



INTERNATIONAL DOCTORAL
SCHOOL OF THE USC

Ismael
Ordóñez Miguéns

PhD Thesis

OBJECTIVITY WITH OBJECTS. A MODAL EXPLANATION OF THE INFINITE

Santiago de Compostela, 2024

Doctoral Programme in Logic and Philosophy of Science



Ph.D. THESIS

OBJECTIVITY WITH OBJECTS. A MODAL EXPLANATION OF THE INFINITE

Ismael Ordóñez Miguéns

Directora: Concha Martínez Vidal

Director: José Ferreirós Domínguez

Tutora: Concha Martínez Vidal

ESCOLA DE DOUTORAMENTO INTERNACIONAL DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO EN LÓXICA E FILOSOFÍA DA CIENCIA

SANTIAGO DE COMPOSTELA
2024



Declaración do autor da tese

D. Ismael Ordóñez Miguéns

Título da tese: Objectivity with Objects. A Modal Explanation of the Infinite

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

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

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En Santiago de Compostela, 30 de Decembro de 2023

Asdo. Ismael Ordóñez Miguéns

How the heart surges when it thinks of you,
Infinite One! How it sinks
when it gazes down upon itself!
Lamenting, it sees but misery, night and death.

Klopstock, *To the Infinite One*

AGRADECEMENTOS

O presente escrito é o resultado dun longo proceso. Durante estes anos varias persoas —en distintos momentos e de distintas maneiras— contribuíron a que sexa posible. Sen elas, non tería visto a luz. E a elas llo quero agradecer.

Quero comezar dándolle as grazas aos meus directores, Concha Martínez e José Ferreirós, polo seu seguemento e atención todo este tempo. Especialmente, pola confianza depositada en min. En segundo lugar, quero expresar o meu agradecemento a Øystein Linnebo, pola súa hospitalidade durante a miña estadía na Universidade de Oslo, por múltiples e ricas conversas e, sobre todo, pola súa inestimable axuda co último capítulo da tese. A José Pedro Úbeda quero agradecerlle por ter espertado en min o interese na teoría de conxuntos tantos anos atrás, e por seguir acompañándome no estudo desta complexa teoría.

Grazas, tamén, a Martín Pereira, polos seus bos consellos docentes, e pola axuda co soporte e presentación final da tese. Non é esaxerado dicir que, sen a súa axuda, este escrito non existiría. Tampouco podo esquecerme dos amigos atopados polo camiño, que fixeron máis levadeiro o que este conleva. Especialmente, grazas aos meus compañeiros de doutorado, Violeta e Alex, con quen compartín tantas conversas, momentos e experiencias enriquecedores en Santiago de Compostela.

Grazas a Oscar Parcerro, Xavier de Donato e Laura Lojo. O seu apoio e bos consellos foron imprescindibles para que esta tese chegue a bo porto. Grazas aos meus pais. E grazas a Cecilia, polo seu apoio e por acompañarme neste camiño.

Finalmente, quero darlle as grazas, tamén, ás institucións que fixeron economicamente posible a presente investigación. En primeiro lugar, ao Ministerio de Educación, Formación Profesional e Deportes, pola financiación a través do Programa Nacional FPU (número de axuda: FPU18/02787). E tamén por financiar a miña estadía na Universidade de Oslo baixo o programa de axudas para estadias no estranxeiro (número de axuda: EST19/00949). En segundo lugar, ao grupo de investigación Episteme, que me financiou baixo os seguintes proxectos de investigación: “Abstract Objects: For and Against”, financiado por FEDER, o Ministerio de Ciencia, Innovación e Universidades, e a Axencia Estatal de Investigación (número de proxecto: FFI2017-82534-P); e “Deflationary views in ontology and metaontology”, financiado pola Axencia Estatal de Investigación” (número de proxecto: PID2020-115482GB-I00).

Abstract

The objectivity of mathematics has been questioned for two reasons: Benacerraf's challenge and the emergence of several limitation theorems in mathematics. This has led to the search for new explanations of mathematical knowledge and alternatives to universalist platonism. Mainly, mathematical knowledge has been reduced to logical knowledge. Against this, this dissertation argues that mathematical knowledge cannot be reduced to logical knowledge. The two best proposals defending this reduction, fictionalism and multiversism, run into serious problems. Both will be challenged to explain how mathematical language acquires its content. Then, it will be argued that they do not satisfy this test. Faced with this problem, an alternative explanation of mathematical language and mathematical knowledge will be developed using abstraction principles. As a result, mathematical knowledge will be explained through linguistic competence. Finally, the proposal will be applied to ZF set theory.

Keywords objectivity, fictionalism, multiversism, abstraction principles, potentialism, set theory

Resumo

A obxectividade das matemáticas foi cuestionada por dúas razóns: o problema de Benacerraf e a emerxencia de numerosos teoremas de limitación en matemáticas. Isto levou á busca de novas explicacións do coñecemento matemático e de alternativas ao platonismo universalista. Principalmente, o coñecemento matemático tentouse reducir a coñecemento lóxico. En contra disto, esta tese argumenta que o coñecemento matemático non pode reducirse ao coñecemento lóxico. As dúas mellores propostas que defenden esta redución, o ficcionalismo e o multiversismo, incorren en serios problemas. As dúas serán desafiadas a explicar como a linguaxe matemática adquire o seu contido. Entón, argumentarase que ningunha satisfai esta proba. Ante este problema elabórase unha explicación alternativa da linguaxe e do coñecemento matemático usando principios de abstracción. Como resultado, a explicación do coñecemento matemático pasa pola explicación da competencia lingüística. Finalmente, a proposta aplicarase á teoría de conxuntos ZF.

Palabras clave obxectividade, ficcionalismo, multiversismo, principios de abstracción, potencialismo, teoría de conxuntos

Resumen

La objetividad matemática ha sido cuestionada por dos razones: el problema de Benacerraf y la emergencia de numerosos teoremas de limitación en matemáticas. Esto ha conducido a la búsqueda de nuevas explicaciones del conocimiento matemático y de alternativas al platonismo univiersista. Principalmente, se ha intentado reducir el conocimiento matemático a conocimiento lógico. En contra de esto, esta tesis argumenta que el conocimiento matemático no se puede reducir al conocimiento lógico. Las dos mejores propuestas a favor de esta reducción, el ficcionalismo y el multiversismo, incurren en serios problemas. Principalmente, las dos serán desafiadas a explicar como el lenguaje matemático adquiere su contenido. Entonces, se argumentará que ninguna pasa esta prueba. Ante este problema se elaborará una explicación alternativa del lenguaje y el conocimiento matemático usando principios de abstracción. Como resultado, la explicación del conocimiento matemático pasa por la explicación de la competencia lingüística. Finalmente, la propuesta se aplicará a la teoría de conjuntos ZF.

