

4. Visual Proofs and Visual Thinking

4.1 Triangular numbers

A triangular number is the sum of the first *n* positive integers, for some positive integer *n*. The first triangular number is 1, the second is 1+2, and in general the n^{th} is 1+2+...+(n−1)+n. Suppose you want to find the n^{th} triangular number, for a certain *n*. If *n* is very small one can just add the numbers, but for large numbers we will use the general formula. Perhaps, with the following diagram, the formula becomes a bit *less miraculous*:[62]



Theorem: 1 + 2 + 3 + ... + n = $n^2/2 + n/2$

This picture is not a traditional rigorous verbal or symbolic proof. It rather looks like some sort of mental or informal mathematical induction, which does not correspond to an inductive proof in the strict sense either. In addition, it is an obviously limited version of the so-called "proof", in that it only shows the case when *n* is no bigger than 5. Still, by looking at the picture and with a little context, we seem to grasp the generality, that is, we can infer that the theorem will be valid for 6, 7… just by *extending the diagram in our minds* (which contributes to our grasping of the generality). Therefore, we become convinced that it is true of all numbers – it is a theorem. This brings the question of whether a picture can constitute a mathematical proof.

4.2 Are "picture proofs" genuine proofs?

James R. Brown (2012) argues that this is a proof in its own right. In fact, he sees the diagram as an argument supporting his idea that we can infer theorems from pictures. That is, by means of mathematical intuition, we can grasp the truth of this theory.

---

[62] Picture *proof* taken from: Brown 2012, *Platonism, Naturalism and Mathematical Knowledge*, Routledge, p. 201.

In order to provide some context to the debate, it can be argued that accepting pictures or diagrams as real proofs might lead to accepting that visual thinking and inference from a diagram or picture amounts to proving in mathematics, equivalently to formal thinking and traditional deductive reasoning.

Interestingly, Brown (2012) argues that the information provided by the picture is not enough for us to grasp the proof. We need to add "mathematical intuition" to the equation, that is, our ability to obtain knowledge of mathematical objects that goes beyond the senses and allows us to have access to the mathematical realm[63]. As Brown argues, the picture is not just a representation in the usual way; it is something like a telescope that helps us "see" things in the platonic realm. In his view, just like devices such as telescopes, the picture proof is a device that facilitates our mathematical intuition.

There are many references to mathematical intuition in the literature. Gödel himself argued that "the assumption of such objects is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a satisfactory theory of our sense perceptions" (Gödel 1944, 456). Also, "we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have any less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception" (Gödel 1947, 484).

What is undeniable is that the case raises a rich debate concerning the issue of visual thinking, and whether a diagram or picture can constitute a proof. However, I would point at several problems with this perspective. Firstly, every tool is created for a purpose, and it is designed according to that purpose. Therefore, in this case, the picture is specifically designed for us to grasp and understand the theorem of triangular numbers with ease, so it is not necessarily about mathematical intuition, but some kind of auxiliary resource to facilitate our reasoning through the mathematical theorem. It is no coincidence that we grasp the truth of the theorem through a particular diagram (and not just *any* diagram), which implies that the facilitating role of the diagram is not independent from our sense experience. The diagram represents a limited number of cases, and it works as an aid in an abstraction process, through which we understand that the same structure could be extended to all natural numbers and, hence, we grasp that the theorem is true of all natural numbers.

In addition, we need not appeal to intuitions to account for this mechanism. It would be easier to explain it in terms of generalization or mathematical (informal) induction, even though it comes from a picture and thus it does not constitute a rigorous proof by induction. If we do not appeal to intuition, this case is not incompatible with a naturalist epistemology and it is not necessary to postulate a means of knowledge about which we know nothing or very little, and its mere existence is more than controversial. Therefore, Brown's argument is not as strong as it may appear in the first place and the appeal to intuition (a *mysterious* means of acquiring knowledge) could be unnecessary.

Of course, Brown's view is extreme and not shared by many. However, the opposed view is also rare, that is, the view according to which pictures cannot be part of proofs, arguing that proofs are syntactic objects constituted exclusively propositions (see Tennant 1986). Most approaches can be placed somewhere in between these extremes, by accepting the idea that a

---

[63] Of course, this idea is directly related to Brown's platonist approach to mathematics, where mathematical objects belong to an abstract realm, which is independent from the empirical world. Mathematical intuition is our way to have access to the abstract realm. This is an alternative way to mathematical platonism that is independent from the EIA debate.

picture or diagram can be a relevant component of a proof, but it needs to be accompanied and complemented by propositions or sentences in order to be considered a proof in the strict sense.

In this vein, Giaquinto argues:

> We can have cognitive grasp of some structures by means of visual representations. For small simple finite structures we can know them through sensory experience of instances of them, somewhat as we can know the butterfly shape from seeing butterflies. This is a kind of knowledge by acquaintance. Though we cannot have knowledge by acquaintance of any infinite structure, some simple infinite structures can be known by visual means, and not merely as the structure of models of this or that theory. The crucial representations in these cases are category specifications, which are not items of conscious experience. Nonetheless, through the images they give rise to they can give us a grasp of structure that is sometimes operative in practice. How useful this resource can be and how it links up with propositional knowledge represented in words or formulas is a matter for future investigation (in Mancosu 2008, 63-64).

Giaquinto (2007, 2008) goes into neuroscience results of experiments concerning different features of our mathematical knowledge. He argues against the prevalent view, according to which visual representations have (only) a facilitating role and, thus, they cannot be a resource for discovery, justification or proof, thus lacking any epistemic value in mathematics.

With this purpose, Giaquinto points out that there are three exclusive possibilities if we grant that there can be superfluous uses of diagrams in following a proof. Those three options are: (1) all thinking that involves a diagram in following a proof is superfluous, (2) not all thinking that involves a diagram in following a proof is superfluous; but if not superfluous, it is replaceable or (3) some thinking that involves a diagram in following a proof is neither superfluous nor replaceable. Giaquinto argues that (3) is correct, so there can be mathematical proofs in which there is visual reasoning playing a relevant (and non-replaceable) role in the proof. He argues against the idea that proofs containing pictures that play an important explanatory role would not be "proofs" in the strict sense.

Note that claiming that visual thinking has a significant role in mathematical reasoning and claiming that a picture or diagram can constitute a proof are two different things. Visual thinking includes external visual representations, such as all kinds of diagrams, computer images, pictures, symbol arrays… but it also includes internal visual thinking, such as imaginations of spatial transformations or visualizing representations of phenomena. This kind of thinking is often accompanied by non-visual reasoning, which provides a framework and concretion to those images and visual reasoning.

One of the main roles of visual thinking is, probably, that of aiding understanding. That role may be related to helping us grasp certain concepts found in a mathematical statement, or even helping us going through a proof, which is the case – in my opinion – in the *picture proof* reproduced at the beginning of this section. This means that this resource can be particularly useful in contexts of discovery.

Back to our *picture proof* and the issue of whether it constitutes a proof by itself or not, while I grant that the picture constitutes a much a more direct and immediate way of grasping the truth of the theorem, I do not endorse the view that it constitutes a proof in a genuine sense. While it is true that the formal proof would be far more complicated and it would employ a higher level of mathematics, the diagram is made to help us grasp the general principle from a limited number of cases and see that it can be applied to all the natural numbers. It is a tool designed for a purpose, and it succeeds at it.

As for explanation, there is no doubt that the picture has a powerful role in explaining the theorem. However, for it to constitute a proof, it needs some kind of context or explanation so that it makes sense and actually convey the generality and thus the truth of the theorem.

It may well be that we need to abandon the traditional idea that mathematical proofs are exclusively that finite sequence of formulae or sentences, each of which is an axiom, an assumption or follows deductively from those by the application of rules of inference, and the last (closing) formula is the conclusion. Or, at least, we might need to be ready to accept that other elements or complements can figure in a proof without making it less formal or less rigorous. The case of proofs containing (or accompanied by) pictures or diagrams would support this idea.

These visual elements have the aim to provide quick and easy comprehension. This fact is relevant to our research in that it highlights that sometimes proofs present complements which increase their explanatory power, and thus they constitute elements facilitating our reasoning and our working through a proof in order to obtain the understanding of a particular result. This explanatoriness may come in the form of making particular proofs easier to grasp or, in some cases, they can show a deeper level of explanatoriness more related to showing *why* a mathematical statement is the case. Therefore, diagrammatic reasoning can be a part of arguments, and visual thinking can have a relevant role in mathematical proving as it clearly does in mathematical explanation.

Finally, we should also keep an open mind as to other non-visual elements that could be part of proofs as well, like sounds or other resources. It is worth noting that visual elements do not always just facilitate our grasping of particular results. In some cases, they constitute substantive parts of the process of obtaining a proof (for example, in geometry).

5. Is mathematical induction explanatory?

One more issue that goes to the point of questioning the explanatoriness of (particular) mathematical proofs concerns mathematical induction.

As we know, a proof by mathematical induction proceeds in the following way:

> For any property P, if $P(1)$ and for any natural number n, if $P(n)$ then $P(\mathrm{n}+1)$, then for any natural number n, $P(\mathrm{n})$.

That proofs based on mathematical induction are valid is not an issue. The question is, rather: are proofs by mathematical induction explanatory?

The main idea that some have pointed out regarding induction and mathematical explanation is that, while induction constitutes a valid proving strategy and works in providing results in mathematics, proofs by mathematical induction fail to be explanatory, that is, they do not shed any kind of light for understanding purposes. In addition, they do not open new ways to understand a discipline or to systematize a particular area of research (in contrast to what we have analysed for Henkin's completeness proof).

For instance, Lange (2010a) argues that arguments by mathematical induction are not explanatory. His strategy is to establish a contrast between mathematical induction as is usually performed and some kind of induction that is upward and downward from a fixed number (different from 1). This form of induction reasoning is deductively equivalent to our ordinary induction:

> If a theorem can be proved by mathematical induction proceeding upwards from n=1, then generally it can also be proved by an argument like mathematical induction, but instead proceeding upwards and downwards from n=k for any other, arbitrarily selected natural number k. Of course, it is usually easier to begin by showing that P(1) rather than by showing, say, that P(5), since the proof of P(1) is often trivial. And it is obviously easier to show merely that if P(k), then P(k + 1), rather than to demonstrate not only this 'upwards' fact, but also the 'downwards' fact that if P(k), then P(k_1), for any natural number k>1. (Lange 2010a, 207)

The explanatoriness of ordinary mathematical induction should be the same in both cases. That is, if mathematical induction is explanatory, so should be this upward and downward reasoning Lange proposes, since the only thing that distinguishes them is where they start. He shows that following this route would ultimately fall into a circularity of mathematical explanations, which is unacceptable in mathematical explanation so, by reduction, proofs by induction are non-explanatory:

> It cannot be that P(1) helps to explain why P(5) holds and that P(5) helps to explain why P(1) holds, on pain of mathematical explanations running in a circle. Therefore, a mathematical induction does not explain the theorem that it proves if that theorem can also be proved by the 'upwards and downwards from 5' rule. Since generally any theorem provable by mathematical induction can also be proved by the 'upwards and downwards from 5' rule, mathematical inductions are generally not explanatory. (Lange 2010a, 209)

His conclusion is that, given that one of the premises about mathematical explanation is non-circularity, then proofs by ordinary mathematical induction cannot be explanatory.

There have been, however, some doubts about Lange's conclusion. For instance, Baker (2009) argues against Lange's claim that the proof by ordinary mathematical induction and the one Lange puts forward are equally explanatory, since the latter is more disjunctive than the former and it contains irrelevant elements to the explanation, so there is some asymmetry between both explanations.

In addition, Cariani 2011 also points at this asymmetry, although he suggests that there is no clear answer as to whether proofs by mathematical induction are explanatory. He argues that the difference between both explanations lies on simplicity, and it is simplicity what makes one proof more explanatory than the other one.

On the other hand, Baldwin (2016) argues that inductive arguments are explanatory and focuses his criticisms of Lange's argument on the assumption that an instance of a universal generalization cannot help explain the universal generalization.

Lange's argument may seem appealing. However, I would raise some doubts as to whether a question concerning explanatoriness can really be susceptible to a *reductio ad absurdum*. Besides, what Lange's version of induction shows is that the property is inherited by the rest of the numbers starting from a given place. However, if we do not assume that induction proofs have some sort of foundational structure, it is not necessary to accept the idea that the reason mathematical induction is non-explanatory is its circularity. It seems like something more needs to be said in order to obtain that conclusion. I am still inclined to agree with Lange that proofs by mathematical induction fail to be explanatory, but probably for different reasons. The hypothesis that this is the case because mathematical induction fails to account for the reason of why a mathematical property holds for all natural numbers, or that it fails to help us understand the relevant regularity, seems more likely[64].

From a different perspective, Hanna (2010) argues for the thesis that rigour is secondary in importance to understanding (see 2010, 6-7) and that a proof is convincing and legitimate only when it leads to mathematical understanding, which amounts to thinking clearly and effectively about mathematics. This ties mathematical proof and mathematical understanding in a very clear way. Hanna even suggests that the issue has crucial consequences in mathematical education, where, even though students are taught the nature and standards of deductive reasoning in order for them to know when a result has been established, mathematical proof makes its greatest contribution only when teachers are able to use proofs that convey understanding (see 2010, 7). In this context, Hanna refers to mathematical induction as one of those *unavoidable* non-explanatory methods, together with contradiction.

The question of whether mathematical induction yields explanatory proofs is still an open question. Even in the context where philosophers and mathematicians disagree about which proofs of a given theorem merely prove the theorem or also provide explanatory answers to *why* the theorem holds, the case of mathematical induction is one of the most controversial. In this subject, there is a great deal of dependence of philosophers and mathematicians on their own intuitions. Some appear to be quite confident that those arguments are explanatory, and others are confident that the contrary is the case, and there are no many *neutral* arguments for and against the issue.

---

[64] Lange (2017) himself has provided what probably constitutes more convincing grounds for the thesis that induction is not explanatory, based on his account of the explanatory power of proofs. This issue will be discussed in Chapter 5, section 3.

6. Conclusions

The main purpose of the present chapter was to introduce a few case studies that will be further analysed in light of a theory of mathematical explanation and of explanatoriness in general, in Chapter 5. For this reason, we find no big conclusions so far. However, there are some preliminary ideas we can extract from the presentation of the cases.

First, the particular cases of mathematical proofs or explanations in this work constitute relevant examples that might help us understand the concept of proof better, and also know the limits of such concept and its relationship with explanatoriness and mathematical explanation.

All these cases highlight different features of mathematical proof and explanation. The completeness case is a good example of explanatory proofs and the search for better and more explanatory proofs in mathematics. The 4CT case shows that mathematical proofs can be non-explanatory. The *picture "proof"* raises the question of diagrammatic thinking and the use of visual elements in mathematical proofs, Both the four-colour problem and the diagrammatic explanation raise the question of mathematical intuitions and *a posteriori* elements in mathematical thinking. Some examples that do not conform to what is traditionally considered a formal proof may lead to some doubts about whether those constitute proofs in their own right, or even some doubts around the concept of proof itself. Finally, issues around mathematical induction are also present in the debate about explanatory and non-explanatory proofs, since it has been pointed at as a *clear example* of proofs which lack explanatory power.

Mathematical proofs might have roles that go beyond establishing that a targeted theorem is true. Some proofs may establish relations between disparate mathematical areas, by unifying, connecting or subsuming them. This issue is also of interest in the study of mathematical explanation.

It is relevant to note that not all explanations in mathematics are proofs. In some cases, they are part of a conceptual recasting of a discipline, or even auxiliary explanations in contexts where the formal established proof (the one that showed the result) is too obscure or difficult to follow.

Gathering part of the discussion on the concept of mathematical proof (which constituted Chapter 3), these cases will be used in the search for an account of mathematical explanation (Chapter 5). They will hopefully help testing possible accounts for mathematical explanation in general, or even the possibility of achieving a unified account for both extra-mathematical and intra-mathematical explanations.

CHAPTER 5: In search of a unified theory: the counterfactual theory of explanation

1. Theories of internal mathematical explanation

As pointed out, one of the main aims of this work is to determine whether it is possible to provide a unified account of mathematical explanation that is suitable for all kinds of mathematical explanation, understood as a whole.

It is important to note that there are two senses in which we can seek a unified theory of mathematical explanation.

The first one comes with the classification presented by Reutlinger (2018) of monist, pluralist and reductionist approaches, related to the non-causal nature of certain mathematical explanations of empirical phenomena, a classification already presented in Chapter 1. However, let us briefly recall those three different strategies or approaches to mathematical explanation:

- Causal reductionism: the thesis that there are no non-causal explanations. This implies that all the alleged examples of non-causal explanations can ultimately be understood as causal explanations or, presumably, dismissed as not genuine explanations.
- Pluralism. In pluralist approaches, causal and non-causal explanations are covered by two (or more) different theories of explanation.
- Monism. There is one single philosophical account capturing both causal and non-causal explanations, since both kinds of explanation share a feature that makes them explanatory.

As mentioned above (see Chapter 1), there are some reasons why a monist account of mathematical explanation would be preferable, if one is available. Monist accounts apply to both causal and non-causal explanations, which connects with the idea that we tend to prefer unified and more general theories that can account for more diverse phenomena. This makes monism a more attractive approach to mathematical explanation, at least in principle.

It is worth noting that Reutlinger's classification is mainly directed at extra-mathematical explanation. However, it can be of interest to explore the possibility of finding or designing a monist account of mathematical explanations that can not only cover causal and non-causal explanations, but also intra-mathematical and extra-mathematical explanations.

This is, precisely, the other sense in which a theory of mathematical explanation can be unified, and it is related to the classification of mathematical explanations depending on their subject or area of application:

- **External mathematical explanations** are mathematical explanations of empirical phenomena, that is, mathematics is used to account for a phenomenon that is not mathematical.
- **Internal mathematical explanations** are mathematical explanations of mathematical phenomena. That is, mathematics is being used internally, to account for other mathematical facts.