Palabras clave objetividad, ficcionalismo, multiversismo, principios de abstracción, potencialismo, teoría de conxuntos

Contents

Resumo	1
1 Introduction	9
1.1 Hypothesis	14
1.2 Objectives	14
1.3 Methodology	15
1.4 Structure of the thesis	16
2 Challenges to Mathematical Objectivity	19
2.1 Introduction	19
2.2 A Realist Conception of Knowledge	19
2.3 Mathematical Platonism	27
2.4 What Mathematical Platonism <i>is Not</i>	37
2.5 Varieties of Independence	46
2.6 Conclusion	50
Appendix 2.A.	51
Appendix 2.B.	54
Appendix 2.C.	57
3 Anti-realist Pluralism	
Fictionalism	59
3.1 Introduction	59
3.2 Fictionalism	59
3.3 The Reliability Challenge	63
3.4 Full Fictionalism and The Language Acquisition Challenge	74
3.5 Algebraic Fictionalism and the Consistency Challenge	89
3.6 Conclusion	95
4 Realist Pluralism	
Multiversism	97
4.1 Introduction	97
4.2 Knowledge in the Multiverse	97

4.3	MultiversimS	100
4.4	Motivating the Multiverse	103
4.5	The Set-Theoretic Multiverse	109
4.6	Full-Blooded Platonism	118
4.7	Conclusion	127
	Appendix 4.A.	127
5	Abstraction and Mathematical Knowledge	131
5.1	Introduction	131
5.2	Abstractionism	131
5.3	Metaontological Minimalism	138
5.4	Challenges to Abstraction	142
5.5	The Metasemantic Explanation of Mathematical Knowledge	159
5.6	Conclusion	171
	Appendix 5.A.	172
6	The Potential Infinite	175
6.1	Introduction	175
6.2	Iterated Abstraction	176
6.3	Potentialism vs Actualism	181
6.4	The Iterative Conception of Sets	183
6.5	Potentialist Set Theory	187
6.6	Reflection Principles	198
6.7	Conclusion	200
7	Completing the Hierarchy	
	Modal Abstraction Principles	203
7.1	Introduction	203
7.2	Varieties of Modal Abstraction	205
7.3	Refurbishing Abstraction	214
7.4	Infinity	218
7.5	Conclusion	221
	Appendix 7.A.	222
	Appendix 7.B.	222
	Appendix 7.C.	225
8	Conclusion	227
	Bibliography	231
	List of Figures	241

Resumo

A obxectividade das matemáticas foi cuestionada por dúas razóns principais: o desafío de Benacerraf e a aparición de diversos teoremas de limitación nas matemáticas contemporáneas. Comecemos polo primeiro. De acordo con unha caracterización sinxela e informal, as matemáticas son a ciencia dos números, funcións, conxuntos, figuras xeométricas e moitos outros obxectos que habitan o ámbito matemático. Todos estes obxectos teñen algo en común: son moi diferentes dos obxectos que nos rodean na nosa vida cotiá. Máis concretamente, o que os fai tan únicos é a súa natureza non espazo-temporal. Os obxectos matemáticos —existan ou non— son obxectos *abstractos*. Están “fóra” do espazo e do tempo. Por exemplo, a xente le libros, corta madeira ou sube montañas. Pola contra, ninguén viu o número 3 e ninguén o verá nunca. Do mesmo xeito, ninguén tocou nunca unha función —inda que os físicos usen funcións e vectores para describir o mundo. En xeral, os obxectos matemáticos non se poden percibir. Non podemos interactuar con eles. A situación é moi diferente á de galaxias distantes ou á dos obxectos demasiado pequenos para ser vistos. Non se trata, simplemente, de que estean causalmente illados de nós ao estar situados nun lugar remoto ou ao ser inaccesibles dalgún xeito. O que ocorre é que son causalmente ineficaces e, pola súa propia natureza, non exemplifican propiedades físicas.

O peculiar estatus metafísico dos obxectos matemáticos suscita algunhas sospeitas sobre a condición epistemolóxica das matemáticas. A primeira vista, estes obxectos son tan diferentes que poden parecer incognoscibles. Isto está en marcado contraste co coñecemento empírico. Temos boas teorías que explican como as persoas adquiren coñecemento sobre os feitos empíricos. As persoas interactúan con obxectos de tamaño medio en condicións perceptivas estables durante longos períodos. Observan o movemento dos corpos celestes, a caída libre de obxectos, o paso das estacións e moitos outros fenómenos. Elaboran teorías para explicar esta complexa rede de acontecementos. E, con moito esforzo, conseguen adquirir información sobre o mundo máis aló da proporcionada polas canles fisiolóxicas. Por suposto, as explicacións do coñecemento empírico son moito máis sofisticadas que estas oracións superficiais. Pero, existe un acordo xeneralizado en que polo menos temos algún coñecemento empírico do mundo. Sintonizar co mundo dun xeito epistemoloxicamente relevante non é doado. Tampouco é imposible. A maioría das explicacións do coñecemento empírico —como mínimo aquelas explicacións das crenzas empíricas xustificadas— sosteñen que as persoas están conectadas de forma causal ou contrafacticamente cos acontecementos empíricos. Son estas conexións tan fortes as que garanten a fiabilidade epistémica e a obxectividade da ciencia natural. A obxectividade é —en parte (23; 24; 25)— o coñecemento correcto do mundo.

Pola súa banda, as matemáticas resisten este tipo de explicacións. A natureza abstracta dos obxectos matemáticos fai difícil pensar que tipo de relación contrafáctica podería existir entre eles e nós. Son causalmente ineficaces. Polo tanto, descartanse as relacións causais. Non os podemos percibir. Ningún experimento nos proporcionará información nova sobre eles. Isto déixanos poucas alternativas. É moito máis difícil atopar unha canle que permita ás persoas sintonizarse co ámbito das matemáticas. Que tipo de relación contrafáctica podería existir entre os feitos matemáticos e as persoas que xustifique as nosas crenzas sobre eles e institúa aos matemáticos en profesionais fiables?

O reto de Benacerraf (7) utiliza a dificultade de dar resposta a esta pregunta para argumentar contra a obxectividade das matemáticas: se as matemáticas son unha ciencia de obxectos abstractos, entón non podemos explicar o coñecemento matemático e a fiabilidade matemática. Os obxectos abstractos serían epistemicamente imposibles de rastrexar. En concreto, o reto baséase en dous puntos. En primeiro lugar, conclúe que non existe unha conexión contrafáctica entre os feitos matemáticos e as persoas. Téñase en conta aquí que as relacións contrafáticas non son necesariamente causais. Para que exista unha relación contrafáctica entre dous tipos de sucesos só se precisa unha coordinación forte —en certo sentido a precisar— entre a variación dun e outro. Neste caso, a variación dos feitos matemáticos debería afectar ás respostas epistémicas dos matemáticos. Practicamente ninguén acepta que isto sexa posible. En efecto, se os feitos matemáticos son necesarios —como afirman moitos autores (74)—, entón non poderíamos dar sentido á posibilidade de variar estes mesmos feitos. Pola contra, necesitaríamos dar sentido a variacións imposibles —(123; 124; 45; 67). Pero, esta é unha idea controvertida e difícil de precisar. En segundo lugar, a maioría dos defensores do desafío (7; 40) afirman que se non hai unha relación contrafáctica, entón non se pode elaborar ningunha explicación adecuada da linguaxe matemática e do coñecemento matemático. Non poderíamos explicar como se fixa a referencia, como o vocabulario matemático adquire o seu significado ou como os matemáticos se converten en profesionais fiables. En consecuencia, dado que os matemáticos sí falan unha linguaxe significativa e nos proporcionan coñecemento matemático —polo menos crenzas matemáticas xustificadas—, debemos repensar a nosa visión das matemáticas. A diferenza das ciencias naturais, que describen un mundo empírico poboado por moitos tipos de obxectos, as matemáticas non describirían un mundo paralelo poboado por todos os habitantes mencionados anteriormente —números, conxuntos, vectores, etc. O coñecemento matemático e a obxectividade matemática non serían unha función do mundo matemático. Isto déixanos dúas alternativas. Ou a obxectividade matemática se consegue por outros medios —deixando de lado toda relación forte cos obxectos abstractos—, ou se abandona a obxectividade matemática.

Os obxectos matemáticos non son indispensables para explicar a obxectividade matemática. Neste punto podemos seguir o *dictum* de Kreisel: o que importa é a obxectividade, non os obxectos matemáticos. Agora ben, a mellor opción para salvar a obxectividade das matemáticas —á vez que facemos unha lectura literal dos enunciados da mesma, mantendo a súa imaxe de ciencia comprometida con obxectos— consiste en usar relacións lingüísticas e lóxicas máis débiles. Trátase de cambiar a perspectiva. En lugar de partir do mundo, partiríamos da lingua. A idea é impoñer restricións lóxicas e teóricas o suficientemente fortes para seleccionar unha teoría matemática, fixar a referencia dos termos e adquirir unha linguaxe significativa. Deste xeito, poderíamos explicar o coñecemento matemático e a fiabilidade dos matemáticos sen utilizar relacións contrafáticas fortes. Por exemplo, estas restricións incluírían a coherencia

da teoría, que esta sexa o suficientemente potente como para incluír razoamentos matemáticos básicos —por exemplo, a aritmética— ou que respecte as metrizacións utilizadas nas ciencias naturais. Estas restricións cumprirían unha dobre función. Por un lado, seleccionarían unha colección de oracións: os axiomas da teoría matemática desexada, é dicir, a teoría matemática obxectivamente correcta. Por outra banda, seleccionarían unha colección de obxectos: os obxectos que satisfagan as propiedades establecidas na teoría. Unha vez fixada a referencia neste dominio de obxectos, os axiomas da teoría resultan ser verdadeiros. Como resultado, o ideal leibniziano podería finalmente facerse realidade: o coñecemento matemático reduciríase a consecuencia dentro desta teoría ideal. De feito, esta imaxe encaixa ben coa *natureza alxébrica* das matemáticas contemporáneas (56; 35; 76). De acordo con esta, os matemáticos seleccionan un conxunto de axiomas —estes caracterizan os conceptos que desexan estudar. Da mesma forma na que un sistema de ecuacións con varias incógnitas selecciona as súas solucións, os axiomas seleccionan os obxectos que os fan verdadeiros. Entón, a tarefa dos matemáticos consiste en estudar as consecuencias destes axiomas.

Aquí é onde irrompe o segundo dos problemas comentados para obxectividade matemática: a aparición de diversos teoremas de limitación na matemática contemporánea. Como o seu nome indica, estes teoremas mostran as limitacións dedutivas das teorías matemáticas, así como as limitacións expresivas das correspondentes linguaxes. Os primeiros exemplos destes teoremas pódense atopar no século XIX, cando Bolyai e Lobachevsky demostraron a independencia lóxica do quinto postulado con respecto aos demais axiomas da xeometría euclidiana (133, Ch. 6-7). Porén, os resultados de independencia máis famosos son os teoremas de incompletitude de Gödel. Este probou que calquera teoría recursivamente axiomatizable —é dicir, unha teoría con un conxunto recursivo de axiomas— que inclúa á aritmética de Peano-Dedekind e sexa consistente, é incompleta: hai unha oración tal que a teoría non proba nin esta oración nin a súa negación. Pero, cal é a relevancia dos resultados de independencia para a obxectividade matemática? Lembremos que estamos tratando de recuperar a obxectividade matemática escollendo unha colección de restricións lóxicas e teóricas que seleccionen a teoría desexada: a teoría matemática obxectivamente correcta. Vexamos en detalle que papel xogan estas restricións. En primeiro lugar, deben seleccionar unha teoría que, en principio, sexa manexable para as persoas. Isto significa que a teoría debe ser recursivamente axiomatizable. Do contrario, non poderíamos saber que oracións forman parte dela. A xente non sería capaz de recoñecer a teoría. En segundo lugar, a teoría debe recuperar a matemática suficiente. Polo menos, debe recuperar a aritmética de Peano-Dedekind —esta é unha teoría mínima incluída na maioría dos razoamentos matemáticos. En terceiro lugar, a teoría debe ser consistente; aquí estámonos ocupando das matemáticas clásicas. En consecuencia, os teoremas de incompletitude dinnos que a teoría resultante será incompleta. Calquera teoría seleccionada por un conxunto de restricións lóxicas e teóricas do tipo anteriormente mencionado será incompleta. Polo tanto, o coñecemento matemático e a obxectividade matemática non poden reducirse á implicación lóxica dentro dunha teoría *ideal*. De feito, sabemos que a maioría das mellores teorías matemáticas, como a teoría de conxuntos de Zermelo-Fraenkel, son altamente incompletas: hai infinitas oracións que estas teorías, así como outras extensións naturais das mesmas, non deciden.

Os resultados de independencia son un serio problema para a obxectividade matemática. Porén, alguén podería argumentar que non todo está perdido. Este tipo de resultados simplemente mostran que

o noso coñecemento matemático está lonxe de ser perfecto. Hai moitas preguntas matemáticas para as que non temos resposta. Quizais nunca a teremos. Pero, isto non significa que a xente non teña coñecemento matemático ningún. De feito, este tipo de resultados son de esperar. Ningunha parcela de coñecemento humano é perfecta. As ciencias naturais tamén albergan un gran número de preguntas sen resposta. Que había antes do Big Bang? Como se orixinou a vida na terra? O universo é finito ou infinito? A falta dunha resposta a todas estas preguntas non nos leva a concluír que a ciencia natural non é obxectiva, polo menos ata o momento. Do mesmo xeito, se unha colección de restricións lóxicas e teóricas é suficiente para seleccionar unha teoría matemática —aínda que sexa unha teoría incompleta— e fixar a referencia dos termos da linguaxe, isto sería suficiente para explicar a obxectividade do coñecemento limitado que teñen os matemáticos. Estas restricións seleccionan a interpretación da linguaxe matemática. Á súa vez, a interpretación selecciona unha teoría completa: a teoría matemática obxectivamente correcta. Deste xeito, a teoría decide todas as cuestións matemáticas. E, isto é así inda que haxa oracións que os matemáticos non saben se pertencen ou non á teoría. Esta resposta aos resultados da independencia é sinxela e eficaz.

Porén, hai un segundo tipo de teorema de limitación que afecta ao poder expresivo da linguaxe matemática e que a invalida: ningunha colección de restricións lóxicas e teóricas é suficiente para fixar a referencia matemática (128; 111). Aquelas teorías de primeira orde que son o suficientemente poderosas como para recuperar a aritmética de Peano-Dedekind pódense interpretar en dominios de obxectos que difiren estruturalmente entre si; as teorías matemáticas de primeira orde non son categóricas. Por outro lado, inda que a lóxica de segunda orde é o suficientemente forte como para facer o traballo, o seu status é moi controvertido (114; 54; 36). En consecuencia, non poderemos illar unha interpretación para a linguaxe matemática enumerando soamente unha colección de restricións lóxicas e teóricas: estas seleccionarán un conxunto de oracións da linguaxe que á súa vez terá múltiples interpretacións. Pero, se non podemos seleccionar unha interpretación privilexiada deste xeito, ningún recurso lóxico seleccionará unha teoría matemática completa. Así, só nos queda un conxunto incompleto de axiomas. E todas as teorías que extendan estes axiomas tamén serán incompletos. Se as matemáticas son unha ciencia de obxectos abstractos, o coñecemento matemático e a obxectividade matemática tampouco son función dos nosos desiderata teóricos.

O desafío de Benacerraf cuestiona o uso de relacións contrafácticas fortes para explicar o coñecemento e a obxectividade matemática. Así mesmo, os teoremas de limitación utilizáronse para cuestionar o uso de relacións máis débiles baseadas na imposición de restricións lóxicas e teóricas sobre as nosas teorías. Estes tampouco son suficientes para seleccionar unha teoría matemática obxectivamente correcta. Así, ambos tipos de argumentos foron usados para cuestionar a caracterización informal e amable mencionada máis arriba: a caracterización das matemáticas coma unha ciencia de obxectos abstractos independentes de nosoutros. Se un universo de obxectos abstractos existe independentemente de nós, entón é moi plausible que non exista unha relación semántica ou epistémica entre ese universo e as persoas. Apoiándose nesta liña de razoamento, ambos argumentos foron empregados para cuestionar o platonismo matemático, a tese de que os obxectos matemáticos existen independentemente de nós e que os matemáticos se dedican a estudalos. En consecuencia, atopámonos de novo ante unha versión reforzada do dilema anterior: ou repensamos a nosa visión das matemáticas —é dicir, o platonismo—, ou renunciámos á obxectividade

matemática.