As has been pointed out, it seems reasonable, at least in principle, to prefer a theory that accounts for all kinds of mathematical explanation, that is, a theory that can be applied to both the cases of mathematical explanations of empirical phenomena and mathematical explanations of mathematical phenomena. This will, hopefully, provide us with some clues on the role of mathematics in explaining and what it is that makes mathematics a useful explanatory tool in general.

That being our aim, let us review some of the theories that we examined in chapters 1 and 2 -those were fundamentally theories of external mathematical explanation – and see if they can be applied to cases of internal mathematical explanation. Among them, the counterfactual theory of explanation looks like a good candidate as a monist account of all kinds of mathematical explanation. Of course, it is yet to discuss whether the theory can cover all of them.

Interestingly enough, internal mathematical explanation has been treated very differently from external mathematical explanation when it comes to designing a theory that accounts for it. That is, if we go to the literature, it may appear that internal and external mathematical explanations are two very different and completely disparate uses of mathematics, which cannot be accounted for with the same kind of approach. Indeed, from the theories that have arisen, both subjects seem to have very little in common, apart from the use of the same structure that is mathematics.

Let us now see some of the classical approaches to internal mathematical explanation that can easily be found in the literature.

1.1 Classical approaches to internal mathematical explanation

If we go through the literature, it is unavoidable to encounter two main classical approaches to mathematical explanation: those put forward by Steiner and Kitcher. Steiner’s account is based on the concept of “characterizing property”, and Kitcher’s is based on unification as the main role of mathematics in explanation.

These accounts have received a good amount of attention in recent years and have been used as starting points or just general frameworks when analyzing the explanatory power of mathematical explanations. A map of views on explanation would definitely be incomplete without referring to these accounts, so let us briefly characterize them:

Steiner’s approach

Steiner (1978) argues that explanatoriness is primarily a *local* property of proofs.

What distinguishes an explanatory proof from a non-explanatory one is that only the former involves a characterizing property. By “characterizing property”, Steiner means a property unique to a given entity or structure within a family or domain of such entities or structures, where the notion of family is taken as undefined. In his own words: “an explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that from the proof it is evident that the result depends on the property” (see Steiner 1978, 144, 147). Also, explanatory proofs exhibit generalizability, that is, varying the relevant feature (and hence a certain characterizing property) in such a proof gives rise to an array of corresponding theorems, which are proved—and explained—by an array of deformations of the original proof.

Therefore, Steiner points at two criteria for explanatory proofs, i.e. (i) the dependence on a characterizing property and (ii) the generalizability through variations of that property.

Kitcher's approach

Kitcher, in contrast, argues that explanatoriness is a *global* property of proofs.

In general, according to Kitcher (1981), to explain is to unify. He argues for an account of scientific explanation as theoretical unification, which also applies to the case of mathematical explanation.

His focus on explanation and understanding is clear when he argues that a theory of explanation should account for how science advances our understanding of the world and help in the evaluation of disputes in science. For this, he proposes a unification account. This account is, in a way, inspired by Friedman: "this is the essence of scientific explanation—science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than one with more" (Friedman 1974, 15). According to this, understanding of the world is achieved by reducing the number of facts we take as brute.

Thus, this is the general schema for Kitcher's account:

Let us take a set *K* of beliefs that are assumed to be consistent and deductively closed. We can think of this, informally, as a set of statements endorsed by an ideal scientific community at a specific moment in time (Kitcher 1981, 75). A systematization of *K* is any set of arguments that derive sentences in *K* from other sentences of *K*. The explanatory store over *K*, E(*K*), is the best systematization of *K* (Kitcher makes an idealization by claiming that E(*K*) is unique). There are different degrees of unification that correspond to different systematizations, and the highest degree of unification is that provided by E(*K*).

Finally, in order to judge which systematizations are the best, there are three factors: (i) the number of patterns, (ii) the stringency of the patterns and (iii) the set of consequences derivable from the unification.

1.2 Alternative approach: explanatory virtues

As we know, there is a different general strategy in the search for an account of mathematical explanations: appealing to explanatory virtues without necessarily placing them into a previously established theory of explanation.

There is a quite detailed analysis of explanatory virtues in Chapter 1, Section 6.1, which we will not repeat here. However, the present section is more than just a reminder of the approach or even the particular explanatory virtues at issue. The relevance of the subject is that explanatory virtues also come to play when we address internal mathematical explanation. Indeed, it is highly likely that the features of mathematics that account for its ability to explain are the same in external and internal mathematical explanation.

Let us just recall the explanatory virtues that were found more referenced in the literature in order to set the framework for the analysis of intra-mathematical explanations.

Two of the most repeated explanatory virtues in accounts of mathematical explanation are scope and topic generality, which can take different names. They make reference to the level of abstraction that mathematics allows in explanations. As we have seen, Baker (2016) argues that topic generality allows for *unificatory power*, so another central explanatory virtue would be unification. Baron (2017) also refers to it as unity and Kitcher has unification as the central concept in his theory of explanation. This gives the idea that the power to unify explanations under a smaller number of principles (or premises) is one of the most important features of good mathematical explanations, and of explanations in a broader sense.

In addition, simplicity and explanatory depth (which is sometimes characterized as what gives us the *why* of the *explanandum*), as we have seen back in Chapter 1, are the other two explanatory virtues that have received most attention in the literature. These capture other relevant features of good mathematical explanations.

The idea to be highlighted in this context is that the explanatory virtues we analysed in the discussion of the application of mathematics to explanations of empirical phenomena will be shown to also apply to the cases where we use mathematics to account for other mathematical facts.

2. The Counterfactual Theory of Explanation

After having motivated the interest in obtaining a version of the counterfactual theory of explanation that can account for all kinds of mathematical explanations, let us recall how Knowles and Saatsi (2019) present their account and which possible ways of manipulating and adapting the theory are available in order to make it suitable for our purposes.

This account makes reference to two senses in which mathematics provides generality (or, in a way, unification): topic generality and scope generality. These properties collapse with the two senses of generality discussed in the previous section about explanatory virtues. However, it is worth noting that the counterfactual theory includes them in the analysis of mathematical explanation. Topic generality refers to the fact that the explanation has nothing to do with the nature of the things involved, and in order to achieve topic generality, the explanation must abstract away from what kinds of things figure in the explanation. On the other hand, scope generality refers to the fact that mathematical explanations are equipped to provide answers to counterfactual questions in order to consider variations of the *explanans* variable. Taking natural numbers, the theory accounts for counterfactual questions in order to consider any natural number *n* as a possible value of the *explanans* variable. Here, the mathematical apparatus makes the difference by making it easier to abstract from the particular case. Both are ultimately ways to unify explanations and make them, in principle, simpler, by using the same resources to explain more particular cases.

As we know Knowles and Saatsi's concept of explanation within the CTE is that explaining consists of providing information of systemic patterns of counterfactual dependence, where counterfactuals capture objective, mind-independent modal connections that show how the value of the *explanandum* variable depends on the value of the relevant *explanans* variable(s). This idea is in correspondence with our way of reasoning by indicating how the *explanandum* would have been different, had the *explanans* been different, so it complies with our intuitions regarding the virtues of explanations in terms of the counterfactual information they provide. According to this, we see as more explanatorily powerful those explanations that provide more answers to what-if questions, compared to other explanations.

The account highlights three explanatory virtues in good mathematical explanations: non-sensitivity (scope generality), the degree of integration (topic generality) and cognitive salience, of which cognitive salience is what ultimately what makes mathematics explanatory.

2.1 The counterfactual theory and the cases

Before taking the cases we presented in Chapter 4 and using them to illuminate relevant aspects of intra-mathematical explanation (and explanation in a more general sense), there are two separate, but related, issues to keep in mind.

On one hand, there is the issue of finding out whether there is one account of mathematical explanation that can be turned into a good approach to those particular cases. On the other hand,

and relatedly, there is the issue of showing that these particular mathematical explanations share the same explanatory virtues.

In the search of a unified account of mathematical explanation that can accommodate both internal and external mathematical explanation, the CTE seems to be a good candidate. It has been shown that it can succeed in extra-mathematical explanation, and now it is the turn of seeing whether it can be applied to intra-mathematical explanation.

Consequently, this section has two main aims.

First, we will try to show that the Counterfactual Theory of Explanation is suitable for all kinds of mathematical explanation. Particularly, in this section we will apply the theory to the cases presented in Chapter 4.

Second, we will examine whether all the cases show a degree of explanatoriness which can be accounted for in terms of the explanatory virtues that Knowles and Saatsi choose for the case of mathematical explanation within the counterfactual theory: cognitive salience, degree of integration and non-sensitivity. Ultimately, simplicity and unification are two of the relevant explanatory virtues that we go for when using mathematics in explanations of both mathematical and empirical events.

2.1.1 The cases in light of the theory

In order to test the theory, let us examine the cases presented in Chapter 4 in the light of the counterfactual theory of explanation to determine if there is a way to account for them from this theory and have a preliminary grasp on whether the theory is a good candidate to account for all mathematical explanation.

2.1.1.1 The completeness theorem

In order to account for the completeness theorem in terms of the CTE, let us apply the structure of explanations as they are characterized within the framework[65].

As we know, the *explanans* includes nomic generalizations $G_1,\ldots, G_m$, statements about initial (or boundary) conditions $IC_1,\ldots, IC_n$, and, further auxiliary assumptions $A_1,\ldots, A_0$.

For Gödel’s completeness theorem, the nomic generalizations include the axiom and assumptions of classical logic and as initial conditions we can find the characterization of the language and the range of variables as well. For instance, this includes the definition of a ‘logical expression’ as a well-formed first-order formula without identity, the definition of “refutable”, “satisfiable” and “valid” expressions. The calculus is specified as being presented in normal form, that is, with the quantifiers occurring at the beginning of the expressions, and the degree of a formula is presented as the number of alternating blocks of quantifiers.

Then, as presented in Chapter 3[66], the proof’s strategy is to show that if the completeness theorem holds for formulas of degree k, it must hold for formulas of degree $k + 1$. Then, it shows that any normal formula $(Q)\varphi$ of degree 1 is either satisfiable or refutable, where “(Q)” stands for a (non-empty) block of universal quantifiers followed by a (possibly empty) block of existential ones. From here, Gödel defines a well-ordering of all tuples of variables that arise from the need to satisfy φ as dictated by (Q).

The *explanandum*, that is, the theorem, states that every valid logical expression is provable. This proof provides an explanation of why, in counterfactual terms, had it been possible to identify a valid argument for which there is no proof in the system, then the logic would have been incomplete.

---

[65] See Chapter 2, section 3.

[66] See Section 2.

In Henkin's completeness theorem, the nomic generalizations include the axioms and assumptions of classical logic (language, calculus…), and the initial conditions include a characterization of the language and the range of variables, which Henkin provides. Among the IC, for instance, there are the conditions that we are working with a first-order language, or the construction and order of the formulae of the language[67]. There are other laws to include in the *explanans*, such as the definition of a consistent set, maximal consistent set and maximal consistent set with witnesses. Then, the proof proceeds by showing how to construct the maximal consistent set with witnesses and, from there, the construction of the model, which ultimately constitutes the core of the proof[68].

As for what the relevant counterfactuals are, the variables that determine whether a proof of completeness in Henkin-style can be put forward have to do with (a) "changing" the cardinality of the language (whether we are working with a finite fragment of the language, an infinite denumerable or a non-denumerable language), (b) expanding the language with second-order quantifiers or modal operators (which would carry the corresponding changes for the calculus or (c) changing the calculus itself (and thus choosing first-order, second-order, modal logic…).

For instance, if we were to include second-order language, the result would be that we would not be able to construct the model.

That the *explanans* logically entails the *explanandum* is shown by the fact that we have a proof of the theorem in both Gödel's and Henkin's version. In Henkin's case, more in particular, the explanatory feature of the *explanans* has to do with the fact that the proof unveils a relationship between certain characteristics of the formal language (and calculus) and the possibility of constructing a model according to Henkin's methodology. The obtaining of such relation allows to establish counterfactual dependencies between the elements of the *explanans* and the *explanandum* (that is, the theorem).

Both completeness proofs establish the theorem and provide some explanation of it but, as we advanced in Chapter 4, Henkin's proof is by far more explanatory than Gödel's. The main difference is that the concepts and formalization that Gödel used in his completeness proof are now obsolete and harder to follow compared to Henkin's, which has to do with our understanding of the proofs, if we take explanatoriness in Brown's weak sense. Moreover, in a *strong* sense, Henkin's proof allows to formulate more counterfactuals in that it is more unified, as we will see in next section.

For now, let us conclude that we can accommodate the case to the structure of explanations provided in the CTE and therefore show that the completeness proof can be accounted for in the context of the counterfactual theory. We shall see in the corresponding section below that the counterfactual theory also provides us with an account of the explanatory power of this proof.

2.1.1.2 Computer proofs: The Four-Colour Theorem

If we take the Four-Colour Theorem and its proof, we cannot ignore the fact that it is a proof by exhaustion inserted into a *reduction ad absurdum*, and this fact will affect our analysis of the proof. Besides, the analysis might bring some light to the study of this kind of proofs.

---

[67] Manzano (1999) explains that the proof can be extended to non-denumerable languages, provided we have the Axiom of Choice, so if we were to extend the proof, we would need to add the Axiom of Choice to our initial conditions. In this case, the axiom would work as a nomic generalization (or law).

[68] See Section 2 from Chapter 4.

Bearing this in mind, the procedure would be that of taking into account all the cases contemplated in the Four-Colour Theorem proof, and thus accounting for proofs by exhaustion in counterfactual terms.

For the Four-Colour Theorem case, we could explain what happens with the (lack of) explanatoriness in counterfactual terms by stating the following: since we cannot revise the proof, that is, go through all the cases, there is no viable way of exploring all counterfactuals. In this case, exploring all counterfactuals amounts to exploring all possibilities, each and every one of the cases being checked separately in order to determine whether the theorem is true or not. In the proof of the theorem, the maps are abstracted to networks, which are the cases checked by computer. The proof's central work is to check each of the relevant cases or conditionals, of the form: "If the map can be represented by network *n*, then it can be coloured with just four colours".

The issue that has been pointed out with this proof by mathematicians and philosophers is that it does not provide a deeper understanding of why the proof works. The fact that we fail to formulate the relevant counterfactuals in an explanatorily significant way has to do with the fact that we cannot go through all the cases of this proof by exhaustion. It is, then, a problem concerning the fact that the proof does not constitute an explanation.

In order to get some clarification of this case, let us, alternatively, consider a toy example. In this case, our toy example will be the case of the triangular numbers explained by the diagram, but isolating the cases of natural numbers from 1 to 5. With this example, we can see that there is a significant difference between just checking all five relevant cases and going for an explanation by using the formula or even the diagram itself.

If we pursue the first option and calculate the sum of the first *n* for every number from 1 to 5 and obtain 1, 3, 6, 10 and 15 as the results, we do obtain the triangular numbers, but this constitutes no explanation at all. We are just calculating sums.

However, if we pursue the second option, the formula ($1 + 2 + 3 + ... + n = n^2/2 + n/2$) provides us with an explanation of the case and, moreover, if we use the picture, we get understanding of how the formula works by getting us to the result for our cases from 1 to 5 (and, by extending it, for any other natural number). That is, the general formula provides a result for each natural number and, having proved the formula, it provides an explanation of why that is the case.

In the context of analysing the contrast that has been shown, Baron's observation that not all not all applications of mathematics are explanatory[69] comes in handy. The mechanism Baron uses to distinguish explanatory from non-explanatory applications of mathematics (for extra-mathematical explanations) is the *informational test*. It is also Baron's objection to the CTE (that it does not allow the distinction between descriptive and genuinely explanatory applications of mathematics), and one of the main objections to the CTE that were analysed in Chapter 2.

Let us see how the informational test would do with proof by exhaustion by applying it to our toy example.

According to the informational test (Baron 2017, 703-704), if we remove the part of the information *I* that is descriptive about a physical system while holding the rest of the informational content of *M* fixed, there are two possibilities. The first one is that the information in *I* exhausts the information that *M* carries regarding the *explanandum*, then *M* fails the test. This is clearly the case for the example of the triangular numbers by exhaustion, since all the informational content of the mathematical proof by exhaustion (that is, checking all the cases)

---

[69] See Chapter 2, Section 4.2 for a presentation of Baron's account.

is descriptive. We are merely describing five cases and, thus, *I* (the description of the cases) collapses with the whole explanation. In this case, according to the test, the *explanation* is only contributing descriptive information.

The second possibility is that the information *I* does not exhaust the information that *M* carries regarding the *explanandum.* In this case, *M* passes the test, so it carries non-descriptive information regarding the *explanandum*. This *extra* information is what ultimately *explains* the phenomenon in question. This is the case for the proof using the formula (which can be combined with the diagram). The formula provides us with a reason why the theorem holds and goes beyond a mere description of the five cases, which contrasts with the proof by exhaustion.

The application of the informational test and Baron's distinction between descriptive and non-descriptive information regarding the *explanandum* is compatible with the version of the CTE that we are using in the present work. In addition, it helps us see how one proof fails to be explanatory while the other one does provide an explanation of the phenomenon. The verdict we can take from the CTE is that when we use mathematics to just measure, these counterfactuals do not present the relevant explanatory features that would make them explanatory. Therefore, in a way, the CTE does provide an answer to this challenge from Baron's informational test.

Given that the CTE can accommodate such a test and Baron's distinction, it is plausible that Baron's objection to the CTE as a good account of mathematical explanation is not a serious one. In the end, what is behind Baron's objection might be mostly his aim at obtaining a picture of mathematical explanation that can be the base of a light-weight platonist approach to mathematical ontology. The CTE, however, is compatible with nominalism as well, a conclusion that Baron wants to avoid.