Con respecto a este dilema, sitúame entre Scila e Caribdis. Por unha parte, este texto é unha defensa dunha tese central e negativa:

O coñecemento matemático non se reduce a coñecemento lóxico.

Consecuentemente, a obxectividade matemática non se reduce a obxectividade lóxica. En contra das novas formas de platonismo pluralista e tamén en contra do fictionalismo, nego que o coñecemento matemático se poida explicar exhaustivamente apelando meramente ao coñecemento da consistencia e da consecuencia lóxica. Asumindo unha concepción eminentemente semántica da linguaxe, argumentarei que o coñecemento lóxico presupón a competencia lingüística. Polo tanto, calquera redución do coñecemento matemático ao coñecemento lóxico presupón dúas cousas: primeiro, que a linguaxe matemática é significativa, e segundo, que os matemáticos son capaces de comprendela. Porén, argumentarei que nin o multiversismo nin o fictionalismo son capaces de explicar isto. É dicir, estas posicións non son capaces de proporcionar unha explicación metasemántica de como o vocabulario matemático adquire o seu significado, e de como os falantes adquiren a súa competencia lingüística. Así, opoñerei ás filosofías pluralistas das matemáticas unha nova versión do Desafío da Adquisición da Linguaxe usado (30) por en contra das matemáticas clásicas —porén, esta nova versión non está directamente relacionada co argumento de Dummett, nin coas súas preferencias intuicionistas.

Por outra parte, o texto é unha defensa dunha tese positiva complementaria:

O coñecemento matemático pode ser explicado mediante a competencia lingüística.

Así, tratarei de complementar as concepcións pluralistas das matemáticas desenvolvendo unha explicación adecuada de como o vocabulario matemático adquire o seu contido, e de como os matemáticos adquiren a súa competencia lingüística e matemática. Para isto, usarei a recente aproximación abstraccionista aos obxectos abstractos deseñada por Linnebo (81; 82). O seu programa fai un forte uso dos principios de abstracción para explicar a referencia e o coñecemento matemático. A relevancia filosófica dos principios de abstracción foi orixinalmente sinalada por Frege, na súa *Grundlagen der Arithmetik*. Estes principios introducen unha perspectiva reducionista en metasemántica que os fai especialmente apropiados para elaborar unha explicación epistemolóxica dos obxectos abstractos: podemos explicar como os falantes adquiren unha linguaxe enteira comprometida con estes obxectos sin presuporner pola súa parte ningún coñecemento dos mesmos. Ademais, a adquisición dunha linguaxe mediante o uso de principios de abstracción proporciona aos falantes un coñecemento non lingüístico. É dicir, a competencia lingüística ven acompañada do coñecemento dun punhado de verdades formuladas nesa linguaxe —non non soamente de verdades lóxicas. Desta forma, o logro cognitivo en matemáticas recibe unha explicación metasemántica seguindo as liñas de (79; 82) e (115; 116). Isto garantiza un mínimo de verdade matemática obxectiva sen abandonar os obxectos matemáticos. Neste sentido, a miña posición é unha modesta vindicación do platonismo matemático —a pesar das aparencias, inda hai un lugar para a combinación de ambos.

No capítulo 2 analizo a explicación máis intuitiva da obxectividade matemática: o platonismo matemático. En primeiro lugar, caracterizo a obxectividade do coñecemento empírico como a conxunción

de dúas teses: Independencia e Reactividade. A primeira afirma que o contido das crenzas, xuízos ou afirmacións é independente das respostas epistémicas dos axentes. A segunda afirma que o coñecemento obxectivo implica unha relación forte e contrafáctica entre os axentes e o mundo. En segundo lugar, caracterizo o platonismo matemático como a tese filosófica de que os obxectos abstractos existen independentemente de nós. De acordo co platonismo, as matemáticas son a ciencia que estuda eses obxectos. En terceiro lugar, explico en que consiste o desafío de Benacerraf e argumento que este funciona só cando se acepta que o coñecemento matemático satisfai Reactividade. Finalmente, explico por que o platonismo non involucra nin unha teoría sobre a estrutura do mundo ou sobre a verdade, e desenvolvo a idea de que os obxectos matemáticos son independentes de nos.

No **capítulo 3** analizo o ficcionalismo, a tese filosófica de que as teorías matemáticas son ficcións, e polo tanto son falsas. O ficcionalismo reduce o coñecemento matemático ao coñecemento lóxico: coñecemento sobre a consistencia e as consecuencias desas ficcións. Neste capítulo desafío o ficcionalismo por tres razóns. En primeiro lugar, argumento que baixo dous supostos razoables non pode explicar a fiabilidade matemática. Entón, intentarei reforzar esta obxección, argumentando que o ficcionalismo non pode explicar a fiabilidade dos matemáticos inda se os devanditas supostos non son aceptados. En segundo lugar, distingo dous tipos de ficcionalismo: o *ficcionalismo pleno* e o *ficcionalismo alxébrico*. O primeiro afirma que, a pesar de que as teorías matemáticas son falsas, a linguaxe matemática é significativa e trata sobre obxectos abstractos. Pola súa parte, o segundo non atribúe ningún contido á linguaxe matemática. Usando esta distinción, defendo que se o ficcionalismo pleno pode explicar como o vocabulario matemático adquire o seu contido e como os matemáticos adquiren a súa competencia lingüística, entón o platonismo tamén pode. Pero, dado que o mellor argumento para o ficcionalismo consiste no fracaso do platonismo, isto é fatal para a primeira posición filosófica. Finalmente, defendo que o ficcionalismo alxébrico non está nunha posición máis vantaxosa. A redución do coñecemento matemático ao coñecemento lóxico presupón a competencia lingüística dos matemáticos. O ficcionalismo alxébrico non pode explicar isto a menos que, á súa vez, sexa reducido ao ficcionalismo pleno. En conclusión, o ficcionalismo debe ser complementado con unha explicación alternativa da competencia lingüística dos matemáticos.

No **capítulo 4** analizo o multiversismo. O multiversismo é unha nova versión pluralista do platonismo matemático que defende a existencia de múltiples universos matemáticos. En lugar dun universo único e privilegiado habería un multiverso. Neste, todos os universos son equivalentes dende o punto de vista da semántica e da metafísica. Como o ficcionalismo, o multiversismo reduce o coñecemento matemático ao coñecemento lóxico. En primeiro lugar, explico que hai moitos tipos diferentes de multiversismo. En segundo lugar, explico os dous principais argumentos a favor do multiversismo: o argumento dende a independencia lóxica e o argumento dende a categoricidade. En terceiro lugar, argumento que as dúas versións máis radicais do multiversismo non son boas explicacións das matemáticas. Ningunha delas impón límites á extensión do multiverso. Por este motivo ambas poden reducir o coñecemento matemático ao coñecemento lóxico. O multiversismo hamkinsiano non pode facer fronte ao desafío de explicar como o vocabulario matemático adquire o seu significado. Máis concretamente, a súa explicación da referencia matemática é circular. Á súa vez, o platonismo *full-blooded* carece de unha noción adecuada de consistencia para explicar o coñecemento matemático. O uso desta noción presupón unha explicación

da competencia lingüística dos falantes. O problema é que o platonismo *full-blooded* tampouco pode proporcionar a devandita explicación.

No **capítulo 5** utilizo o enfoque abstraccionista dos obxectos matemáticos desenvolvido por Linnebo (81; 82) para explicar a competencia lingüística dos matemáticos e o coñecemento matemático. En primeiro lugar, explico que son os principios de abstracción. Estes introducen un enfoque metasemántico reduccionista permitindo explicar como unha comunidade de falantes é capaz de adquirir a linguaxe matemática sen presuporner ningunha relación contrafáctica ou lóxica entre os seus membros e os feitos matemáticos. En segundo lugar, explico en que consisten os dous principais problemas para as perspectivas abstraccionistas: o problema do Xulio César e o problema da mala compañía. Despois, mostro como estes problemas poden ser resolto. Unha vez que se demostre que os principios de abstracción están en boas condicións, mostrarei como tamén poden ser empregados para explicar o coñecemento matemático. O proceso de abstracción mediante o cal os falantes estenden a súa linguaxe tamén lles proporciona coñecemento sobre unha colección de verdades matemáticas—e non soamente verdades lóxicas. Así, tal e como Linnebo (79; 82) e Rayo (115; 116) afirman, o coñecemento matemático ven da man da comprensión lingüística. Fronte á perspectiva pluralista do ficcionalismo e do multiversismo, o enfoque abstraccionista permítenos reintroducir un mínimo de obxectividade matemática.

No **capítulo 6** explico como os principios de abstracción son suficientes para recuperar a teoría de conxuntos de Zermelo-Fraenkel (81, Ch 3,12). En primeiro lugar, explico como a iteración dos intentos de abstracción xera unha serie de expansións do dominio. Así mesmo, explico como se poden expresar as expansións do dominio a nivel da linguaxe de obxectos usando operadores modais. En segundo lugar, contrapoño as concepcións potencialista e actualista da teoría de conxuntos. O actualismo afirma que o universo conxuntista forma unha pluralidade de obxectos. Á súa vez, o potencialismo nega isto. Entón, argumento que o segundo está en mellores condicións que o primeiro. En terceiro lugar, introduzo a concepción iterativa dos conxuntos. Unha vez feito isto, explico como a aplicación repetida dos principios de abstracción proporciona unha explicación das súas operacións primitivas: a operación ‘conxunto de’ e o conxunto potencia. Esta explicación implica o potencialismo en teoría de conxuntos. Deste xeito, expoño, por unha parte, os principios modais que describen o proceso iterativo e, por outra, aqueles principios que caracterizan a natureza específica dos conxuntos. Combinados, todos estes principios forman a teoría de conxuntos potencialista. Finalmente, mostro como esta teoría interpreta a teoría de conxuntos de Zermelo-Fraenkel sen os axiomas de Infinitude e Reemprazo. Para recuperar ambos axiomas, a teoría de conxuntos potencialista debe complementarse con principios adicionais máis fortes.

Finalmente, no **capítulo 7** desenvolvo un novo tipo de principios de abstracción modal. A través deles, explico como os falantes adquiren a capacidade de conceptualizar o infinito en acto mediante un proceso de abstracción. O principal beneficio da abstracción modal é que se trata dunha forma de abstracción carente de especificacións. É dicir, o proceso de abstracción non ten lugar no mundo presente, sobre un dominio de obxectos relacionados mediante a relación de unidade. En cambio, ten lugar sobre o que é posible relativamente a esta relación. Así, a mera posibilidade de que dous obxectos estean relacionados por ela é suficiente para que a abstracción introduza o correspondente obxecto abstracto especificado por eles. En primeiro lugar, descarto algunhas formas insatisfactorias de enunciar estes principios. Despois, estando a linguaxe da teoría de conxuntos potencialista, introducindo unha serie de operadores modais.

Unha vez feito isto, os principios de abstracción modal pódense formalizar apropiadamente. Finalmente, mostro como os principios de abstracción modal interpretan o axioma do Infinito cando se engade á teoría de conxuntos potencialista. Deste xeito, o abstraccionismo pode explicar como os falantes captan a noción de infinito.

O obxectivo principal desta investigación é o de recuperar un mínimo de obxectividade non lóxica para o coñecemento matemático ao mesmo tempo que se defende a súa condición de ciencia de obxectos abstractos, independentes de nós. O coñecemento matemático implica unha forma moi particular de logro cognitivo: a correcta comprensión da linguaxe matemática. As mellores explicacións pluralistas das matemáticas, o ficcionalismo e o pluralismo radical, non son capaces de proporcionar unha explicación satisfactoria da competencia lingüística dos matemáticos. Fronte a estas, a defensa feita do programa abstraccionista pode entenderse coma unha defensa do platonismo universista. Porén, isto non debe ser entendido coma unha refutación do ficcionalismo ou do multiversismo. Unha vez introducido un mínimo de obxectividade matemática, estas concepcións pódense defender para outras teorías matemáticas adicionais. O debate entre as dúas posicións filosóficas mencionadas está fóra do alcance do presente texto e, de ser posible a súa solución, terá que ser resolto mediante o uso de outros argumentos filosóficos. A introducción dun mínimo de obxectividade matemática só constitúe o punto de partida na exploración da multiplicidade de ficcións ou universos alternativos.

1 Introduction

The objectivity of mathematics has been called into question because of two main reasons: Benacerraf's challenge and limitation theorems. Let's start with the former. According to a simple and informal characterization, mathematics is the science of numbers, functions, sets, geometrical figures and many other objects inhabiting the mathematical realm. All these objects have something in common: they are very different from those objects that surround us in our daily lives. More specifically, what makes them so unique is their non-spatiotemporal nature. Mathematical objects —whether they exist or not— are *abstract* objects. They are “out” of space and time. For example, people read books, people chop wood, or they climb mountains. On the contrary, nobody has seen the number 3 —and nobody will see it ever. Similarly, nobody has touched ever a function —even if real-valued functions and vectors are used by physicists to describe the world. In general, mathematical objects cannot be perceived. We cannot interact with them. The situation is very different from that of distant galaxies or objects too small to be seen. It is not simply that they are causally isolated from us because they are remote or inaccessible. They are causally ineffective and, by their very nature, do not exemplify physical properties.

The peculiar metaphysical status of mathematical objects raises some suspicions about the epistemological condition of mathematics. At first glance, they are so different that they might seem to be unknowable. This is in stark contrast to empirical knowledge. We have good theories that explain how people acquire knowledge about empirical matters. People interact with medium-sized objects under stable perceptual conditions for long periods. They observe the movement of celestial bodies, the free fall of objects, the passing of seasons, and many other phenomena. They elaborate theories to explain this complex web of events. And, with a lot of effort, people manage to acquire information about the world beyond that provided by physiological channels. Of course, explanations of empirical knowledge are much more sophisticated than these superficial phrases. But the point is that there is agreement that we at least have some empirical knowledge of the world. Tuning into the world in an epistemologically relevant way is not easy. Nor is it impossible. Most explanations of empirical knowledge —at least of empirically justified belief— contend that people are causally or, generally, counterfactually connected to empirical events. And, these strong connections guarantee their epistemic reliability and the objectivity of natural science. Objectivity is —in part (23; 24; 25)— the correct knowledge of the world.

In turn, mathematics resists these explanations. The abstract nature of mathematical objects makes it difficult to think what kind of counterfactual relation could exist between them and us. They are causally ineffective. Therefore, causal relations are ruled out. We cannot perceive them. No experiment will provide us with new information about them. This leaves us with few alternatives. It is much more difficult

to find a channel that allows people to tune in to the mathematical realm. What kind of counterfactual relation could exist between mathematical facts and people that justifies our beliefs about them and makes mathematicians reliable professionals? Benacerraf's challenge (7) uses the difficulty of providing an answer to this question against the objectivity of mathematics: if mathematics is a science of abstract objects, then we cannot explain mathematical knowledge and mathematical reliability. Abstract objects would be epistemically untraceable. Specifically, the challenge is based on two points. First, it concludes that there is no counterfactual connection between mathematical facts and people. Note that counterfactual relations are not necessarily causal. All that is needed for there to be a counterfactual relation between two types of events is a strong—in a sense to be specified—coordination between the variation of one and the other. In this case, the variation of the mathematical facts should affect the epistemic responses of the mathematicians. Virtually no one accepts that this is possible. Indeed, if mathematical facts are necessary—as many authors claim (74)—, then we could not make sense of *possible variations* of these facts. Instead, we would need to make sense of *impossible variations*—(123; 124; 45; 67). But, this is a controversial and difficult idea to pin down. Second, most advocates of the challenge (7; 40) contend that if there is no counterfactual relation, no suitable explanation of mathematical language and mathematical knowledge can be elaborated. We could not explain how reference is fixed, how mathematical vocabulary acquires its meaning, and how mathematicians become reliable professionals. Consequently, since mathematicians speak a meaningful language and provide us with mathematical knowledge—at least with justified mathematical beliefs—we must rethink our view of mathematics. Unlike the natural sciences, which describe an empirical world populated by many kinds of objects, mathematics would not describe a parallel world populated by all the aforementioned inhabitants—numbers, sets, vectors, etc. Mathematical knowledge and mathematical objectivity would not be a function of the mathematical world. This leaves us with two alternatives. Either mathematical objectivity is achieved by other means—leaving strong relations to abstract objects aside—, or mathematical objectivity is abandoned.