Besides, this analysis gives us an idea that proofs by exhaustion are not explanatory. This is perhaps even worse for cases where we cannot even check the relevant cases, as happens with the 4CT, which contributes to the idea that the proof is not explanatory since it provides no understanding of why it works, even though it is a valid proof accepted by mathematicians.

As has been advanced, the counterfactual theory should also allow us to get a grasp on why the Four-Colour Theorem proof cannot be considered explanatory. In terms of the theory, there is no way, with this proof, to make the relevant counterfactuals explicit. In other words, the proof does not provide us with the relevant information regarding the different situations that the theorem covers if we change the variables. Besides, the CTE framework establishes that the laws (or the initial conditions, in general) need to establish the relations between the *explanans* and the *explanandum*. In this case, however, we have no laws to connect *explanans* and *explanandum* in a way that shows this counterfactual dependence (recall that the task is performed by a computer). Therefore, we cannot think of the proof as being explanatory.

Moreover, the proof seems to lack the features that make proofs explanatory. Perhaps we could replicate the procedure for other similar theorems (such as the five-colour case, for instance), or other topics that require checking a large number of cases. This would require creating an analogous computer program, which is ultimately the one doing the work of going through all the cases.

To sum up, it is not hard to account for the non-explanatoriness of this proof (and proofs by exhaustion, in general) in the framework of the counterfactual theory.

2.1.1.3 ***Visual thinking: the picture "proof"***

If we proceed to the analysis of the *picture proof* or, more accurately, the *picture explanation*, we can account for the case in terms of the counterfactual theory of explanation.

In addition, as we shall see in the following section, there is a way in which we can argue counterfactually that it is an explanatory picture.

Independently of the debate on whether the picture presented in Chapter 4 constitutes a proof or not, it shows the case from natural numbers from 1 to 5. If we think about it counterfactually, we could ask, for instance: “What if the number to consider were 10?” In this case, we would mentally *extend* the picture and get to the idea that the formula holds for 10 as well (so, it would provide the relevant degree of scope generality and cognitive salience, as we shall see later). There is something related to the shape of the diagram which enables us to grasp the relevant property of these numbers and the generality (*law*) the picture represents. These, with some interpretation, are what constitutes the explanation.

Again, since we are dealing with an explanation of a mathematical statement that holds for all natural numbers and the explanation only shows the truth of the statement for a limited number of cases, our relevant counterfactuals here will be the ones that result from varying the number we take into consideration, and asking whether the diagram can be *extended* in order to account for that number. The counterfactuals for this example will range, as is expected, over all natural numbers.

If this approach is correct, then this means that the counterfactual theory can also account for mathematical explanations of mathematical phenomena that are presented in non-formal ways (the visual element), and that are not proofs in the strict sense. There is a way in which the picture explanation has more explanatory power than the formal one based on induction, and it goes back to the idea that we can mentally extend the diagram and, by *visual means*, grasp the generality.

2.1.1.4 Proofs by mathematical induction

It seems hard to imagine a way in which we can account for mathematical induction in terms of the counterfactual theory of explanation. It might have something to do with the generally accepted thesis that mathematical induction is not explanatory, because it fails to convey the *why* of the *explanandum* or shed some light in explanatory terms. In the end, we put our *trust* in mathematical induction because we know it works, but when it comes to obtaining explanations and understanding, mathematical induction is probably little more than just a matter of *mathematical faith* or a mere case of checking that the theorem works for an infinite set, an idea resembling the procedure of proofs by exhaustion.

Some have argued that there are explanatory proofs by mathematical induction. If this were the case, then we must find a way to account for this explanatoriness in terms of the counterfactual theory of explanation. Perhaps there is a way in which inductive proofs can be explanatory if the step from n to n +1 is illuminating in some relevant way, or if we find a way to establish that the initial conditions (more specifically, the laws) identify relations between *explanans* and *explanandum* and highlight the relevant counterfactual dependence, in which case the proof is something more than just *exhaustion* of the cases, in that they provide a general method for getting from n to n + 1. However, it is doubtful that the task can be done for proofs by mathematical induction, so we have some evidence that these proofs generally provide no explanation of the phenomena they establish.

2.1.2 The cases and their explanatory power

From the analysis of the cases presented in Chapter 4 and the explanatory power of each one of them, we can draw some conclusions regarding what explanatoriness means in mathematics. With this in mind, let us go through our chosen cases again but this time focusing on the explanatory virtues that make those explanations more or less successful and, particularly, those explanatory virtues they have in common. Ultimately, the aim of this section is to show that all cases share the same explanatory virtues, that is, their degree of explanatoriness comes by virtue of the same features.

2.1.2.1 The completeness theorem

For the case of the Completeness Theorem, we can observe an increase of generality when we go from Gödel's original proof to Henkin's. The role of mathematics in this case is, not only a method of achieving a great level of generality (or abstraction) in the obtainment of the proof, but also the proof is an example of how we can, through mathematics, achieve different levels of generality, given that we are presented with two different proofs, but both of a mathematical nature. Besides, the fact that we prefer one of them to the other is quite illuminating, in that it is clear that we prefer the one that accounts for more cases, that is, the one that is applicable to more areas of mathematics and model theory. This suggests that we do prefer explanations that account for more cases. In other words, we choose theories that exhibit more scope and topic generality and, therefore, answer more what-if questions in the sense that they explain completeness for more than just one system.

Moreover, Henkin's proof shows a higher level of cognitive salience if we compare it to Gödel's proof, in the sense that it is accessible to a wider audience and it uses a technical apparatus that is much more accessible. This particular feature of Henkin's proof is key when it comes to choosing one proof over another. It is also behind the fact that nowadays Gödel's proof is not accessible to most students of logic and getting to understand the proof often comes with big efforts to grasp all relevant concepts and background logic Gödel used, both obsolete nowadays. This shows the importance of cognitive salience and context-dependence in proofs, a feature that is characterized as the main advantage of mathematics in explanations by Knowles and Saatsi's version of the CTE.

It is also worth noting that the explanatory power of Henkin's proof is not just a matter of cognitive salience or understanding (Brown's weak sense of explanation). We can say that the proof also explains in a deeper sense, by showing *why* the first-order logic is complete, that is, explaining in the strong sense. By constructing the canonical models, we can generalize the proof to other languages and obtain completeness for different modal logical systems, for instance. This, of course, has to do with the scope and topic generality mentioned above, and the feature of mathematics that provides us with the reasons why a theorem is true. Therefore, with this proof we obtain explanation in both senses presented in Chapter 3.

Following this line of thought, it seems like it would not be too hard to accommodate the case of the completeness theorem into a counterfactual picture of mathematical explanation, and account for its explanatory power.

2.1.2.2 Computer proofs: The Four-Colour Theorem

As has been pointed out, the Four-Colour Theorem Proof can hardly be thought of as an explanatory proof, that is, a proof that, apart from establishing the truth of the theorem, illuminates relevant aspects about *why* the theorem is the case. It has also been pointed out that this is due to the fact that the proof includes a great deal of the work that is performed by a computer and, moreover, cannot be reproduced by merely human means.

In terms of topic and scope generality, we can point out that the proof achieves a suitable level of abstraction in the task of establishing that we can colour any map with just four colours.

However, we should indicate that the same proof sketch does not necessarily apply to the result for five or more colours.

As we have pointed out above, the main issue with the Four-Colour Theorem and its explanatory power is directly related to its failure to provide a real explanation of the theorem, in both senses of the term. First, it does not tell us *why* the theorem holds, for the reasons provided in section 2.1.1.2, related to why proofs by exhaustion are not explanatory in contrast to other kinds of proofs. Second, it does not illuminate relevant aspects of the proof or provide understanding of the fact that we can colour any map with just four colours, so the proof cannot be thought of as cognitively salient. We ultimately believe that the theorem is true, but we obtain no explanation or understanding of any kind.

This is, of course, connected to our inability to revise and check all the relevant cases[70] of the proof, which gives us a grasp of why the proof fails to be a good explanation of the theorem.

It is, however, worth noting that, even though the proof is not verifiable by humans (in an algorithmic sense), Appel argues that it is "easily replicable", in the sense that the program can be reproduced in any computer for the work of checking the cases: "Most mathematicians who have some computer background would be satisfied with a copy of the program, some way of verifying that the inputs were typed in correctly, and some output that indicated that the program ran to its end. *Any reader who remained worried could easily program a matrix multiplication algorithm for his or her own computer, type in the inputs and run a check program.* What qualms remain would very likely result from the possibility that both author and checker had misread some of the input entries, but certainly this is less worrisome than the difficulties involved in performing similar tasks by hand. I would call such a proof *an easily replicable proof*" (Appel, 1984, 36). Of course, the issue of the use of computer programming for performing the proof still stands.

Following this line of thought, let us consider that, even though it is true that we cannot reproduce the proof by only *human resources* (if we do not consider a computer program a mere extension of our resources) and, therefore, a big part of the proof is obscure to us, there is another way in which we can consider the proof explanatory, and that is by understanding the reasoning behind the construction of the structure of the proof, the process previous to the use of the computer and the interpretation of the computer results. If we consider these, then it looks like there is a glimpse of hope for the explanatoriness of the proof, in that we get to understand why we construct such a proof (containing such a large number of combinations that it is materially impossible for us to check each and every one of the cases). However, the fact that we know what we need to check does not by itself make the proof explanatory, since we obtain no reason or understanding regarding the result. Rather, this would still leave us far from grasping all the relevant aspects behind the proof and the result. If we take the proof of the Four-Colour Theorem to be an explanation of the theorem, we must either consider it a *bad* explanation in that it leaves us with no understanding of *why* the theorem is true, or reject it altogether as an explanation in the genuine sense. If we do the latter, then the proof will still be a proof for it does establish the truth of the theorem, but we would refuse to consider it a proof that provides an explanation of the given result. In other terms, the proof is not an explanation of the theorem even though it succeeds in establishing its truth.

With regard to explanations that gather a large number of cases, my intuition is that the level of cognitive salience decreases when the number of cases gets higher to the point where we lose the ability to cognitively grasp the whole set of them and see any pattern in it. If these cases are part of a proof by exhaustion, it might not be an issue of a decrease of cognitive

---

[70] See Section 3.3.1.

salience after all since, as we have seen, examining all cases by itself does not provide any cognitive salience (or explanation) at all. It is true that the Four-Colour Theorem presents *too many* cases to check (1476, in fact) and, as we know, as humans, cannot possibly check them by ourselves, so we *delegate* the checking of the proof to a computer, instead of using more traditional methods such as asking other mathematicians to check the proof, for instance. On one hand, we remain not knowing *why* the theorem is true in the *strong* sense and, on the other hand, the proof does not clarify any aspects of the fact that we can colour any map with just four colours in the sense related to understanding. Perhaps, we would even need of artificial intelligence (that is, more computer elements) in order to get such an explanatory pattern for us to obtain an explanation of the theorem. This explanation could be an adaptation of the proof or, alternatively, an explanation of why we can colour any map with just four colours that is not a proof. But, once again, we are relying in artificial intelligence instead of our *human colleagues* for this task.

At most, we can believe in the process that got to the proof (which includes the sketch of the proof and the parts that were designed by humans) and in the truth of the theorem in question. Still, the fact that the proof establishes the truth of the theorem but it does not illuminate any aspects of how that is the case looks more like an issue concerning non-explanatory proofs such as the Four-Colour Theorem proof, and not a question about proof by exhaustion or even the Counterfactual Theory of Explanation itself. Indeed, as we have seen, the counterfactual theory allowed us to get a grasp on why the proof cannot be thought of as an explanatory one.

Perhaps a relevant question concerns why we have not (at least yet) come up with an alternative proof that does satisfy the criteria for being explanatory. The answer might have to do with the proof being too complex in that it involves too many cases to process, so we have not been able to get a global picture of what goes on in the proof in a way that yields understanding.

This situation even led to the rejection of the theorem by some, for believing that proofs that involve computers are "pseudo-mathematics", such as Bonsall (1982), who argued that "if the solution involves computer verification of special cases […] such a solution does not belong to mathematical science at all. […] We cannot possibly achieve what I regard as the essential element of proof –our own personal understanding – if part of the argument is hidden away in a box"[71]. Tymoczko, as we have already seen, argues that the viability of a proof depends on it being surveyable, convincing and formalizable, three properties that the Four-Colour Theorem proof does not meet, due to the use of the computational work, which would be against rigorous mathematical arguments. According to Tymoczko (see 1979, 58), the inclusion of computers in mathematics has some serious philosophical implications, such as the fallibility of the theorems and results, and mathematics no longer being purely *a priori* knowledge.

In contrast, I agree with the opposite view, according to which mathematical methods are diverse and some of them include *experimental* techniques, apart from of the use of computers[72]. The use of computers could be just an extension of our cognitive capacities. As we know, our methods need not be infallible, but just reliable, and this is the case of computers, which are at least as reliable as mathematicians. What the use of computers may be introducing in this case, though, is some obscurity into the proof in that we cannot replicate the computer's work (and, thus, loss of cognitive salience). However, proofs can still be "obscure" independently of the use of computers or any kind of "experimental" resource, and perhaps computer errors are not so different from human errors. The program used in the proof was

---

[71] F. F. Bonsall, "A down-to-earth view of mathematics", American Mathematical Monthly 89, 1982, 8-15.

[72] See Chapters 3 and 4.

written in a highly formal language and that can indeed be checked by humans. The computer, in the end, is limited to performing the operations that are given to it.

The question of whether computer aided proving challenges the alleged infallibility of mathematical knowledge is probably an overreaction to the case, since computers are unlikely to lead to more errors than "ordinary" mathematical practice and, besides, it is probably more a question of elegance or cognitive salience. The use of computers would probably just represent an aid to our cognitive capacities, at least in cases like the Four-Colour Theorem.

The way I look at it, the main problem raised by the Four-Colour Theorem and computer proving is that, while a computer can have a role in establishing the truth of a theorem, using computers in proofs can sometimes leave us with no more understanding of why the theorem is true than we had before the proof, as will happen with any proof by exhaustion. In this particular case, the work performed by the computer, is *hiding* in some way the explanatory features of the proof, which include the checking of all combinations. At least, this is the case to this day. We may get a real explanation of the Four-Colour theorem someday, perhaps even with the help of artificial intelligence.

Moreover, this issue is also related to the question of whether we maintain an idealized concept of proof and try to make (or even *force*) every particular proof in mathematical practice conform to it, or we should rather try to reflect mathematical practice itself and see what is actually going on in it. This dichotomy is nothing strange in philosophy of science and mathematics is no exception. There is also some sort of intermediate route, which is the idea that it works in both directions. However, the debate on whether the computer proof is a genuine proof is independent from the debate on its explanatoriness or its qualification as an explanation. The point in the present work is that, whether it is or not a proof in the strict sense, it does not succeed as an explanation of the Four-Colour Theorem in either of the two different meanings of explanation with which we are working here.

2.1.2.3 Visual thinking: The picture "proof"

The "explanation" of the theorem for triangular numbers can be easily given by using the diagram from Chapter 3. By making use of this tool, the theorem is grasped by quite immediate means. Besides, the diagram makes the mathematical theorem accessible to a wide audience as long as it has some very basic concepts of basic arithmetic.

For these reasons, it may be argued that the explanation consisting of the picture - accompanied by some orientations – has a high explanatory power. Moreover, the counterfactual theory can help us see how the explanatory virtues considered in the theory (fundamentally generality, simplicity and cognitive salience) are present in the explanation.

By mentally *extending* the diagram to account for other cases than natural numbers from 1 to 5, we can almost (literally) *see* that the formula holds for all the cases, so we find that the explanation provides the relevant degree of scope generality and it does so in a very simple way, by visually showing it with a rather simple diagram. Therefore, generality and simplicity are present in the explanation, which, once again, constitute two of the fundamental explanatory virtues for which mathematical explanations are valued.

The fact that we can grasp the regularity that the formula conveys with ease and that the diagram illuminates relevant aspects that help our understanding of the mathematical statement sheds some light in terms of cognitive salience on the explanatoriness of the explanation. There is no doubt that the picture, with some interpretation, helps our understanding of how the theorem reflects a regularity in the natural numbers, and putting it in this fashion shows its scope (all natural numbers) and provides us with a high level of cognitive salience, in Knowles and Saatsi's terms.

This has to do with the shape of the diagram, which is very effective at enabling us to grasp the relevant property of the numbers, which constitutes the theorem. There are, of course, many other examples of mathematical explanations that are more or less visual, namely in geometry, for obvious reasons. For instance, the Pythagoras Theorem can be easily explained with visual resources:

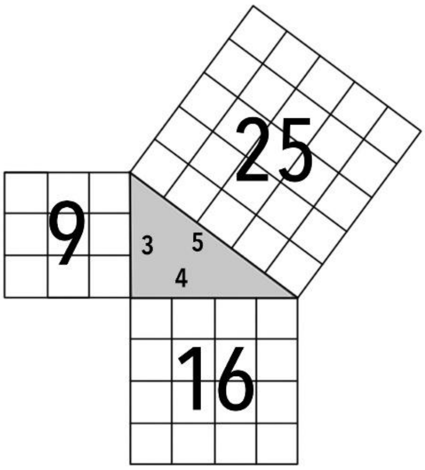


Diagram of the Pythagoras Theorem

This diagram, very much used when teaching mathematics, is an effective way of conveying the theorem. With little interpretation, it makes it easy to understand this particular case, $3^2 + 4^2 = 5^2$, but also any other right angled triangle, since we can mentally adapt the picture to other cases. The picture of the triangles with the squares constitutes a very useful explanatory tool in this diagram, and there are countless more examples like this one, since this kind of resource is very common in geometry and other mathematical areas.