Mathematical objects are not indispensable for mathematical objectivity. At this point, we can follow Kreisel's *dictum*: what matters is mathematical objectivity, not mathematical objects. Now, the best option to save the objectivity of mathematics—while reading it at face value, and keeping the image of a science of objects—is to resort to weaker linguistic and logical relations. It is a matter of turning the question around. Instead of starting from the world, we would start from language. The idea is to impose enough logical and theoretical constraints to select a mathematical theory, fix mathematical reference, and acquire a meaningful language. In this way, we could explain mathematical knowledge and the reliability of mathematicians without using strong, counterfactual relations. For example, these constraints would include the consistency of the theory, that it be powerful enough to include basic mathematical reasoning—such as arithmetic—, or that it respects the metrizations used in the natural sciences. These constraints would serve a dual function. On one side, they would select a collection of sentences: the axioms of the desired mathematical theory—the objectively correct mathematical theory. On the other side, they would select a collection of objects: the objects that satisfy the properties coached in the theory. Once reference is fixed in this domain of objects, the axioms of the theory turn out to be true. As a result, the Leibnizian ideal could finally become reality: mathematical knowledge reduces to consequence inside this ideal theory. Indeed, this picture fits well with the *algebraic nature* of contemporary mathematics (56; 35; 76).

Mathematicians select a set of axioms. These characterize the concepts they wish to study. As a system of equations with several unknowns selects its solutions, these axioms would select the objects that make them true. Then, mathematicians study what follows from these axioms.

Here is where the second problem for mathematical objectivity bursts in: limitation theorems. As their name suggests, these theorems show the deductive limitations of mathematical theories as well as the expressive limitations of mathematical languages. We find the first examples of these theorems in the 19th century when Bolyai and Lobachevsky proved the logical independence of the fifth postulate with respect to the other axioms of Euclidean geometry (133, Ch. 6-7). However, the most famous independence results are Gödel's incompleteness theorems. He proved that any recursively axiomatizable theory —i.e. a theory whose set of axioms is recursive— that includes Peano-Dedekind arithmetic and is consistent, is incomplete. There is a sentence such that the theory proves neither it nor its negation. But, what is the relevance of independence results for mathematical objectivity? Remember, we were trying to recover mathematical objectivity by laying down a collection of logical and theoretical constraints that select our desired theory —the objectively correct mathematical theory. Let's see in detail what role these restrictions play. First, they must select a theory that is in principle manageable for people. This means a recursively axiomatizable theory. On the contrary, we would not be able to know which sentences are part of it! People would not be able to recognise the theory. Second, the theory must recover enough mathematics. At least, it must recover Peano-Dedekind arithmetic —this is a minimal theory involved in most mathematical reasonings. Third, the theory must be consistent —we are dealing with classical mathematics. Consequently, such a theory will be incomplete. Any theory selected by a set of logical and theoretical constraints of the above-mentioned kind will be incomplete. Consequently, mathematical knowledge and mathematical objectivity do not reduce to logical entailment inside an *ideal* theory. Indeed, we know that most of our best mathematical theories, such Zermelo-Fraenkel set theory, are highly incomplete: there are infinitely many sentences that these theories or natural extensions do not decide.

Independence results are a serious drawback for mathematical objectivity. However, somebody could argue that not everything is lost. This kind of results simply show that our mathematical knowledge is far from being perfect. There are many mathematical questions for which we have no answer. Perhaps we never will. But this does not mean that people do not have mathematical knowledge. Indeed, these sort of results should be expected. No parcel of human knowledge is complete. Natural science also harbours a large number of unanswered questions. What was there before the Big Bang? How did life on earth originate? Is the universe finite or infinite? But, the lack of an answer to all these questions does not lead us to conclude that natural science is not objective —at least up to some point. Therefore, if a collection of logical and theoretical constraints selects a mathematical theory —even if an incomplete theory— and fixes mathematical reference, this would be enough to explain the objectivity of mathematicians' limited knowledge. These constraints select the interpretation of the mathematical language. In turn, the interpretation selects a complete theory —the objectively correct mathematical theory. As a result, the theory decides all mathematical questions. And, this is so even if there are sentences that mathematicians do not know whether or not they belong to the theory. This response to independence results is simple and effective.

Nevertheless, there is a second kind of limitation theorem affecting the expressive power of mathematical language that tells against it: no uncontroversial collection of logical and theoretical constraints is enough to fix mathematical reference (128; 111). Those first-order theories which are powerful enough to recover Peano-Dedekind arithmetic can be interpreted in domains of objects which differ structurally from each other—first-order mathematical theories are not categorical. In turn, while second-order logic is strong enough to do the job, this is highly controversial (114; 54; 36). Consequently, we will not be able to isolate one of these interpretations just by listing a collection of logical and theoretical constraints on theories: these will only select a set of sentences from the language, which in turn will have multiple interpretations. But, if we cannot select a privileged interpretation in this way, nothing else will select a complete mathematical theory. We are just left with an incomplete set of axioms. And all theories extending them will be on a par. If mathematics is a science of abstract objects, mathematical knowledge and mathematical objectivity are not a function of our theoretical desiderata either.

Benacerraf's challenge questions the use of strong, counterfactual relations to explain mathematical knowledge and mathematical objectivity. In turn, limitation theorems have been used to question the use of weaker relations based on the imposition of logical and theoretical constraints on our theories. These are not enough to select an objectively correct mathematical theory either. Thus, both kinds of arguments have been used to question the informal and amenable characterization of mathematics as a science of abstract objects independent from us. If a universe of abstract objects exists independently of us, then it is plausible that there is no semantic or epistemic relation between it and people. In this way, both arguments have been used to question *mathematical platonism*, the thesis that mathematical, abstract objects exist independently of us and that mathematicians aim to study them. Consequently, we are faced again with a tightened version of the above dilemma: either we rethink our view of mathematics—i.e., platonism—or, or we give up mathematical objectivity.

All philosophical accounts of mathematics that have been proposed so far have embraced at least one horn of this dilemma. Those philosophies of mathematics that seek to preserve a minimal notion of mathematical objectivity—i.e., mathematical objective truth and knowledge—have chosen the first option, rethinking what mathematics is about. Some of them deny that mathematics is a science of objects. Instead, mathematicians would be involved with complex, modal facts (106; 55; 8) or with structures (119; 125; 56; 65). Others, while accepting that mathematics is a science of objects, have reconsidered their metaphysical nature. For example, some authors understand mathematical objects as symbolic objects, explaining mathematical objectivity in terms of intersubjectivity within a community (34; 35; 37; 38). Alternatively, mathematical platonism has not disappeared. Nevertheless, it has changed its form, becoming a plural metaphysical thesis (2; 54). Faced with the difficulties of justifying the existence of a single correct theory of mathematical objects—the objectively correct mathematical theory—its advocates have accepted that there are a multiplicity of correct theories. All of them are equivalent from the point of view of semantics and metaphysics. Instead of a mathematical universe, platonism contends the existence of a mathematical *multiverse*. This metamorphosis of platonism comes with a new understanding of mathematical knowledge. The explanation of mathematical knowledge no longer needs a relation between mathematical facts and speakers. On the contrary, this reduces to logical knowledge—knowledge of consistency and entailment—and mathematical objectivity reduces to logical objectiv-

ity. Finally, though not exhaustively, other accounts of mathematics embrace both horns of the dilemma. This is the case of fictionalism (76; 16; 5). Fictionalism accepts that mathematical theories are about mathematical, abstract objects. However, it denies that such objects exist. According to it, mathematical theories are fiction. And mathematics, rather than a science of abstract objects, is the study of what happens in those fictions. Since all fictions are semantically and metaphysically on a par, fictionalists also deny the existence of *the* objectively correct mathematical fiction. The single privilege of mathematical theories is to be studied historically by mathematicians. In this way, fictionalists, as the new platonists, reduce mathematical knowledge to logical knowledge —knowledge about consistency and entailment in mathematical theories.

In this dilemma, I will place myself between Scylla and Charybdis. On one side, this text is a defence of a central, and negative thesis:

Mathematical knowledge does not reduce to logical knowledge.

Consequently, mathematical objectivity does not reduce to logical objectivity. Against the new pluralist versions of platonism as well as against fictionalism, I deny that mathematical knowledge can be explained exclusively by appealing to knowledge of consistency and logical consequence. Assuming an eminently semantic conception of language, I will argue that logical knowledge presupposes linguistic competence. Therefore, any reduction of mathematical knowledge to logical knowledge presupposes two things: first, that mathematical language is meaningful, and second, that mathematicians understand it. Nevertheless, I will argue that neither multiversism, nor fictionalism can explain this. That is, they cannot provide a metasemantic explanation of how mathematical vocabulary acquires its meaning, and how speakers acquire their linguistic competence. Thus, I will oppose to pluralist philosophies of mathematics a new version of the Language Acquisition Challenge used by (30) against classical mathematics —however, this new version is not directly related to Dummett’s argument, nor with his intuitionist preferences.

On the other side, the text is a defence of a complementary, positive thesis:

Mathematical knowledge can be explained through linguistic competence.

Thus, I will try to complement those pluralist accounts of mathematics by developing an adequate explanation of how mathematical vocabulary acquires its content, and how mathematicians acquire their linguistic and mathematical competence. For this, I will use Linnebo’s (81; 82) recent abstractionist approach to abstract objects. The program makes strong use of abstraction principles to account for mathematical reference and mathematical knowledge. The philosophical relevance of abstraction principles was originally brought out by Frege, in his *Grundlagen der Arithmetik*. These introduce a reductionist perspective in metasemantics that makes them especially suited to elaborate an epistemological account of abstract objects: we can explain how speakers acquire a whole language committed to abstract objects without presupposing that they have any prior knowledge of these objects. Moreover, the acquisition of a language through abstraction provides speakers with some non-linguistic knowledge. That is, linguistic competence comes with knowledge of a handful of truths formulated in that language —and not just logical truths. As a result, cognitive achievement in mathematics receives a metasemantic explanation

following the lines of (79; 82) and (115; 116). This guarantees a minimum of objective mathematical truth and knowledge without abandoning mathematical objects. In this sense, my position is a modest vindication of objectivity and mathematical platonism—despite appearances, there is still a place for the combination of both.

However, this must not be understood as a refutation of fictionalism or multiversism. Once a minimum of mathematical objectivity has been introduced, these conceptions can be defended for other mathematical theories. The debate between them is beyond the scope of the present text, and—if possible—will have to be solved by other philosophical arguments. The minimum of mathematical objectivity introduced here only constitutes the starting point in the exploration of the multiplicity of alternative mathematical fictions or universes.

1.1 HYPOTHESIS

The reduction of mathematical knowledge to logical knowledge proposed by fictionalism and the most extreme forms of multiversism entails a plural conception of mathematics and the abandonment of a proper notion of mathematical objectivity: all consistent mathematical theories are correct. However, this fails as an explanation of mathematical knowledge. Logical knowledge—i.e. knowledge of consistency and logical entailment—presupposes the linguistic competence of speakers. But, neither fictionalism nor multiversism can explain how mathematical vocabulary acquires its content, and how speakers understand it. An alternative explanation is required. This explanation can be developed using abstraction principles. In addition, the use of abstraction principles will show how mathematical knowledge is acquired to some extent at the same time as linguistic competence is acquired.

1.2 OBJECTIVES

The main objective of this research is to explain how a minimum of non-logical objectivity can be recovered for mathematics. In particular, the recent abstractions approach of (81; 82) will be used to show how cognitive achievement in mathematics can be explained as a form of linguistic knowledge. The achievement of this objective will be articulated through the following sub-objectives:

1. To elaborate an analysis of the most intuitive explanation of mathematical objectivity: mathematical platonism. Also, identify the main characteristics of objectivity in the natural sciences and show that the main argument against platonism, Benacerraf's challenge, is biased by attributing them to mathematical knowledge.
2. To analyse the explanation of mathematical knowledge proposed by fictionalism. The analysis will show how the reduction of mathematical knowledge to logical knowledge presupposes the linguistic understanding of speakers.
3. To analyse the explanation of mathematical knowledge proposed by Hamkinsian multiversism and full-blooded platonism. As in the previous sub-objective, the analysis will show how the reduc-

tion of mathematical knowledge to logical knowledge presupposes the linguistic understanding of speakers.

4. To use abstraction principles to explain linguistic competence in mathematics as well as to explain mathematical knowledge. The reductionist perspective introduced in metasemantics by these principles shows that cognitive achievement in mathematics is a form of linguistic competence.
5. To recover a mathematical theory powerful enough to develop advanced mathematical reasoning: Zermelo-Fraenkel set theory. In addition, the abstractionist explanations developed so far will be improved. Abstraction principles will be modified in such a way that they are sufficient to prove the axiom of Infinity and, thus, to explain how speakers acquire the concept of actual infinity.

1.3 METHODOLOGY

The methodology underlying this research project is eminently interdisciplinary, embracing a rich pool of resources characteristic of the fields of the humanities —specifically, contemporary analytic philosophy— and of the formal sciences, such as logic and mathematics. This interdisciplinary approach is required by the objectives outlined above —in general terms, to obtain a better understanding of the epistemology, metasemantics and metaphysics of mathematical theories. In the first place, this effort requires the development of theories that provide us with adequate explanations of the relevant philosophical phenomena. In turn, this demands both an analytical methodology, capable of designing or elucidating the concepts involved in such theories, and an argumentative methodology by which the different proposed explanations are evaluated and compared. In this way, it will be possible to give reasoned answers to the questions on which the present project focuses. Furthermore, the task presupposes a detailed understanding of the logical properties of both these philosophical theories and the mathematical theories under discussion. Given the wide range of resources that fall under this description, it will be useful to classify them according to their general scope —being present in a wide range of philosophical and formal investigations— and to those resources that are relevant for the specific purposes of this project.

General Methodology in Philosophy (i) Analysis —understood in a broad sense— of those concepts that appear in the philosophical explanations under discussion. These include epistemological concepts, such as *objectivity* or *mathematical knowledge*, semantic concepts, such as *reference*, metaphysical concepts, such as *mathematical object*, or logical concepts, such as the concept of *abstraction principle*. (ii) Analysis and description of those theories and philosophical positions considered in the dissertation: mathematical platonism, fictionalism, multiversism and abstractionism. (iii) Construction of arguments for the evaluation and comparison of these positions and the relevant explanations they offer.

Specific methodology in Philosophy (i) Construction of arguments against pluralist conceptions of mathematics (fictionalism and multiversism). (ii) Use of predicative abstraction principles to elaborate an explanation of justified mathematical belief and a metasemantic explanation of mathematical language compatible with platonism. (iii) Development of a stronger abstraction principle to improve the proposal.