Back to our initial case, it has been argued that the picture by itself does not constitute a proof, and even with some orientations in form of sentences, it would probably still fail to constitute a proof. The *real* proof, in the *strict* sense, would be a classical deduction from the axioms. Nevertheless, there is no doubt that the diagram is illuminating and it helps us grasp *why* the theorem holds for all natural numbers in a very simple and elegant fashion - undoubtedly much simpler than a classical deductive proof- and also providing understanding of the theorem.

One of the concerns that can be found in the literature regarding visual thinking and diagrammatic explanations is that if they play a relevant role in the process of grasping mathematical concepts and acquiring mathematical knowledge, then mathematics may need to be considered an *a posteriori* science, since sensory experience would play a role that goes beyond just aiding understanding and "enabling" and more the "evidential" kind of role in the process of acquiring knowledge. This, of course, is notable since it contradicts the traditional idea that mathematics is an *a priori* science. However, we can still argue that we are acquiring *a priori* knowledge since the result does not depend on a particular (in this case) diagram or picture. The picture is just an aid to grasp a mathematical concept. The same applies to the Pythagoras Theorem and so on. For instance, counting on our fingers belongs to this kind of evidential use of sensory experience in mathematics. Counting can be an *a posteriori* way of getting mathematical knowledge (which can later be generalized and may lead to more abstract thinking). Another example concerns the question of how many vertices a cube has[73]. Since we have the background knowledge that cubes do not vary in shape and material cubes do not differ from geometrical cubes in number of vertices, we can know the answer by looking at a material cube, or even by using our imagination and mentally visualizing a cube, and we will extract the information that the vertices of a cube are exactly the vertices of these two quadrangular faces.

The key aspect of these examples is that the visual experience contributes to our belief in the answer. Therefore, visual resources are a relevant part of the process of our acquiring mathematical knowledge in many cases (particularly salient in geometry). Still, this is nothing strange in mathematics. Whenever we abstract, we abstract *from something*, so that radical rejection of all kinds of experimental thinking in mathematics may be unjustified.

As Hanna (2010, 16) puts it when she addresses the question of to what extent visual representations can be used not only as evidence for a mathematical statement, but also in justification, she acknowledges that diagrams and visual aids have been used as heuristic tools to facilitate understanding, but more recently they have even been considered substitutes for traditional proof, which led to much controversy. In fact, it is quite plausible that the diagram can constitute a proof provided that it is accompanied by the relevant indications of how it should be interpreted. This route, however, would take us to a debate that goes beyond the purposes of this work. Let us, for our current purposes, just state that we can take the picture to be an explanation of the formula, whether it can constitute a proof or not.

To sum up, it is undeniable that the picture helps us grasp the relevant concepts and, perhaps more importantly, the truth of the theorem with limited resources, which, in terms of explanatoriness, is of great value. The diagram allows a generalization from a limited number of cases, from which it is quite easy to grasp that the formula applies to all natural numbers. As an explanatory tool, the picture is highly successful at providing quick and easy understanding, and also aiding our grasping of the reason why the theorem holds. From this, it makes sense to conclude that diagrammatic reasoning has a relevant role in mathematical explanation.

2.1.2.4 Proofs by mathematical induction

Mathematical induction can be seen as a way of shortening a proof by exhaustion for natural numbers, which would need to consider infinite cases. As we know, there is some disagreement on the question of whether proofs by mathematical induction are explanatory or not. Some authors find that inductive proofs can be insightful and provide real understanding (see Baker 2009, Cariani 2011, Baldwin 2016). In contrast, many others (Lange (2010a) and Hanna (2010), for instance) have argued that proofs by mathematical induction fail to be explanatory.

Indeed, there is little doubt that – at least – some proofs by mathematical induction are non-explanatory, at least, in an epistemic sense. The case of the formula to obtain triangular numbers we considered in Chapter 4 is an example. Apart from the picture *proof* or picture explanation, there is a proof by induction that establishes the truth of the theorem but leaves us wondering how such a formula is found (that is, we get no illumination or understanding of the formula, even though the fact that we have a formula provides a *why* in the *strong* sense of explanation). In contrast, as has been pointed out, the picture is far more explanatory, even if it does not constitute a proof by itself. This is so in both senses of explanation: the picture is of cognitive help, and it explains in a *strong* sense (in that it depicts a geometric reason). Similarly, the formulas for the sum of the first n positive integers ($n(n+1)/2n(n+1)/2$) or the sum of the squares

---

[73] For more on this example, see: Starikova, I. and Giaquinto, M. (2017), “Thought Experiments in Mathematics”, Stuart, Michael T., Yiftach Fehige, and James R. Brown (eds.), *The Routledge Companion to Thought Experiments*, London and New York: Routledge.

of the first *n* positive integers (n(n+1)(2n+1)/6n(n+1)(2n+1)/6) can be both proved by mathematical induction. However, these proofs are not explanatory. Taking Brown's distinction, they do not provide a satisfactory explanation in any of the senses of the term he distinguishes. On one hand, they do not account for *why* those formulas work or how they were found and, on the other hand, they do not provide us with understanding in the sense of the proof being *user-friendly.*

Indeed, even from an educational perspective, mathematical induction can come as an obscure mechanism for students who first approach it, since, no matter how easy the principle statement is to learn, it may look a bit opaque in the sense that from the statement of the induction principle we do not grasp how or why it proves the case for all natural numbers. More importantly, it gives us no clue as to why the particular mathematical statements we prove by using the principle are true in the first place. When we prove it by mathematical induction, it looks like we are checking a result that was found, but in that checking process we find no explanation for the result. Like many other proofs by mathematical induction, these proofs do not explore any ideas or explanations, they are done by using merely algebra resources. From this point of view, it is little more than a *trick* that allows us to check an infinite set by exhaustion. This is plausibly also behind the fact that sometimes, even though we have a proof by induction for a particular result, mathematicians do not quit the job of looking for different proofs of the result, for considering the original proof not explanatory enough. The reason that these proofs lack explanatory power lies in the fact that they do not illuminate why the relevant property is inherited, that is, why if the property holds for *n* it must hold for *n* + 1.

As we know, Lange (2010a) has a detailed discussion about this subject, where he argues that arguments by mathematical induction are not explanatory. This discussion has been addressed in Chapter 4, so let us limit ourselves to just picking some key features of his analysis.

Lange follows the strategy (see Lange 2010a, 207-210) of establishing a contrast between mathematical induction and a different *version* of induction that works upward and downward from a fixed number (different from 1), which is equivalent to ordinary induction.

The explanatory power of both inductions should be the same, since the only thing that distinguishes both procedures is the starting point. From here, Lange argues that following this path will fall into an unacceptable circularity and, given that any theorem provable by mathematical induction can also be proved by the upwards and downwards rule, by applying reduction, proofs by mathematical induction cannot be considered explanatory (see Lange 2010a, 209).

As we have seen in Chapter 4, there has been some criticism directed to Lange's proposal and its conclusions, such as that of Baker (2009) and Cariani (2011), who point out that there is some asymmetry between traditional induction and Lange's upwards and downwards version of induction, or Baldwin's (2016) thesis that inductive arguments can be explanatory and his criticism of the idea that an instance of a universal generalization cannot help explain the universal generalization.

The question of whether mathematical induction yields explanatory proofs is still an open question, to which complete extensive works could be dedicated. In fact, it is one of the most controversial debates concerning the explanatoriness of mathematical proofs, since philosophers and mathematicians tend to rely on their own intuitions to *decide* the question. The result is a debate with very confident viewpoints from philosophers and mathematicians but with no many strong arguments for and against the ability of mathematical induction to provide us with explanation and understanding.

What is clear is that some inductive proofs seem to be explanatorily sterile, such as the ones mentioned above. As it has been argued, a proof by induction provides a method to check that the (infinite) set of natural numbers presents a particular property. Beyond that, it could be that some of these proofs by mathematical induction provide a reason why all natural numbers have the property, so perhaps some mathematical inductions are explanatory after all. What is clear is that the existence of examples of non-explanatory inductive proofs does not suffice to argue for the non-explanatoriness of all inductive proofs or, more importantly, it does not support the thesis that the lack of explanatory power that those proofs present is due to their being proofs by mathematical induction. However, the ideas provided above related to the comparison with mathematical exhaustion can help to clarify why mathematical induction tends to be non-explanatory.

2.2 Does the counterfactual theory of explanation do the job?

After reviewing our case studies both in their explanatory power and whether they can be accounted for in counterfactual terms, it is time to evaluate the counterfactual theory itself and draw some conclusions as to whether it is a suitable account of internal mathematical explanation in general.

As we have illustrated with our case studies, the counterfactual theory seems to be quite successful at accounting for the explanatory power of internal mathematical explanations, in both senses. The analysis of the *weak* (epistemic) sense of explanation is already in Knowles and Saatsi's (2019) account of the CTE (though applied only to extra-mathematical explanations), and corresponds to cognitive salience. As for the *strong* (objective) sense of explanation, we can find in the fact that proofs provide us with *why* a theorem holds, it is compatible with a counterfactual reading of explanations, so it can be accounted for in terms of the CTE as well. Still, this is often accompanied by other explanatory virtues (such as generality and understanding) and it fits within the counterfactual account.

If this approach is correct, it constitutes an extension of the framework as Knowles and Saatsi had designed it, in that the same theory applies to both external and internal mathematical explanation.

This analysis also highlights that explaining is a human activity, the goal of which – or, at least, one of the main goals – is the provision of explanatory understanding. This introduces epistemic and pragmatic aspects to explanation. The fact that, in the account, cognitive salience is one of the central roles of mathematics goes along with this idea. Mathematics has the virtue of presenting the relevant information in a more accessible and understandable way, and counterfactuals help in that they help us measure explanatoriness in terms of how much modal explanatory information an explanation provides to a particular audience.

As for scope generality and topic generality in the case of intra-mathematical explanation, they become perhaps even more relevant, since the increase in scope and, more importantly, topic generality might say something about how some mathematical structures and entities relate to each other and respond to similar patterns. This reveals aspects of mathematics itself that are of interest in mathematical practice.

2.2.1 Can the CTE be improved?

So far, we have explored the possibility of using one unified theory of mathematical explanation to account for all cases where we use mathematics to explain both mathematical and non-mathematical phenomena. We addressed the counterfactual theory of explanation as one of the most promising theories for that purpose from those found in the literature. Now, it might be reasonable to introduce some minor changes to the theory as it is in order to make it more successful when it comes to analysing mathematical explanation.

As we will see below, many authors have highlighted the importance of taking into account aspects related to context dependence and even the audience dependence of mathematical explanations. Even though Knowles and Saatsi, in a way, share this viewpoint by attributing such an important role to cognitive salience as a concept which already incorporates this idea that understanding is crucial and it can vary depending on the context or the subject, we could perhaps enhance the view by incorporating some more specific concepts concerning this context dependence. It might be worth exploring how the counterfactual theory of explanation can benefit from some features of other theories. This idea of the identification of the relevant properties, presented by Frans and Weber and going back to Steiner, can be one of them[74].

Another possibility here is to go to a quite recent account of mathematical explanation put forward by Bueno and Colyvan (2011), the inferential account of mathematical explanation. This proposal has the advantage of being ontologically *neutral* (as the CTE is), in that it does not commit us to any particular ontological approach to mathematical entities, that is, it is intended to be compatible with both realists and anti-realists approaches to mathematical ontology. The difference will lie in the way in which each view interprets the success in obtaining inferences (see Bueno and Colyvan 2011, 366-367).

What makes this view interesting here in order to complete or enhance the counterfactual theory is the role attributed to choice and context dependence, where “at the interpretation stage we need to make a decision about how to interpret the sum obtained, because, in the uninterpreted mathematics, the sum is just a real number” (Bueno and Colyvan 2011, 355). In cases involving idealizations, there is no full mapping between the empirical set up and the mathematical structures so, in cases of this sort, there are partial mappings between the empirical and mathematical structures. The notions of partial structure and partial relation can potentially help the counterfactual theory account for all kinds of mathematical explanations and help its success in that aim, particularly in some extra-mathematical explanations, when the phenomena are not completely in correspondence with the relevant mathematical model and thus formulating all the relevant counterfactuals can be a difficult (or, perhaps, impossible) task.

2.2.2 Counterpossibles

One of the potential problems some have detected for the counterfactual theory, and that could bring some doubt about the possibility of applying the counterfactual theory to internal mathematical explanations, is counterpossibles and counterpossible reasoning.

We often find it quite natural to formulate counterfactuals in our ordinary and scientific reasoning. For instance, we might wonder what would happen had Bolsonaro lost the Brazilian election. Mathematical thinking is no exception to this, and we often formulate mathematical counterfactuals. The peculiarity with mathematical counterfactuals is that a good part of them are counterpossibles, which introduces the issue of counterpossible reasoning. Counterpossible reasoning consists of reasoning from assumptions or conditional antecedents that are impossible, that is, necessarily false. For instance, this kind of reasoning could take the form of asking what would be the case if 2 were an odd number, when we wonder what would be the

---

[74] Frans, J. and E. Weber, 2014, “Mechanistic Explanation and Explanatory Proof in Mathematics”, *Philosophia Mathematica,* 22: 231–248.

case had 2 been an odd number, we are placing ourselves into an impossible situation, since mathematical truths are necessarily true and mathematical falsities are necessarily false, such as this one where number 2 is odd.

In these cases, we intuitively tend to accept some conclusions, but not all of them, or the reasoning would be completely trivial as anything follows from a contradiction. This would probably lead us to operate in non-classical logics, depending on our assumptions. As we know, in classical logic a conditional with a false antecedent is true. So, all counterpossibles are, according to classical logic, true. This is problematic since it does not allow a distinction between the following two kinds of conditional statements:

C1. Were it the case that 13/4 = 3, then cicadas would probably not have a 13-year life cycle.

C2. Were it the case that 13/4 = 3, then cicadas would still have a 13-year life cycle.

The issue with these examples is that C1 is clearly true, and even explanatorily relevant for the cicada case, whereas we would probably want C2 to be false, even though both C1 and C2 share the same impossible (thus, false) antecedent. The difference between both counterpossibles is that C2 seems to be explanatorily impotent, or even simply wrong reasoning.

Also, there could be counterfactuals like:

C3. Were it the case that 13/4 = 3, then Australian people speak English.

Intuitively, we will want to consider C3 false, but for a different reason than the one we gave for C2. The reason for that is related to the subject matter, since the antecedent and the consequent have nothing to do with each other. This counterfactual would easily be regarded as an absurd.

Therefore, it seems that, even though according to classical logic all counterpossibles are trivially true, this contradicts the intuitions guiding us to look a bit more into the content of such counterfactuals[75]. The traditional semantics for counterfactuals is due to Stalnaker (1968) and Lewis (1973), who use the framework of compositional intensional semantics, possible world semantics, in particular. Logics of a Lewis-Stalnaker fashion are not completely successful in accounting for counterpossibles, since they still entail that counterfactuals with impossible antecedents are vacuously true.

However, we sometimes need to hypothesize and reason about theories knowing that they are not only false, but cannot possibly be true or correct, so it looks like counterpossible reasoning might still be relevant in mathematical conjectures and even metaphysical debates. Sometimes, in metaphysical research, we postulate clearly impossible worlds, and we might operate with counterpossibles in order to achieve the task.

Relatedly, Timothy Williamson (2007, 2017)**[76]** has addressed the subject of counterpossibles from a particular perspective. He defends orthodoxy (the thesis that all counterpossibles are true) against some objections that have been put forward. Very briefly, he (2007) argues that all counterpossibles are true (so, against the non-trivial approach to counterpossibles) on the grounds that treating counterfactuals as them not being trivially true leads to the failure of several logical principles which hold in the standard approach to counterfactuals (the Lewis-Stalknaker approach). For instance, it would lead to impossible worlds and create opaque contexts (that is, contexts in which the substitutivity of co-referential terms does not hold).

---

[75] The fact that conditionals with false antecedents are true is probably the most surprising and counterintuitive principle that students of logic encounter when they first approach the discipline.

[76] For criticism of Williamson's approach, see Berto, F., French, R., Priest, G. *et al.*, 2018, "Williamson on Counterpossibles". *J Philos Logic* 47**,** 693–713.

Back to the subject of explanation, if we intend to go to intra-mathematical explanations and see how the counterfactual theory and, in particular, counterpossibles do, we will find that perhaps it is not that easy to connect counterpossibles with an explanatory picture of counterfactuals. Whenever we use a counterpossible, we immediately lose track of the *truth* of the statements we are using. This is due to our intuitive idea that the antecedents of those conditionals (and, sometimes, the conditionals themselves) do not even make sense. For instance "If it were the case 3 + 6 = 7, then either 3 or 6 would have to be an odd number" is plausibly true, but with an impossible antecedent there is something strange about the reasoning itself.

The fact that it seems counterintuitive to treat mathematical objects counterfactually (that, is, counterpossibly) may be due to the structure we think about when we reason in mathematics. For instance, the possibility that number 2 is odd might seem a strange thought for it implies that then number 2 would not occupy the place that it actually has in the number line or the number structure. That is, the scenario where number 2 has a different property than it actually has would contradict the whole mathematical structure thus making this thought, at least, very hard to imagine. From all the counterpossibles we can use in our counterfactual reasoning, perhaps the mathematical ones are of the most difficult to picture.

Besides, not every counterfactual in explanations of mathematical phenomena can result in an adequate tool in order to understand an explanation better. If considering a mathematical structure where 13 is not prime does not provide an adequate frame to analyse the cicada case, in the same vein we will not always obtain more unification or generality with the use of mathematical counterfactuals that vary the mathematical structure.

As much as the debate on counterfactuals has interest on its own, considering counterpossibles from a general perspective, I do not see how these concerns regarding counterpossibles really affect the application of the counterfactual theory to internal mathematical explanation. In Chapter 2[77], we addressed the question of the use of counterfactuals in Baron's and Knowles and Saatsi's account. We analysed that, in extra-mathematical explanations, it seemed odd to formulate counterpossibles as the relevant counterfactuals, since it is more reasonable to direct our counterfactuals to what we want to explain (the empirical phenomena) and leave mathematics *as it is*. Of course, in intra-mathematical explanations if there is anything to *vary*, it has to be mathematics itself.

However, in the end, when we formulate counterfactuals within mathematics, we are not really wondering about what would happen to the mathematical structure had, for instance, 4 been prime. What we are after, when we formulate such counterfactuals (more accurately, counterpossibles), is a better understanding of the mathematical structure as it actually is. Therefore, counterpossibles in the context of the reasoning from the CTE do not constitute much more than mental work or some sort of a "thought experiment", which does not commit us to anything even though the antecedent is necessarily false. After all, thinking about impossible situations is an activity in which we engage on a regular basis in order to highlight the relevant conditions in a particular situation in need to be explained. In mathematics it is quite frequent to reason from counterpossible antecedents. We function in that way, for instance, in proof by contradiction. This contributes to the idea that, at least in some situations, counterpossible reasoning makes sense within mathematics.

For those who would argue that counterpossibles are a problem for the counterfactual theory in the case of internal mathematical explanation, I would suggest considering non-classical logical systems, where avoiding counterpossibles altogether can be done with relative

---

[77] See Section 6.1.

ease. However, it might make sense to accept that counterfactuals are part of counterfactual reasoning, given we quite often reason over impossible situations, just to get more clarification in our original subject.

To sum up, for addressing the question of whether we can use counterfactuals in a meaningful way in internal mathematical explanation, we are faced with two options. The first one is to take into consideration the possibility of formulating mathematical counterfactuals which do not have a necessarily false antecedent, that is, avoiding counterpossibles if that is viable. The second alternative is choosing to accept counterpossibles in the cases where they show to be explanatorily relevant or helpful.

I prefer the second option. Perhaps mathematical impossibilities are not that different from physical impossibilities after all. We can understand counterfactuals simply as some kind of thought experiment that is, necessarily, partial. What I mean by this is that, when we wonder how would a non-prime 13 affect the duration of cicadas' life-cycle, we are not considering all the consequences of 13 being prime (which would ultimately *destroy* the number system as we know it), but only those that affect the case on which we are focused. In the same way, when we formulate counterfactuals (or counterpossibles) whose antecedent affects the empirical world, we only consider those consequences that produce a difference in the *explanandum*. Some counterfactuals posit situations that would make the empirical reality impossible. However, this goes beyond the scope of the present work.

I think it is safe to assume that we only formulate counterfactuals (and counterpossibles) with particular targets when we are trying to get a successful explanation, so automatically assuming that counterpossibles are problematic just because of what they imply is a wrong approach to the subject if we are to take the counterfactual theory seriously. Moreover, perhaps we can conceive of counterfactuals as providing *epistemic* counterfactual dependence, rather than some kind of *ontic* counterfactual dependence. This relates the counterfactual theory of explanation directly to our understanding (or explanation in the *weak* epistemic sense).

3. Lange’s approach: symmetry and simplicity

Mark Lange has a vast work on mathematical explanation, where many relevant aspects of mathematics’ explanatory role are highlighted. Part of this work is collected in a book titled *Because without Cause* (2017). Lange’s view was addressed when the particular issue of mathematical induction was discussed. Our purpose here is far from providing a detailed discussion of Lange’s work, but rather choosing the relevant features that will help us complete the CTE or simply understand internal mathematical explanation better. Let us take Lange’s work on internal mathematical explanations, in particular, his emphasis on the role of symmetry and simplicity, as well as his view on how mathematical proofs can explain.

At the end of this section, we will be able to apply this framework to the cases we are considering in this work and test whether Lange’s theory does a good job at accounting for them, compared with the CTE.

The central idea in Lange’s work is that the difference between proofs that explain and proofs that do not is that the explanatory proof exploits some feature in the setup that is salient in the result: “what it means to ask for a proof that explains is to ask for a proof that exploits a certain kind of feature in the setup— the same kind of feature that is outstanding (i.e., salient) in the result” (see Lange 2017, 255). It is the salience of that feature what makes certain proofs explanatory.

Very often, this property will be symmetry, but there are other cases in which it is simplicity. Still, Lange acknowledges that there can be others, or even some proofs do not present any notable features (see Lange 2017, 264).

One of the main examples (of symmetry as the salient feature) Lange provides is that of the biased coin (see Lange 2017, 234-238). Lange presents this example in the following way[78]:

> A number *p* between 0 and 1 is generated randomly so that there is an equal chance of the generated number’s falling within any two intervals of the same size inside [0,1]. Next a biased coin is built so that *p* is its chance of landing heads. The coin is then flipped 2000 times. What is the chance of getting exactly 1000 heads?

As Lange indicates, our thoughts about this problem might be that if *p* is close to 1, then a sequence with a vast majority of heads is likely (and if it is close to 0, a sequence with a vast majority of tails is likely). This being the case, the chance of getting 1000 heads on 2000 tosses is very small. However, *p* is random, so if it is in the middle between 0 and 1, then the sequence becomes more likely. This is the likely reasoning of the problem. Nevertheless, the result is that there are 2001 possible results for 2000 tosses, from 0 heads to 2000 heads, so the fact that 1000 heads is in the middle does not count in favour or against its chances. Each of the possible outcomes has the same probability, which may strike us as surprising[79].

The relevance of the example is the fact that every possible outcome has the same chance, and that it can be explained in terms of symmetry (see Lange 2017, 237-238). There is symmetry at the setup of the proof (“when the $n + 1$ generated numbers (from the same random-number generator) are listed from smallest to largest, each possible position on the list is equally likely to be occupied by the first number that is generated” (Lange 2017, 238)). According to

---

[78] He takes the presentation from the Bay Area Math Meet, San Francisco, April 29, 2000.

[79] See Lange’s presentation of the example in Lange 2017, 234-235.

his view, the symmetry in the setup accounts for the same symmetry in the outcome, to the point that our curiosity was caused by that symmetry as well. A different proof that does not exploit that symmetry in the setup ends up being non-explanatory (see Lange 2017, 238).

His conclusion from the example is that it is quite frequent that “a mathematical result that exhibits some striking symmetry is explained by a proof showing how it follows from a similar symmetry in the problem. Each of these symmetries consists of some sort of invariance under a given transformation; the same transformation is involved in both symmetries” (Lange 2017, 239).

Lange provides many other examples from different areas, such as probability, real analysis, number theory, complex numbers, etc. He includes, for instance, D’Alembert’s theorem (see Lange 2017, 239-242) or several geometric explanations that exploit symmetry (*ibidem*, 245-254).

Mathematical coincidence also has a relevant place in Lange’s analysis. One of the examples he employs is that, from the keyboard of a calculator (minus the zero), we can form a six-digit number by taking the three digits on any row, column and diagonal in forward and in reverse order (for instance, 123321 or 357753), and each one of those numbers (a total of 16) is divisible by 37 (see Lange 2017, 276).

7	8	9
4	5	6
1	2	3

Of course, one way of proving that this is the case is by checking each one of the cases (a proof by exhaustion) but, according to Lange, this treats the result as a coincidence and does not constitute an explanation (see Lange 2017, 276). This idea is not too different from our thesis concerning proofs by mathematical exhaustion.

According to Lange, a proof should proceed from some property that is common to all 16 numbers. Indeed, the proof exploits the fact that these numbers “can be expressed as $10^5 a + 10^4 (a+d) + 10^3 (a+2d) + 10^2 (a+2d) + 10(a+d) + a$ where a, $a + d$, $a + 2d$ are three integers in arithmetic progression” (Lange 2017, 277). These three integers are the ones that generate the calculator number from the calculator keypads (taken forward and backward). The fact that the numbers can be expressed in this form is the property that the proof tackles and it is what actually explains the mathematical fact.

Lange treats this issue in terms of “unity”, that is, the relevant property unifies all the cases by treating them in the same way and, thus, making them non-coincidental. This acknowledgement of the importance of unification in his account resembles Kitcher’s approach. However, he takes some distance from Kitcher’s view according to which, roughly, all kinds of explanations involve unification in one or another way, and sees his account of mathematical explanation as being a better one by doing justice to “the fact that (as I have shown in various examples) we can appreciate a proof’s explanatory power (or impotence) just from examining the details of that proof itself, without considering what else could be proved by instantiating the same scheme (or “proof idea”) or how much coverage the given proof adds to what’s covered by proofs instantiating other schemes” (Lange 2017, 308). That is, Lange regards unification as one sort of explanatory power (among others), which operates where the salient feature of the explained phenomenon involves unification and, besides, not every case of mathematical unification yields explanatory power.

Lastly, Lange’s analysis can be joined to those accommodating the idea that a proof’s explanatory power can depend on the context: “On my account (the “big Lange theory”), whether a proof qualifies as an explanation of the theorem being proved depends on what feature of the theorem is salient. In different contexts, different features may be salient. Therefore, a proof that qualifies as an explanation in one context may fail to do so in another" (Lange 2017, 298-9). Nevertheless, this does not mean that there is no objectivity in mathematics, since the context-sensitivity of a proof's explanatoriness does not imply that a proof does not reveal how things really are (see Lange 2017, 303).

3.1. Lange's approach and the CTE

There is a way in which Lange's approach is not entirely incompatible with the counterfactual theory. It may even contribute to highlighting relevant features that the CTE should take into account.

Lange's main idea that the salience of that relevant feature that is exploited in the setup of a proof is what makes the proof explanatory is not too far from the idea that a good explanation establishes patterns of counterfactual dependence between the *explanans* and the *explanandum*, by identifying the relevant variables. Those relevant counterfactuals can be tied to a relevant feature that the proof exploits. Symmetry and simplicity are good candidates as the main examples of those features. In fact, we have already looked at the role of simplicity in internal mathematical explanation from a counterfactual framework.

Needless to say, Lange's emphasis in unification is already in the CTE (and, in fact, many other theories of explanation), since the power to unify diverse phenomena (or their cases) is very often a way to grasp why a theorem holds or just get more understanding of the theorem through a proof that exploits that unity. For instance, Henkin's proof provides us with unification by generalizing completeness to other domains, which is what makes it a good example of an explanatory proof. In the case of the Four-Colour Theorem, the lack of that salient feature may be what is behind its non-explanatoriness. We find, in principle, no way to exploit any symmetry, simplicity or any other relevant feature in the proof. The picture explaining the theorem of triangular numbers, even though it is hardly a proof, can be seen as displaying some sort of symmetry, which is the feature that makes it so easy for us to grasp the truth of the theorem and to extend the picture to all natural numbers and *see* that the formula works.

Other aspects, such as the idea that the relevant feature may depend on the context, as is easily inferred from our analysis in this work, is already within our account of the counterfactual theory.

As for the question of which theory does better in accounting for mathematical explanations, it looks like both the counterfactual theory and Lange's account point at relevant aspects of mathematical explanation. There is, in principle, no problem with accepting both theories as good ones, which would ultimately leave us with some sort of pluralism of theories of mathematical explanation, where it would not be too problematic to have different perspectives to account for the same issues. Both accounts capture interesting features of internal mathematical explanations and, moreover, they even seem to highlight similar features of explanations, such as unification and simplicity. This points at the idea that these two approaches are compatible, in some sense. The choice of the most suitable account of explanation might even depend on the particular case that is being considered.

In addition, I find no problem in thinking that one approach can benefit from the other one. Following this line of thought, Lange's analysis of symmetry (this being its main novelty) could be incorporated in the CTE in order to make it a more exhaustive account of mathematical explanation. This is particularly useful for the case of internal mathematical explanation, which is somewhat outside of Knowles and Saatsi's scope. However, in the present work it is argued that the CTE can also account for mathematical explanation of mathematical phenomena.

4. Internal and external mathematical explanation: two cases of the same phenomenon?

A question that arises given our analysis so far is that of the contrast and separation between internal and external mathematical explanations. It might seem like, ultimately, the debate at issue is the same for internal and external mathematical explanation. From the present work and the fact that we can account for both kinds of mathematical explanation within the same framework, we come to the question of whether those two separate explanations can actually be just two sides of the same phenomenon or, at least, be not that far apart. The idea that, in the end, internal and external mathematical explanation are two sides of the same phenomenon is appealing, given that we usually prefer accounts and explanations that can unify separate phenomena by finding common principles in them, so let us very briefly address the issue of internal and external mathematical explanation being two occurrences of, at least, a similar phenomenon.

Explaining, taken in a broad sense as a human activity, is present in all areas of research and knowledge, and in many contexts of our ordinary lives. It has to do with our need to get explanation of the world both in the sense of obtaining answers to *why*-questions, and also in the sense of obtaining understanding.

Of course, the fact that both kinds of explanation can be accounted for in the same framework does not imply that they are the same thing. However, it makes sense to consider explanation and understanding as something *global*, closely related to human nature and adopting diverse forms depending on the subject. Thus, I am not claiming that internal and external mathematical explanation are the same phenomenon. However, I do think that there is a sense in which explanation is a whole and the diverse forms it adopts depending on the subject must have something in common. In the same way, mathematical explanation unifies internal and external mathematical explanations, in that they share the mathematical apparatus as the central source of explanatory power. This also gives us an idea of the usefulness of mathematics in providing us with explanations of the most separate phenomena, unifying (at least, partially) some aspects of empirical and mathematical facts. This fact illuminates at least two features of mathematical explanation: (1) it says something about the phenomena that are explained by using mathematics and that there are some structural features in common between empirical and mathematical facts and (2) it says something about mathematics itself, since mathematics, as the discipline that deals with structures, can help account for very different phenomena, which gives an idea of how explanatorily powerful this tool is.

In conclusion, without aiming at unifying mathematical explanation in a strong sense, mathematics provides some sort of unification in explanations, given that the same apparatus provides us with answers to *why*-questions and understanding about two, in principle, separate areas: internal and external applications of mathematics, so both mathematical and empirical facts and events.

5. Conclusions

In this chapter, a broad overview of the classical approaches to intra-mathematical explanation was provided, with the aim to introduce the subject as it has been treated in the literature. The main aim of the chapter was to provide an adaptation of the CTE to mathematical explanations within mathematics and, for that, we have tested the CTE in light of the cases that were presented in Chapter 4. The result is that the CTE is a good framework to account for the cases and also their explanatory power, or lack of explanatory power, that being the case.

The CTE shows how we can think of mathematical explanations of mathematical phenomena in counterfactual terms and account for their explanatory features.

In terms of the relevant explanatory features, for instance, as the case of the completeness theorem shows, some mathematical explanations present a relevant degree of topic generality. This is the case of Henkin's proof in contrast to Gödel's original version. Henkin's proof can be generalized to many other domains, which directly translates into a higher degree of topic generality, in that it proves completeness for more systems than just first-order logic. The proof integrates different subject matters, thus providing a high level of generality. That is, in terms of allowing new what-if questions to be asked and answered, or makes such questions easier to answer. This is what we mean when we say that a theory is explanatorily valuable in terms of the counterfactual theory.

Moreover, it has been seen that the framework can account for two meanings of 'explanation' we distinguished in Chapter 3: a strong sense of explanation related to the *why* of the *explanandum*, and a weaker sense of explanation, more related to our understanding (and cognitive salience, in Knowles and Saatsi's terms). Thus, mathematical proofs, as the main forms an internal mathematical explanation can take, can explain not only by providing us with (epistemic) understanding, but also by telling us why some mathematical statements hold.

Two of the main advantages of the account, as we have seen, are monism and lightweightness. That is, it can account for both kinds of mathematical explanation and still not forcing a platonistic vision regarding the question of mathematical ontology[80]. Therefore, this view can be attractive to those wishing to account for mathematics' explanatory role and their intuition that there is some sense of "objectivity" working in mathematics, and still deny or remain agnostic about the existence of mathematical objects.

As much as the CTE seems to be a good candidate to account for mathematical explanation, there is room for some improvement. In the present chapter, we have found that the CTE can benefit from some concepts and ideas from other approaches. Here, Bueno and Colyvan's (2011) inferential account provides some ideas on how to account for the role of choice and context dependence, by including a relevant role of interpretation of the mathematical structures we use to explain.

Besides, the framework is flexible enough to be compatible with interesting features from other theories, such as the ones pointed out above and Lange's account on why proofs are (or are not) explanatory based on the fact that proofs exploit a property at the setup of the proof that is then relevant to the result to be explained.

---

[80] See the analysis in Chapter 2.

Lastly, we have briefly addressed the issue of internal and mathematical explanations being two sides of the same coin or, at the very least, just have some relevant features in common. The use of mathematics by itself as an explanatory tool is a good indicator that internal and external mathematical explanations share crucial features and sometimes work in the same way.

Final remarks

After this journey through how mathematical explanations work and also exploring the possibility of constructing a unified theory of mathematical explanation for both internal and external mathematical explanations, there are some conclusions that we can now gather. Undoubtedly, there are many more cases that can be taken into account and many more analyses to be made on the subject, but from the research above we can already draw some tentative conclusions concerning the role of mathematics in explanations. Still, this work provided an analysis of mathematical explanation from diverse perspectives.

A significant part of the dissertation (Chapters 1-2) was dedicated to extra-mathematical explanation, both to a revision of how the debate of the indispensability of mathematics developed in recent years, and an assessment of the EIA debate and its ontological implications more generally, focusing on the analysis of theories of extra-mathematical explanation.

Chapters 3, 4 and 5 were dedicated to intra-mathematical explanation. A broad overview of the classical approaches to intra-mathematical explanation was provided, as well as a deeper analysis of the concept of proof and a study of cases concerning the application of the CTE to particular cases of intra-mathematical explanations and an analysis of their explanatory power.

The Counterfactual Theory of Explanation

As we know, the CTE was examined across the dissertation in order to determine whether it is suitable to account for mathematical explanation, both in applications to empirical phenomena and within mathematics, given that one of the aims driving the study of the counterfactual theory has to do with the goal of establishing a unified theory of explanation, that is, a theory that can account for all kinds of mathematical explanation. Having a unified and systematized theory of explanation that offers a more complete view of mathematical explanation is probably a better option than to simply rely on explanatory virtues to identify cases of intra- and extra-mathematical explanations. In this sense, approaches such as that of Knowles and Saatsi's look much more promising than the enumeration of explanatory virtues or other theories with a much narrower scope.

After assessing its role in applications of mathematics, the framework was adapted in order to accommodate cases of intra-mathematical explanations, with the aim to see whether the CTE can be a suitable framework to account for extending it to all kinds of explanations. This included a study of several cases of mathematical proofs and their explanatory power, or lack thereof.

Two of the main advantages of the account, as we have seen, are monism and lightweightness, in the sense that the view does not force a platonic perspective regarding mathematical ontology. The CTE accounts for how we can think of mathematical explanations in counterfactual terms, and also for their explanatory features. Besides, it can account for two meanings of explanation we distinguished in Chapter 3 following Brown's characterization, a strong (more objective) sense related to the *why* of the *explanandum*, where explaining in terms of the CTE turns out to be particularly useful, and a weaker sense of explanation, more related to understanding and the easiness with which we grasp concepts and regularities.

The view has been analysed as a good candidate for accounting for mathematical explanation, but it can be improved. For instance, we found that the CTE can benefit from some concepts and ideas in other approaches, such as Bueno and Colyvan's focus on the importance of the selection and interpretation of the models we use in order to explain, in their inferential account. Apart from the role of providing inferences, these authors point at other roles, such as the role of unification (of disparate theories), aiding in the obtainment of novel predictions, providing explanations of empirical phenomena, and so on. In particular, for extra-mathematical explanation, the view includes a salient role of choice and interpretation, as well as the dependence on the context, in the sense that, when applying a mathematical theory to explain an empirical phenomenon, we need to choose the right structure or mapping (which can be partial) between the empirical set up and the mathematics (see Bueno and Colyvan 2011).

Moreover, the counterfactual account can benefit from Lange's contribution, according to which the relevant feature that is exploited in the setup of a proof is what makes the proof explanatory. This idea, as has been argued (see Chapter 5) is compatible with the thesis that a good explanation establishes patterns of counterfactual dependence between the *explanans* and the *explanandum* by identifying the relevant variables, since the relevant counterfactuals can be tied to a relevant feature in the proof. Lange also points at symmetry and simplicity as the main examples of those features, and also stresses the relevance of unification, just as the CTE does

As for the CTE and the use of counterfactuals, which has been thought of as a problem for the account, it becomes clear that formulating counterfactuals that *change* mathematical facts does not correspond to our way of reasoning when we think counterfactually. Indeed, we do not need to formulate that sort of counterfactuals (counterpossibles) in order to apply mathematics to our *explananda*. Rather, in explanations of empirical phenomena we tend to keep mathematics *as it is* and apply the relevant structures to the phenomenon in question. In extra-mathematical explanations, our counterfactuals are directed to what we want to explain, so the range of variables is in the physical side of the explanation, and not in the mathematical apparatus we are using to explain that phenomenon. It is certainly odd to treat mathematical objects counterfactually in intra-mathematical explanation. This is probably for reasons related to the structure of mathematics itself, which makes counterpossible reasoning certainly not the kind of operation we do when we are trying to explain physical or even mathematical facts. The formulation of counterpossibles may be useful as a kind of *thought experiment* in some contexts, but it is not the usual counterfactual thinking in explanations. It is hard to imagine how antecedents picturing an even 3 or a prime 12 would help us understand other mathematical facts or facts of the world, since these are not adequate frames to obtain more unification or generality in explanations. In general, the relevant counterfactuals for intra-mathematical cases need not be counterpossibles, but rather a variation of cases. For instance, "what would be the case had we chosen the case for $n = k$ rather than $n = m$?".

This provides us with clues about what is exactly doing the explanatory work. If our relevant counterfactuals for extra-mathematical explanations are within the empirical information, that is, they leave the mathematical apparatus unchanged, then it looks like it is not the mathematical entities the ones doing that explanatory work, but rather what is happening is that the mathematical structures help us see explanatory relations that were already in the physical facts.

From this perspective, the issue of counterpossibles should not represent a problem for the success of the counterfactual theory of explanation, which might still have potential in accounting for mathematical explanations, both internal and external. Some alternatives can be explored, such as non-classical logical systems, where the issue of counterpossibles can be avoided. However, counterfactuals are part of our reasoning, and sometimes we even reason over impossible situations, just to get more clarification in our original subject and without literally *meaning it*. Besides, we formulate counterfactuals (and counterepossibles) with specific targets with the purpose of clarification, so the assumption that all counterpossibles are unacceptable is simply the wrong approach. To sum up, the use of counterpossibles can be accepted in mathematical reasoning without carrying a problem for the account.

The CTE and internal mathematical explanation

Since the CTE had already been applied to external mathematical explanation, in Knowles and Saatsi's work, our focus went to adapting the framework to accommodate internal mathematical explanations. This was the main goal of our case study (Chapters 4 and 5, fundamentally). Our case study helped us analyse several features of internal mathematical explanation and, more particularly, mathematical proof.

For instance, the completeness case illustrates that, sometimes, the fact that we have a proof for a particular theorem does not stop the mathematicians' search for better and more explanatory proofs. The contrast between Gödel's and Henkin's proof shows that the latter is more explanatory and presents far more power of unification, since it is applicable to a broader spectrum of mathematics.

The debate concerning the Four-Colour Theorem made clear that some mathematical proofs fail to be explanatory. Together with our toy example of the triangular numbers for n smaller or equal to 5, it shows that proofs by exhaustion, while successful in establishing mathematical facts, yield no explanation of the phenomena in question.

The picture explanation opens the debate of whether visual elements such as pictures and diagrams can constitute proofs by themselves, and the use of visual elements in mathematical proofs, more generally. It also goes to the point that the criteria for constituting a proof are not clear and there might be some mathematical explanations that are not proofs.

Indeed, in this study, it becomes clear that not all mathematical proofs explain the mathematical fact in question, and also that there is more to internal mathematical explanation than mathematical proof. Some proofs might have results that go beyond that of establishing the truth of a theorem. These results may include establishing relations between disparate mathematical areas by unifying, connecting or subsuming them, which is also of interest in the analysis of mathematics' explanatory role. Some explanations in mathematics are part of a conceptual recasting of a discipline or even auxiliary elements in context where the formal proof is too difficult to follow (this is where picture explanations come in handy, for instance).

In addition, the analysis of mathematical induction as a method of proof in mathematics has brought up the debate on whether those proofs can be considered explanatory, since many authors have argued for their non-explanatoriness.

Besides, the concept of proof as is traditionally taken has been challenged by mathematical practice, where a less strict (or *more informal*) concept of proof is at work. From our analysis of mathematical proof (see Chapter 3, fundamentally, but also Chapters 4 and 5), we got to the conclusion that mathematical proof, as performed by mathematicians, goes beyond logical deductions from axioms to conclusion. In this concept, the standards are far less strict and there is room for incorporating diagrammatic thinking, computer aids and so on. Another conclusion, which was already an accepted idea at the outline, is that there is a contrast between proofs that are limited to establishing a mathematical fact, and proofs that also *explain* that mathematical fact.

The EIA and the ontological question

As has been seen, one of the main motivations behind most current studies concerning mathematical explanation is the ontological question about mathematical entities. The present dissertation provides an assessment of the current state of the debate and some ideas of what its development should be.

Departing from the Quine-Putnam IA debate, we saw both main responses from the nominalist side: Field's program and diverse *easy-road* approaches to the issue. Eventually, platonists developed a new strategy by focusing on the substantial explanatory role of mathematics in particular extra-mathematical explanations (such as the cicada case) resulting in the EIA debate. Here, we assessed the debate between Baker, on the platonist side, and Knowles and Saatsi on the opposite side, to conclude that the kind of substantial explanatory role that mathematics provides is not one that, by itself, commits us to the existence of mathematical entities. Therefore, this kind of role that mathematics plays in explanations is compatible with a thin realist or a nominalist view on abstract entities. It is compatible to argue that mathematical objects do not exist in a strong platonist sense and still believe that they have explanatory power. In this sense, since these views can make sense of mathematical entities without a strong ontological commitment, it is reasonable to drop the strong platonist view (such as heavy-duty platonism) and rather go for a light-weight platonist or nominalist approach.

Given that mathematical explanations of empirical phenomena of the sort of the cicada case can be accounted for within the CTE, which does not imply a platonistic approach to mathematical entities, there is no reason to hold a strong platonist view on mathematical ontology, while acknowledging that the view is not ruled out either.

One of the key questions, regarding the issue of mathematical ontology in light of the EIA debate is the question of whether there is a fact of the matter about which one of the different approaches is right. Here, perhaps the main conclusion of the present dissertation regarding this matter is that there is no fact of the matter as to which ontological picture accounts for mathematical entities better, at least as far as the EIA debate goes.

Besides, the idea that if mathematics had been different then the empirical facts would have been different is highly counterintuitive. It probably makes more sense to think that if empirical facts had been different, then the mathematical structures we use to describe and explain them would have been different as well. This idea is in tune with a concept of mathematics where it is a product of human reasoning. Following this path, there is a way in which the concept of primeness can be thought as just a way of measuring the periods of time for the cicada case.

According to Maddy's theses, we can believe a theory and still remain agnostic or instrumentalist about the existence of its objects, so the same happens in natural sciences, with posits such as infinite wires and average families. Those posits constitute tools that we use in order to facilitate our analyses. The case of mathematics could be analogous, in the sense that the same attitude we have towards average families could be behind our relationship with mathematical entities.

An explanation of the cicada case, agreeing with Leng (2010), could be a good one in that it respects facts about the years and their succession and it succeeds in accounting for the phenomenon, but being realist about periods of time is something different from being a realist about mathematical objects. The acceptability of mathematical statements would come from their being consequences of accepted axioms, so the fictionalist approach is still an alternative.

In a scenario where we have no strong reason to hold a platonism such as a heavy-duty platonist approach, it makes sense to drop that view. Even though there are no conclusive evidence to directly refute heavy-duty platonism, what we found out is that we have no compelling reasons to take mathematical explanations as involving mathematical objects in a strong platonic sense. This being the case, it is reasonable to reject heavy-duty platonism on the grounds that a lighter metaphysics, whenever possible, is preferable. However, from the approaches analysed here (with Baron on one side and Knowles and Saatsi on the other) between a lighter realism and a nominalist approach the question is ultimately of choice or interpretation, which would probably ultimately depend on our intuitions.

This provides us with an important conclusion, which is that we have no reason to believe that mathematical objects exist in a strong sense. Another issue is whether Thin Realism or Arealism is the best account of mathematical ontology. As Maddy points out, the difference between thin realism and nominalism is just a matter of interpretation, as we have seen from Leng's and Maddy's analysis (Chapter 1). If there is no *fact of the matter*, then it could be that Maddy had a point when she argued that the question of whether numbers exist or are just fictions has no deep answer, given that thin realism and arealism are equally successful at accounting for the nature of pure mathematics.

We have gone through other views. Baron, for instance, argues that mathematical structures do make a difference, in the sense that mathematical facts carry non-descriptive information about facts of the world. Therefore, we should believe in the existence of mathematical facts. This does not carry a commitment to the existence of mathematical *objects*, so the view is compatible with both heavy-duty and thin platonism. However, if this is the case, given that the EIA does not provide a clear distinction between the ontological picture drawn from thin platonism and nominalism, then Baron's picture of mathematical explanation is also compatible with nominalism. This leaves his proposal compatible with any ontological view, which is certainly far from his original purpose. For us, it is more evidence that there is no reason to hold a strong nominalist approach.

Furthermore, the significance of this side of Baron's approach for our purposes is that his picture of mathematical explanation does not settle the ontological question. The fact that both light-weight platonism, heavy-duty platonism and even nominalism fit into his account of mathematical explanation is far from his purpose (of deciding the question for some sort of light-weight platonism). Perhaps no account settles the ontological question at all. This is not to say that Baron's approach does not provide some interesting elements of analysis to our study of mathematical explanation. For instance, his analysis of the importance of relevance issues, which highlight that the explanation should not include irrelevant information or tools such as the informational test, which was even applied in our analysis of internal mathematical explanation in order to clarify the non-explanatoriness of the 4CT (Chapter 5).

In addition, we find no substantial difference from the ontological commitments that stem from internal and external mathematical explanation, so we can conclude that in both cases the ontological commitment we need is the same, so the same ontological picture for mathematical ontology works for all mathematical explanations.

This analysis, however, has a downside: we cannot really tell thin realism and nominalism apart, at least not in the context of the EIA debate. For those wanting a radical nominalist approach, this does not work. The same happens for those wanting thin realism as the only correct view.

Those who argue that there is no fact of the matter (such as Maddy and Leng) between a thin form of realism and arealism have a point. If this is the case, then it makes sense to let it depend on our intuitions. Baron's strategy of providing a theory of explanation that accommodates a lightweight approach to mathematical entities did not work in forcing a light-weight platonism. Indeed it only works by allowing any other ontological perspective, there are indicatives that those looking for a well-established independent theory of mathematical explanation that is neutral on the ontological side have a point. This leads to the idea that the ontological question is not solvable via the analysis of the explanatory role of mathematics, and thus undermines the EIA strategy itself. However, it gives us the opportunity to choose our preferred view.

To sum up, one of the main conclusions of this work is that it is highly likely that the EIA debate is not the right place to look for an answer to the ontological question. Ultimately, no account for mathematical explanation seems to settle the case regarding which is the correct view about the existence of mathematical entities. In the end, it looks like the choice of our preferred theory of mathematical ontology will depend on our interpretation, which heavily relies on our metaphysical intuitions, as we have seen with the approaches presented by the several authors analysed in this work.

Explanatory and non-explanatory proofs

One question is whether mathematics is explanatorily indispensable to account for empirical phenomena. Another question is the kind of explanatory role that mathematics presents. In the present work, some evidence has been put forward as to the kind of explanatory role that mathematics plays in explanations, following Brown's distinction between a strong (objective) and a weak (epistemic) sense of understanding. There is also an attempt to provide a study of the fundamental explanatory role of mathematics. Here, several issues can be put forward.

For instance, the distinction between explanatory and non-explanatory proofs (or explanations, in general) has to do with cognitive ease, but also with the ability of the proofs to convey answers to our *why*-questions regarding phenomena. Besides, being cognitively salient is related with our access to a particular mathematical fact, that is, our ability to grasp the relevant *truths* with relative ease. Of course, there are many variables in this process, related with our background knowledge, the context, the balance between simplicity and detail of the proof, and so on. Therefore, there are multiple cases for which we cannot say that mathematics' role is merely that of providing us with tools for the analysis. This is opposite to Brown's thesis that abstract entities are not explanatory in a substantial way and are thus limited to providing models and analogies. Our conclusion is quite different since, as has been seen, there has to be something beyond mere cognitive salience, in that mathematics explains in a more objective way. Just to give an example, when addressing Henkin's proof, it was shown that its explanatory power is not just a matter of improving cognitive salience or understanding, since the proof also explains why first-order logic is complete, that is, it also explains in a deeper sense.

This does not undermine the epistemic role of mathematical explanation, though. The notion of cognitive salience itself, as one of the main features of mathematics in explanations, highlights the dimension of explanation that has to do with our (limited) cognitive capacities as human beings, as well as our preferences and abilities to explain and understand based on our background and the dependence on the context or the audience. In the end, explaining, as a human activity, is related to our need for answers when we wonder about facts of the world. This dimension of mathematical explanation is compatible with a (related) more objective explanatory role, as we have seen. Our account of mathematical explanation does not rule out mathematical objects or properties bearing objective, explanatorily relations to phenomena. In fact, we found that those two senses of explanation distinguished by Brown do take place in mathematical explanations of empirical and mathematical facts, so we find a more *subjective* explanatory role connected to understanding, and a more objective one, related to establishing facts of the world and answering the question of *why* those empirical phenomena or mathematical facts are the case.

As for the features that make explanations successful or *good explanations*, we found that the vast majority of accounts point at the same basic explanatory properties or virtues that a proof can show, in various degrees. These main explanatory features include simplicity, generality, unification and cognitive salience.

Both simplicity and unification point at the same basic features of good mathematical explanations. Moreover, as stated above, there is a clear connection between unification and abstraction, in the sense that explanations that can account for more than just one phenomenon are usually the product of abstractions. Abstraction makes explanations simpler, at least in principle, which leads to easier explanations to grasp and understand, contributing to their cognitive salience. As Weber and Verhoeven (2002, 303) point out, "one of the generally accepted aims of explanation in the natural sciences and social sciences is *unification*. Unifying events consists in showing that two or more different events are instances of the same (set of) law(s) of nature", so unified explanations are usually preferred among scientists and mathematicians. Ultimately, simplicity and unification combined are two of the central explanatory virtues pointed at in virtually all accounts and also present in explanations known for their presenting a high explanatory value. This is the case for both empirical and mathematical contexts. Indeed, the connection between unification and abstraction is relevant in a way that affects the debate about abstraction, which has received much interest in the literature as an explanatory tool and also as a means to achieve new results both in mathematics and natural sciences.

Other relevant, though minor, conclusions of this work have to do with the different standards mathematicians seem to use when evaluating proofs. As has been seen, Colyvan examined different proofs of the same result to conclude that several standards can coexist and they can point at several ways in which mathematics has explanatory value. There is a tendency to accept the idea that many different explanatory features may have a role in explanations, in various degrees. Simplicity, the degree of abstraction, their power to unify different domains, and other explanatory virtues may all have a relevant role in explanations, as well as subjective issues and the particular context or audience to which we present an explanation. From the contrast between Steiner's and Kitcher's approaches to mathematical explanation, it may well be that both of them capture opposite but equally relevant features of mathematical explanation: that of identifying the relevant property and ability to obtain generalizations by varying the explanation, and that of unifying diverse phenomena under the same explanatory pattern.

Relatedly, from our study of mathematics from an educational viewpoint, some conclusions can be drawn. Firstly, it provides more support to the thesis that there are no fixed standards when evaluating proofs. Second, it is in correspondence with the thesis that unification has a crucial role in internal mathematical explanation, since the unification of diverse mathematical facts under the same explanatory patterns makes them easier to understand by students. Third, it highlights the importance of the dimension of understanding in contrast to the commonly stressed feature of rigour. Hanna (2010) claims that rigour is secondary in importance to understanding, with the idea that it is crucial that we obtain convincing proofs which lead to mathematical understanding, rather than proofs limited to establishing the truth of the relevant statement. This is in line with the idea that, often, the concept of proof at work in mathematical practice is a less formal one, with providing understanding in the horizon, and which can incorporate various resources such as computation or visual elements.

A, somewhat, open conclusion

It has been pointed out that finding or constructing a unified theory of explanation would be of great value, since then all kinds of mathematical explanations can be accounted for in the just one general theory or framework. If the CTE can occupy that position, then all kinds of mathematical explanation can be analysed from a counterfactual perspective.

There are reasons to believe that internal and external mathematical explanations have many features in common. This is not to say that they constitute the same phenomenon, but there is a sense in which both internal and external explanations are unified under the same explanatory tool, that is, the mathematical apparatus. By using mathematics, both mathematical explanations of empirical and mathematical facts share a substantial explanatory core. Besides, explanation and understanding can be considered something *global* in that they are part of human nature. There is a sense in which the diverse forms explanations adopt depending on the subject have something in common. The fact that we use mathematics as a central source of explanatory power also unifies, in a way, internal and external mathematical explanation.

From this, there are two relevant features of mathematical explanation that can be pointed out, regarding the phenomena that are explained and also mathematics itself. The first one is that there are some structural features in common between empirical and mathematical facts, since we use (the same) mathematical structures to explain them, and the second one is that the mathematical structure can account for very different phenomena, which says something about its high explanatory power.

Therefore, there is still some hope for the aim of providing a unified account of mathematical explanation, which works for both the internal and the external uses. The basic principles of the counterfactual theory seem to apply to both internal and external mathematical explanation.

However, we have also seen some room for improvement of the framework and even other accounts, such as Lange's, that might be also suitable to account for at least some cases of mathematical explanations.

If we are to answer which theory does a better job, it might be that both theories, the CTE and Lange's account, highlight relevant features of mathematical explanation. This being the case, there is no problem in accepting both accounts as good ones, given that Lange's approach is not incompatible with the counterfactual theory. Furthermore, it may even contribute to highlighting relevant features that were initially missing in the CTE. There is no problem in thinking that one approach can benefit from the other one. For instance, Lange's analysis of symmetry (which does not appear in Knowles and Saatsi's counterfactual account) can be incorporated into the CTE with relative ease, by simply taking it as another relevant property of good mathematical explanations that can be put in counterfactual terms to explain the cases showing symmetry as a salient property. This feature would make the theory a more exhaustive view on explanation, in particular, for intra-mathematical explanation. Besides, both accounts seem to point at similar features in explanations, such as unification and simplicity, which contributes to strengthen the idea that, in a way, these two approaches are compatible after all.

In the end, our approach to mathematical explanation might need to be broad enough to accommodate these multiple features that can make a proof explanatory, or even acknowledge that, perhaps, the right approach to mathematical explanation has to be an open one, compatible with the idea that several frameworks work in order to account for the explanatory role of mathematics. If that is correct, then we are accepting that some kind of pluralism might make sense in order to account for mathematical explanations.

This would ultimately leave us with some sort of an open landscape, so it might even be the case that Baron was not completely wrong when he pointed at a pragmatist approach as a possibility for accounting for mathematical explanation, even though he ended up rejecting this approach. Several theories can have a role in explaining mathematical explanation, and some will be more suitable than other depending on the particular case we are considering. It does not seem too problematic to have different perspectives from which to analyse diverse mathematical explanations.

Still, there is no doubt that thinking counterfactually is an activity in which we engage on a regular basis in order to highlight the relevant conditions in a particular situation in need to be explained. It is in tune with our ordinary reasoning when we –informally – think in terms of "what would happen if *x* were the case?", or "what if things had been different in such and such ways?". In this sense, it is a theory that works according to our intuitions.

It looks like, while the CTE can account for both internal and external mathematical explanation and, thus, it can be a good unified theory of mathematical explanation, it is, nevertheless, not the only available account. Other approaches may still have a place in the debate.

To conclude, there is still much work to be done in order to obtain a satisfactory account of mathematical explanation. The present dissertation aimed at providing more elements of analysis and some clarification concerning several issues on mathematical explanations, both applied to empirical and mathematical facts. It is now clearer what an explanatory explanation is, as well as the role of some key features such as simplicity and unification. Similarly, as far as the ontological question is concerned, according to what has been seen, the debate is far from closed, so it is safe to assume that the interest that explanation in mathematics has is an indicative that many more analyses on the subject are yet to come.

References

- Achinstein, P. (1983). *The Nature of Explanation*, New York: Oxford University Press.
- Appel, K. (1984). “The Use of the Computer in the Proof of the Four Color Theorem”. *Proceedings of the American Philosophical Society, 128*(1), 35-39.
- Atiyah, M. (1984). “Interview with Michael Atiyah". *Mathematical Intelligencer*, Volume 6 Issue 1, 9-19.
- Azzouni, J. (1997). “Applied mathematics, existential commitment and the Quine-Putnam indispensability thesis”. *Philosophia Mathematica,* 5(3), 193–209.
- Azzouni, J. (1994). *Metaphysical Myths, Mathematical Practice*, Cambridge: Cambridge University Press.
- Azzouni, J. (2004). *Deflating Existential Commitment: A Case for Nominalism*. Oxford: Oxford University Press.
- Azzouni, J. (2009). “Why Do Informal Proofs Conform to Formal Norms?”, *Foundations of Sciences*, 14, 9-26.
- Azzouni, J. (2010). *Talking About Nothing: Numbers, Hallucinations, and Fictions*. Oxford: Oxford University Press.
- Azzouni, J. (2012). “Taking the Easy Road Out of Dodge”. *Mind*, 121(484), 951–965.
- Baker, A. (2003). “Does the Existence of Mathematical Objects Make a Difference?”. *Australasian Journal of Philosophy*, 81:2, 246-264.
- Baker, A. (2005). “Are there genuine mathematical explanations of physical phenomena?”. *Mind*, 114(454), 223–238.
- Baker, A. (2009a). “Mathematical Explanation in Science”. *British Journal for the Philosophy of Science*, 60: 611–633.
- Baker, A. (2009b). “Mathematical Induction and Explanation”. *Analysis*, 70: 681–690.
- Baker, A. (2016). “Parsimony and Inference to the Best Mathematical Explanation”. *Synthese*, 193: 333–350.
- Baker, A. (2017a). “Mathematics and Explanatory Generality”. *Philosophia Mathematica*, 25:2, 194–209.
- Baker, A. (2017b). “Mathematical Spandrels”, *Australasian Journal of Philosophy*, 95: 779–793.

- Baker, A. and Colyvan, M. (2011). “Indexing and Mathematical Explanation”. *Philosophia Mathematica*, 19 (3), 323–334.

- Balaguer, M. (1996a). “A Fictionalist Account of the Indispensable Applications of Mathematics”. *Philosophical Studies*, 83, 291–314.

- Balaguer, M. (1996b). “Towards a Nominalization of Quantum Mechanics”. *Mind*, 105, 209–26.

- Balaguer, M. (1998a). *Platonism and Anti-Platonism in Mathematics*. Oxford: Oxford University Press.

- Balaguer, Mark, "Fictionalism in the Philosophy of Mathematics", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/fictionalism-mathematics/>.

- Baldwin, J. (2016). “Foundations of Mathematics: Reliability AND Clarity: The Explanatory Role of Mathematical Induction”. in J. Väänänen et al. (eds.), *Logic, Language, Information, and Computation, Lecture Notes in Computer Science*, Berlin/Heidelberg: Springer, 68–82.

- Baldwin, J. (2018). “The Explanatory Power of a New Proof: Henkin’s Completeness Proof”. *FilMat 2016 Studies in the Philosophy of Mathematics*. 10.1007/978-3-319-93342-9_9.

- Bangu, S. I. (2008). “Inference to the Best Explanation and Mathematical Realism”. *Synthese*, 160, 13–20.

- Baron S. (2016). “Explaining Mathematical Explanation”. *The Philosophical Quarterly*, 66/264, 458–480. Doi: https://doi.org/10.1093/pq/pqv123

- Baron, S. (2019). “Mathematical Explanation by Law”, *British Journal for the Philosophy of Science*, vol. 70, 683-717. (First online 28 November 2017. doi: 10.1093/bjps/axx062.)

- Baron, S., Colyvan, M. and Ripley, D. (2017). “How Mathematics Can Make a Difference”. *Philosophers' Imprint*, Vol. 17, No. 3 (January 2017), 1–19.

- Batterman, R.W. (2000). “Multiple Realizability and Universality”. *British Journal for The Philosophy of Science*. 51. 115-145.

- Batterman, R. W. (2002). *The Devil in the Details. Asymptotic Reasoning in Explanation, Reduction, and Emergence.* Oxford: Oxford University Press.

- Batterman R.W. (2010). “On the explanatory Role of Mathematics in Empirical Science”. *British Journal for the Philosophy of Science*, 61, 1-25.

- Berto, F., French, R., Priest, G. *et al.* (2018). “Williamson on Counterpossibles”. *J Philos Logic* 47**,** 693–713.

- Berto, F. and Jago, M., “Impossible Worlds”. *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/impossible-worlds/>.
- Bokulich, A. (2008). “Can Classical Structures Explain Quantum Phenomena?”. *British Journal for the Philosophy of Science*, 59 (2), 217–235.
- Bonsall, F. F. (1982). “A down-to-earth view of mathematics”. *American Mathematical Monthly,* 89, 8-15.
- Brown, J. B. (2012). *Platonism, Naturalism and Mathematical Explanation,* New York, Routledge.
- Bueno, O. (2012). “An Easy Road to Nominalism”. *Mind,* 121: 967–982.
- Bueno, O. and Colyvan, M. (2011). “An Inferential Conception of the Application of Mathematics”. *Noûs,* 45(2) (June 2011), 345–74.
- Burgess, J. and Rosen, G. (1997). *A Subject With No Object*, New York: Oxford University Press.
- Campaner, R. (2012). “Commentary. *Scientific Explanation and the Causal Structure of the World* by Wesley Salmon”. *Humana.Mente. Journal of Philosophical Studies*, 2012, Vol. 21, 197 – 204.
- Cariani, F. (2011). “Mathematical Induction and Explanatory Value in Mathematics”, unpublished typescript. URL: http://cariani.org/files/MIEV.cariani.pdf (Accessed: 2nd November 2019.)
- Chirimuuta, M. (2014). “Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience”. *Synthese*, 191: 127–153.
- Chirimuuta, M. (2017). “Explanation in Computational Neuroscience: Causal and Non-causal”. *The British Journal for the Philosophy of Science* (first online 14th March 2017, doi: 10.1093/bjps/axw034).
- Chirimuuta, M. (2018). “The Development and Application of Efficient Coding Explanation in Neuroscience”, in A. Reutlinger and J. Saatsi (eds.), *Explanation beyond Causation*. Oxford: Oxford University Press, 164–184.
- Colyvan, M. (1998). “Can the Eleatic Principle be Justified?”. *The Canadian Journal of Philosophy*, 28(3): 313–336.
- Colyvan, M. (2001). *The Indispensability of Mathematics*. Oxford: Oxford University Press.
- Colyvan, M. (2002). “Mathematics and Aesthetic Considerations in Science”. *Mind*, 11: 69–78.

- Colyvan, M. (2006). "Scientific Realism and Mathematical Nominalism: A Marriage made in Hell", in C. Cheyne and J. Warral (eds.), *Rationality and Reality: Essays in Honour of Alan Musgrave*, Dordrecht: Kluwer, 225–237.
- Colyvan, M. (2010). "There is No Easy Road to Nominalism". *Mind*, 119: 285–306.
- Colyvan, M. (2012). "Road Work Ahead: Heavy Machinery on the Easy Road", *Mind*, 121: 1031–1045.
- Colyvan, Mark. (2018). "The Ins and Outs of Mathematical Explanation", *The Mathematical Intelligencer*. 40. 10.1007/s00283-018-9799-1. Also found in: http://www.colyvan.com/papers/tiaoome
- Colyvan, M., "Indispensability Arguments in the Philosophy of Mathematics", *The Stanford Encyclopedia of* Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/mathphil-indis/>. (Accessed on the 1st of August, 2019)
- Colyvan, M., Cusbert, J. and McQueen, K. (2018). "Two Flavours of Mathematical Explanation", in A. Reutlinger and J. Saatsi (eds.), *Explanation Beyond Causation*, Oxford University Press, 231–49.
- Craver, C. F. and Povich, M. (2017). "The Directionality of Distinctively Mathematical Explanations". *Studies in History and Philosophy of Science*, 63: 31–38.
- Daly, C., and Langford, S. (2009). "Mathematical Explanation and Indispensability Arguments". *The Philosophical Quarterly*, 59: 641–658.
- De Villiers, M. (1990). "The role and function of proof in mathematics". *Pythagoras* 24, 17–24.
- Detlefsen, M., and Luker, M. (1980). "The four-color theorem and mathematical proof", *Journal of Philosophy* 77(12):803-820.
- Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.
- Dreyfus, T. (1991). "Advanced mathematical thinking processes". In Tall, D. (Ed.) *Advanced Mathematical Thinking*, 25–41. Dordrecht: Kluwer.
- Field, H. (1980). *Science without Numbers: A Defence of Nominalism*. Oxford: Blackwell. (Second edition, 2016.)
- Field, H. (1989). *Realism, Mathematics, and Modality*, Oxford: Blackwell.
- Frans, J. and Weber, E. (2014). "Mechanistic Explanation and Explanatory Proof in Mathematics". *Philosophia Mathematica*, 22: 231–248.
- French, S. and Saatsi, J. (2018). "Symmetries and Explanatory Dependencies in Physics" in A. Reutlinger and J. Saatsi (eds.), *Explanation beyond Causation*, Oxford: Oxford University Press, 185–205.

- Friedman, M. (1974). “Explanation and Scientific Understanding”. *Journal of Philosophy*, 71: 5–19.
- Giaquinto, M. (2007). *Visual Thinking in Mathematics: An Epistemological Study*, Oxford University Press, 2007.
- Giaquinto, M. (2008). “Cognition of Structure”. *The Philosophy of Mathematical Practice*, DOI*:* 10.1093/acprof:oso/9780199296453.003.0003, 43-64.
- Giaquinto, M. (2016). “Mathematical Proofs: The Beautiful and The Explanatory”. *Journal of Humanistic Mathematics*, Volume 6, Issue 1 (January 2016), 52-72.
- Gödel, K. (1929). “Über die Vollständigkeit des Logikkalküls” in 1986, *Collected Works. I: Publications 1929–1936*. S. Feferman, S. Kleene, G. Moore, R. Solovay, and J. van Heijenoort (eds.), Oxford: Oxford University Press.
- Gödel, K. (1930). “The completeness of the axioms of the functional calculus of logic”, in Feferman, S., J.W. Dawson et al, (eds) (1986). *Kurt Gödel Collected Works Vol I and II*, Oxford: Oxford University Press.
- Gödel, K. (1944). “Russell's mathematical logic”, *The Philosophy of Bertrand Russell* (Library of Living Philosophers), P. Schilpp (ed.), New York: Tudor, 1951, pp. 123–153. Reprinted in Gödel 1990, 119–141.
- Gödel, K. (1944). “What is Cantor's continuum problem?”, *Amer. Math. Monthly*, 54: 515–525. Reprinted in Gödel 1990, 176–187.
- Gould, S. J. (1977). *Ontogeny and Phylogeny*. Cambridge: Harvard University Press.
- Hafner, J. and Mancosu, P. (2005). “The Varieties of Mathematical Explanation”, from book *Visualization, explanation and reasoning styles in mathematics*. Based on the meeting on mathematics as rational activity, Roskilde, Denmark, November 1–3, 2001 (pp.215-250).
- Hanna, G. (1990). “Some pedagogical aspects of proof”, *Interchange* **21**(1), 6–13.
- Hanna, G. (1995). “Challenges to the importance of proof”, *For the Learning of Mathematics,* **15**(3), 42–49.
- Hanna, G. (2010). “Proof, Explanation and Exploration: An Overview”, *Educational Studies in Mathematics* 44(1):5-23.
- Hanna, G. (2018). “Reflections on proof as explanation”, in Stylianides, A.J. & Harel, G. (Eds.) *Advances in Mathematics Education Research on Proof and Proving: An International Perspective (ICME-13 Monographs)* pp. 3-18. New York: Springer.
- Hellman, G. (1989). *Mathematics without numbers: Towards a modal-structural interpretation*. Oxford: Clarendon Press.

- Hempel, C. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.
- Henkin, L. (1949). "The Completeness of the First-Order Functional Calculus", *The Journal of Symbolic Logic* 14(3), 159-166.
- Henkin, L., (1950). "Completeness in the theory of types", *The Journal of Symbolic Logic*, 15, 81{91.
- Henkin, L. (1967). "Truth and provability" and "Completeness" Morgenbesser, S. *(ed). Philosophy of Science today*, New York: Basic Books.
- Henkin, L. (1996). "The discovery of my completeness proofs" *The Bulletin of Symbolic Logic* 2(2), 127.158.
- Hersh, R. (1997). "Prove - once more and again", *Philosophia Mathematica*, 5, 153-165.
- Hilbert, D. and Ackermann, W. (1928). *Grundzüge der theoretischen Logik*, Berlin: Springer-Verlag.
- Hitchcock C. and Woodward J. (2003). "Explanatory Generalizations, Part II: Plumbing Explanatory Depth." *Nous* 37:2, 181-199.
- Hitchcock, C. (1995). "Discussion: Salmon on Explanatory Relevance", *Philosophy of Science*, 62: 304–20. http://www. sjgarchive.org/library/ontogeny.html (Accessed 13 February 2019).
- Huneman, P. (2010). "Topological Explanations and Robustness in Biological Sciences", *Synthese*, 177: 213–245.
- Inglis, M., Mejia-Ramos, J.P., Weber, K., & Alcock, L. (2013). "On Mathematicians' Different Standards When Evaluating Elementary Proofs", *Topics in cognitive science, 5 2*, 270-82.
- Jansson L. (2015). "Explanatory asymmetries: laws of nature rehabilitated", *Journal of Philosophy*, 112 (11). 577-599. ISSN 1939-8549.
- Jansson, L., and Saatsi, J. (2017). "Explanatory abstractions". *British Journal for the Philosophy of Science*, 0, 1–27.
- Keas, M. N. (2017). "Systematizing the theoretical virtues." *Synthese* 195, (2017): 2761-2793.
- Kennedy, J., (2018). "Kurt Gödel", *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/goedel/>.
- Kistler M. (2013). "The Interventionist Account of Causation and Non-Causal Association Laws", *Erkenntnis*, 78 (1): 1–20.

- Kitcher, P. (1981). “Explanatory Unification”, *Philosophy of Science*, 48: 507–531.
- Knowles, R. (2015). “Heavy-Duty Platonism”, *Erkenntnis* 80/6: 1255-1270.
- Knowles, R. and Liggins, D. (2015). “Good Weasel Hunting”, *Synthese*, 192: 3397-3412.
- Knowles, R., and Saatsi, J. (2019). “Mathematics and explanatory generality: Nothing but cognitive salience”. *Erkenntnis*, 1–19. https://doi.org/10.1007/s10670-019-00146-x.
- Koepke, P. (2007). *Gödel’s Completeness Theorem with Natural Language Formulas,* Logik, Begriffe, Prinzipien des Handelns, Logic, Concepts, Principles of Action, Paderborn. http://www.math.uni bonn.de/people/koepke/Preprints/Goedels_completeness_theorem_with_natural_language_formulas.pdf (Accessed: 14th June 2019).
- Komatsu, K., Fujita, T. and Jones, K. (2018), “Explanatory unification by proofs in school mathematics”, *For the Learning of Mathematics*, *38*(1), 31-37.
- Lakatos, I. (1979). “What does a mathematical proof prove?” In Tymoczko 1998 (ed.), *New Directions in the Phylosophy of Mathematics*, USA, Princeton University Press.
- Lange, M., (2010a), “Why Proofs by Mathematical Induction are Generally not Explanatory”, *Analysis*, 69: 203–211.
- Lange, M. (2010b). “What are Mathematical Coincidences (and Why does it Matter)?”, *Mind*, 119: 307–340.
- Lange, M. (2013a). “How to Explain the Lorentz Transformations”, in S. Mumford and M. Tugby (eds.), *Metaphysics and Science*, Oxford: Oxford University Press, 73–98.
- Lange, M. (2013b). “Really Statistical Explanations and Genetic Drift”, *Philosophy of Science*, 80: 169–188.
- Lange, M. (2013c). “What Makes a Scientific Explanation Distinctively Mathematical”, *British Journal for the Philosophy of Science*, 64: 485–511.
- Lange, M. (2014). “Aspects of Mathematical Explanation: Symmetry, Unity and Salience”, *Philosophical Review*, 123 (4): 485–531.
- Lange, M., (2016). “Explanatory Proofs and Beautiful Proofs”, *Journal of Humanistic Mathematics*, 6 (1): 8–51.
- Lange, M., (2017). *Because Without Cause: Non-causal Explanations in Science and Mathematics*, Oxford: Oxford University Press.
- Lange, M. (2018) “Explanation by Constraint”, in A. Reutlinger and J. Saatsi (eds.), *Explanation beyond Causation*, Oxford: Oxford University Press, 15–38.

- Larvor, B. (2012), “How to think about informal proofs”, *Synthese*, 187(2), 715-730. https://doi.org/10.1007/s11229-011-0007-5.
- Leng, M. (2005). “Mathematical Explanation”, in C. Cellucci and D. Gillies (eds.), *Mathematical Reasoning and Heuristics*, London: King’s College Publications, 167–189.
- Leng, M. (2010). *Mathematics and Reality*, Oxford: Oxford University Press.
- Leng, M. (2012). “Taking it Easy: A Response to Colyvan”, *Mind*, 121 (484): 983–995.
- Leng, M. (2020). “Is There a Fact of the Matter about the Existence of Abstract Objects?” in Falguera, J. L., Martínez-Vidal, C. (eds.), *Abstract Objects: For and Against*, Synthese Library, Springer.
- Lewis, D. (1973). *Counterfactuals*, Oxford: Blackwell.
- Lewis, D. (1986). “Causal explanation”, *Philosophical papers* vol. II. (pp. 214–240). New York: Oxford UP.
- Liggins, D. (2012). “Weaseling and the Content of Science”. *Mind*, Volume 121, Issue 484, October 2012, pp. 997–1005.
- Liggins, D. (2016). “Grounding and the Indispensability Argument”, *Synthese*, 193: 531–548.
- Lipton, P. (2004). “What Good is an Explanation”, in J. Cornwell (ed.), *Explanations. Styles of Explanation in Science*, Oxford: Oxford University Press, 1–21.
- Lyon, A. and Colyvan, M. (2008). “The Explanatory Power of Phase Spaces”, *Philosophia Mathematica*, 16(2): 227–243.
- Maddy, P. (1992). *Realism in Mathematics*, Oxford: Oxford University Press.
- Maddy, P. (1995). “Naturalism and Ontology,” *Philosophia Mathematica*, 3: 248–70.
- Maddy, P. (1997). *Naturalism in Mathematics*, Oxford: Clarendon Press.
- Maddy, P. (2011). *Defending the Axioms: On the Philosophical Foundations of Set Theory*, Oxford: Oxford University Press.
- Malament, D. (1982). ‘Review of Field’s Science Without Numbers’, *Journal of Philosophy*, 79: 523–34.
- Mancosu, P. (2008) ed., *The Philosophy of Mathematical Practice*, Oxford: Oxford University Press.
- Mancosu P. (2018). “Explanation in Mathematics”, *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/mathematics-explanation/>.
- Manzano, M. (1999). *Model Theory*, Oxford: Oxford University Press.

- Manzano, M. and Alonso, E. (2014). “Completeness: from Gödel to Henkin”, *History and Philosophy of Logic*, 35(1): 1-26. https://doi.org/10.1080/01445340.2013.816555.
- Marcus, R. (2018). “The Indispensability Argument in the Philosophy of Mathematics”, in *The Internet Encyclopedia of Philosophy*. https://www.iep.utm.edu/indimath/.
- **Martínez-Vidal, C. (2018).** “Putnam and contemporary fictionalism”, *THEORIA. An International Journal for Theory, History and Foundations of Science, 33*(2), 165-181.
- Martínez-Vidal, C. and Rivas de Castro, N. (2020). “Description, Explanation and Ontological Commitment”, in Falguera, J. L., Martínez-Vidal, C. (eds.), *Abstract Objects: For and Against,* Synthese Library, Springer.
- Melia, J. (2000). “Weaseling Away the Indispensability Argument,” *Mind*, 109: 455–79.
- Melia, J. (2002). “Response to Colyvan,” *Mind*, 111: 75–79.
- Morrison, M. (2018), “The Non-Causal Character of Renormalization Group Explanations”, in A. Reutlinger and J. Saatsi (eds.), *Explanation beyond Causation*, Oxford: Oxford University Press, 206–230.
- Nerlich, G., (1979). ‘What Can Geometry Explain?’ *British Journal for the Philosophy of Science*, 30: 69–83.
- Parsons, C. (1980). “Mathematical Intuition”, *Proceedings of the Aristotelian Society*, 80: 145–168.
- Parsons, Ch. (1995). “Platonism and mathematical intuition in Kurt Gödel's thought”, *Bulletin of Symbolic Logic*, 1(1): 44–74.
- Paulson, L. (2014). “A Machine-Assisted Proof of Gödel’s Incompleteness Theorems for the Theory of Hereditarily Finite Sets”, *The Review of Symbolic Logic*. 7. 484-498.
- Pexton, M. (2014). “How Dimensional Analysis Can Explain”, *Synthese*, 191: 2333–2351.
- Pincock, C. (2011). “Mathematical Explanations of the Rainbow”, *Studies in History and Philosophy of Modern Physics*, 42(1): 13–2.
- Pincock, C. (2012). *Mathematics and Scientific Repesentation*, Oxford: Oxford University Press.
- Pincock, C. (2015). “Abstract Explanations in Science”, *The British Journal for the Philosophy of Science*, Volume 66, Issue 4, December 2015, Pages 857–882.
- Pirie, S. & Kieren, T. (1994). “Growth in mathematical understanding: how can we characterise it and how can we represent it?” *Educational Studies in Mathematics* **26**(2), 165–190.

- Plebani, Matteo. (2015). “Non-Factualism Versus Nominalism”, *Pacific Philosophical Quarterly* 98: 344 – 362.
- Plebani M. (2016). “Nominalistic content, grounding, and covering generalizations: Reply to grounding and the indispensability argument”, Synthese 193:2, 549–558.
- Putnam, H. (1975). “What is Mathematical Truth?”, *Historia Mathematica*, 2, 529-533.
- Putnam, H. (2012). *Philosophy in an age of science. Physics, mathematics and scepticism,* Cambridge, Mass.: Harvard University Press.
- Putnam, H. (1971). *Philosophy of Logic,* New York, Harper & Row.
- Quine, W.V.O. (1939). “Designation and Existence.” In Feigl and Sellars, *Readings in Philosophical Analysis,* Appleton-Century-Crofts, Inc., New York: 1940.
- Quine, W.V.O. (1948). “On What There Is”, In Quine 1980.
- Quine, W.V.O. (1951). “Two Dogmas of Empiricism”, *The Philosophical Review,* Vol. 60, No. 1. (Jan., 1951), 20-43. (Also in Quine 1980.)
- Quine, W.V.O. (1955). “Posits and Reality.” In *The Ways of Paradox*. Cambridge: Harvard University Press, 1976.
- Quine, W.V.O. (1958). “Speaking of Objects.” In Quine 1969.
- Quine, W.V.O. (1960), *Word and Object,* Cambridge, MA: MIT Press.
- Quine, W.V.O. (1980). *From a Logical Point of View.* Cambridge: Harvard University Press
- Quine, W.V.O. (1981), “Five Milestones of Empiricism”, in *Theories and Things*, Cambridge, MA: Harvard University Press, 67–72.
- Quine, W.V.O. (1986). Philosophy of Logic, 2nd edition. Cambridge: Harvard University Press.
- Rav, Y. (1999), “Why do we prove theorems?” *Philosophia Mathematica*, 7(3), 5-41.
- Resnik, M. D. (1992), “Proof as a source or truth”, in: M. Detlefsen (ed.), *Proof and Knowledge in Mathematic*s, London, Routledge.
- Reutlinger, A. (2014). “Why is There Universal Macro-Behavior? Renormalization Group Explanation as Non-Causal Explanation”, *Philosophy of Science*, 81: 1157–1170.
- Reutlinger A. (2016). “Is There A Monist Theory of Causal and Non-Causal Explanations? The Counterfactual Theory of Scientific Explanation” *Philosophy of Science*, 83:5, 733-745.
- Reutlinger A. (2017). “Extending the Counterfactual Theory of Explanation” In A. Reutlinger and J. Saatsi (Eds.) 2017, 74-95.

- Reutlinger, A., Hangleiter, D., Hartmann, S. (2018). “Understanding (With) Toy Models”, *The British Journal for the Philosophy of Science*, 69: 1069-1099.
- Reutlinger A. and Saatsi, J. (2017) (Eds.). *Explanation beyond Causation: Philosophical Perspectives on Non-Causal Explanations*. Oxford: Oxford University Press.
- Rice, C. (2015). “Moving beyond Causes: Optimality Models and Scientific Explanation”, *Noûs*, 49: 589–615.
- Rizza, D. (2011). “Magicicada, Mathematical Explanation and Mathematical Realism”, *Erkenntnis*, 74: 101–114.
- Ross, W. D. (1924). *Aristotle's Metaphysics*, Oxford: Clarendon Press.
- Saatsi, J. (2007). “Living in Harmony: Nominalism and the Explanationist Argument for Realism”, *International Studies in the Philosophy of Science*, 21 (1): 19–3
- Saatsi J. (2011). “The enhanced indispensability argument: Representational versus explanatory role of mathematics in science. *The British Journal for the Philosophy of Science 62*(1), 143–154.
- Saatsi, J. (2016). “On the ‘Indispensable Explanatory Role’ of Mathematics”, *Mind*, Volume 125, Issue 500, October 2016, 1045–1070.
- Saatsi, J. and Pexton, M. (2013). “Reassessing Woodward's Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations”, *Philosophy of Science* 80 (5): 613-624.
- Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Salmon, W. (1989). *Four Decades of Scientific Explanation*, Minneapolis: University of Minnesota Press.
- Salmon, W. (1994). “Causality without Counterfactuals”, *Philosophy of Science*, 61: 297–312.
- Skow, B. (2014). ”Are there non-causal explanations (of particular events)?” *The British Journal for the Philosophy of Science*, 65, 445–467.
- Sober, E. (1993). “Mathematics and Indispensability.” *The Philosophical Review* 102: 35-57.
- Stalnaker, R. (1968). “A Theory of Conditionals”, *American Philosophical Quarterly*, Monograph Series 2, repr. in Harper et al. (eds.), 1981, *Ifs*, Dordrecht: Reidel: 41–55.
- Starikova, I. and Giaquinto, M. (2017). “Thought Experiments in Mathematics”, Stuart, Michael T., Yiftach Fehige, and James R. Brown (eds.), *The Routledge Companion to Thought Experiments*, London and New York: Routledge.

- Steiner, M.. (1978). “Mathematical Explanation”, *Philosophical Studies*, 34: 135–151.
- Teller, P. (1980). “Computer Proof”, *The Journal of Philosophy*, Vol. 77, No. 12, 797-803.
- Tennant, N. (1986). “The withering away of formal semantics?”, *Mind and Language* 1, 302-318.
- Tymoczko, T. (1998). “The Four-Color Problem and Its Significance”, T. Tymoczko (ed.), *New Directions in the Phylosophy of Mathematics*, USA, Princeton University Press, 243-266.
- Urquhart, A. (1990). “The Logic of Physical Theory”, in A. D. Irvine, ed., (1990): *Physicalism in Mathematics* (Dordrecht: Kluwer): 145-154.
- Van Fraassen, B. (1980). *The Scientific Image.* Oxford: Oxford University Press.
- Villiers, M. (1990). “The Role and Function of Proof in Mathematics”, *Pythagoras.* 24. 17-24.
- Weber, E., and Verhoeven, L. (2002). “Explanatory Proofs in Mathematics”, *Logique Et Analyse* 179: 299-307.
- Weslake, B. (2010). “Explanatory Depth”, *Philosophy of Science, 77*(2), 273-294.
- Williamson, T. (2007). *The Philosophy of Philosophy*, Oxford: Blackwell.
- Williamson, T. (2017). “Counterpossibles in Metaphysics”, in Armour-Garb, B., & Kroon, F. (Eds.), *Philosophical fictionalism*. Forthcoming.
- Woodward, J. (1989). “The Causal/Mechanical Model of Explanation”, in *Scientific Explanation*. *Minnesota Studies in the Philosophy of Science, volume 13*, P. Kitcher and W. Salmon (eds.), 357–383.
- Woodward, J. (2003), *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Woodward, J. and Hitchock, C. (2003a). “Explanatory Generalizations, Part 1: A Counterfactual Account.” *Noûs*, 37: 1–24.
- Woodward, J. and Hitchock, C. (2003b). “Explanatory Generalizations, Part 2: Plumbing explanatory depth”, *Noûs* 37: 181–199.
- Yablo, S. (1998), “Does Ontology Rest on a Mistake?”, *Aristotelian Society* (Supplementary Volume), 72: 229–261.
- Yablo, S. (2002). “Go Figure: A Path Through Fictionalism,” *Midwest Studies in Philosophy*, 25: 72–102.
- Yablo, S. (2005). “The Myth of the Seven,” in *Fictionalism in Metaphysics*, M. Kalderon (ed.), New York: Oxford University Press, 88–115.

- Yablo, S. (2010). *Things: Papers on Objects, Events, and Properties*: Oxford University Press.
- Yablo, S. (2012). “Explanation, Extrapolation, and Existence.” *Mind* 121, no. 484 (October 1, 2012): 1007–1029.
- Ylikoski and Kuorikoski, (2010). “Dissecting explanatory power.” *Philos. Stud 148*(2), 201–219.