General Methodology in the Formal Sciences (i) Use of classical logic and some of its extensions—modal logic and higher-order logic—in the construction of arguments and explanations of relevant philosophical phenomena. (ii) Use of methods of mathematical proof in these arguments and explanations. (iii) Use of well-known mathematical theories as case studies.

Specific Methodology in the Formal Sciences (i) Use of classical logic in the construction of the above arguments against plural conceptions of mathematics. (ii) Use of arithmetic and Zermelo-Fraenkel set theory as a case study. (iii) Use of modal logic and higher-order logic—plural logic—in the development of a partial interpretation of Zermelo-Fraenkel set theory into a modal-plural theory describing the iterated application of abstraction principles. (iv) Use of additional resources of modal logic in the design of new modal abstraction principles.

1.4 STRUCTURE OF THE THESIS

Chapter 2 I will analyze the most intuitive explanation of mathematical objectivity: mathematical platonism. First, I will characterize the objectivity of empirical knowledge as the conjunction of two theses: Independence and Responsiveness. The former claims that the content of those beliefs, judgements, or statements involved in objective knowledge is independent of agents' epistemic responses. In turn, the second thesis claims that objective knowledge involves a strong, counterfactual relation between agents and the world. Second, I will characterize mathematical platonism as the philosophical thesis that abstract objects exist independently from us. According to it, mathematics is the science that studies those objects. Third, I will explain Benacerraf's challenge. I will argue that the challenge is only conclusive against platonism if mathematical knowledge satisfies Responsiveness. Finally, I will explain that platonism does not involve a particular thesis about the metaphysical structure of the world or a theory of truth, and elaborate on the idea of some objects being epistemically independent from agents.

Chapter 3 I will analyze fictionalism, the philosophical thesis that mathematical theories are fictions, and thus untrue. Fictionalism reduces mathematical knowledge to logical knowledge: knowledge about the consistency and consequences of those fictions. I will challenge fictionalism for three reasons. First, I will argue that under two reasonable assumptions, it cannot explain mathematical reliability. Then, I will try to do better, arguing that even if these assumptions fail, fictionalism cannot explain mathematical reliability. Second, I will distinguish between two types of fictionalism: *full fictionalism* and *algebraic fictionalism*. The former claims that, although mathematical theories are untrue, mathematical language is a meaningful language about abstract objects. In turn, the latter does not attribute any content to mathematical language. I will argue that if full fictionalism can explain how mathematical vocabulary acquires its content and how mathematicians acquire their linguistic competence, then so can platonism. Since the failure of platonism is the best argument for fictionalism, this is fatal for it. Finally, I will argue that algebraic fictionalism is in no better position. The reduction of mathematical knowledge to logical knowledge presupposes the linguistic competence of mathematicians. But, if algebraic fictionalism tries to explain this, then it reduces to full fictionalism.

Chapter 4 I will analyze multiversism. This is a pluralist form of mathematical platonism that contends that there are many universes of mathematical objects. Instead of a single, privileged universe

there is a *multiverse*. All the universes are equivalent from the point of view of semantics and metaphysics. As fictionalism, multiversism reduces mathematical knowledge to logical knowledge. First, I will explain that there are many different types of multiversism. Second, I will explain the two main arguments in favour of multiversism: the argument from logical independence and the argument from categoricity. Third, I will argue that the two more radical versions of multiversism are not good accounts of mathematics. I will argue against Hamkinsian multiversism, opposing the challenge of explaining the mathematical vocabulary. Then, I will argue that full-blooded platonism lacks an appropriate notion of consistency for explaining mathematical knowledge. Finally, I will argue that on the face of a realist ontology, the reduction of logical knowledge to mathematical knowledge imposes unreasonable restrictions on the latter.

Chapter 5 I will use the abstractionist approach to mathematical objects advocated by Linnebo (81; 82) to explain the linguistic competence of mathematicians as well as mathematical knowledge. First, I will explain what abstraction principles are. These principles introduce a reductionist approach in metasemantics, allowing us to explain how a community of speakers acquires mathematical language without presupposing any counterfactual or logical relation between them and mathematical facts. Second, I will explain the two main challenges of abstractionism: the Caesar problem and the Bad Company problem. Then, I will show how these challenges can be overridden. Once we are sure that abstraction principles are in good condition, I will show how they can be used to explain mathematical knowledge too. The abstraction process by which speakers extend their language also provides them with knowledge of a collection of truths—and not only logical truths. Thus, as Linnebo (79; 82) and Rayo (115; 116) claim, mathematical knowledge comes with the understanding of language.

Chapter 6 I will explain how abstraction principles are sufficient to recover Zermelo-Fraenkel set theory (82, Ch. 3, 12). First, I will explain how predicative abstraction principles allow to iterate the abstraction attempts, generating a series of domain expansions. Then, I will explain how domain expansions can be expressed at the level of the object language using modal operators. Second, I will contrast set-theoretic potentialism and actualism. Set-theoretic potentialism is the view that the universe of sets forms a plurality of objects. In turn, set-theoretic potentialism denies this. I will argue that the latter is in a better condition than the former. Third, I will introduce the iterative conception of sets. I will show how the iterated application of abstraction principles provides us with an explanation of its two primitive operations: the Set of and the Powerset operations. This explanation entails set-theoretic potentialism. Thus, I will expose those modal principles that govern the process of iteration and those that characterize the nature of sets. Together, all these principles form the *potentialist set theory*. Finally, I will show how the potentialist set theory interprets Zermelo-Fraenkel set theory—without the axioms of Infinity and Replacement. To recover both axioms the potentialist set theory must be supplemented with additional, stronger principles.

Chapter 7 I will develop a new type of abstraction principle: modal abstraction principles. Through them, I will explain how speakers acquire the ability to conceptualize the actual infinite through the process of abstraction. First, I will discuss how modal abstraction principles should be formalised. For this, I will extend the modal language by introducing two new kinds of operators: the black modal operators, and the backtracking and unbacktracking operators. Second, I will modify and accommodate all elements

of the abstraction process to the new principles of modal abstraction. Once this is done, abstraction can be used again to extend the language and secure reference to additional objects. Third, I will use abstraction principles to prove the axiom of Infinity. Finally, the requirements and difficulties of generalizing the proposal will be assessed.

2 Challenges to Mathematical Objectivity

2.1 INTRODUCTION

In this chapter, I will analyze the most intuitive explanation of mathematical objectivity: mathematical platonism. According to mathematical platonism, mathematics is a science of abstract objects. This is a simple and appealing thesis, but it has been heavily criticized. The main argument against it, Benacerraf's challenge, claims that if platonism is true, then mathematical knowledge cannot be explained. Those who accept the challenge consider it an insurmountable obstacle to platonism.

First, I will characterize the objectivity of empirical knowledge as the conjunction of two theses: Independence and Responsiveness. The former claims that the content of those beliefs, judgements, or statements involved in objective knowledge is independent of agents' epistemic responses. In turn, the second thesis claims that objective knowledge involves a strong, counterfactual relation between agents and the world. Second, I will characterize mathematical platonism as the philosophical thesis that abstract objects exist independently from us. According to it, mathematics is the science that studies those objects. Third, I will explain Benacerraf's challenge. Then, I will argue that the challenge is only conclusive against platonism if mathematical knowledge satisfies Responsiveness. However, this is unjustified. Fourth, I will explain what platonism is not. It does not involve a thesis about the metaphysical structure of the world or a theory of truth. Finally, I will elaborate on the idea of some objects being epistemically independent from us.

2.2 A REALIST CONCEPTION OF KNOWLEDGE

Imagine a sailor on an ancient boat, sailing unknown seas. This sailor has a unique and special task: to watch for land, ships or any danger, as well as to report to the other members of the crew about these. Lookouts were extremely important a few centuries ago. These persons possessed excellent vision and had to be extremely reliable. Any mistake could endanger the whole crew. Thus, our sailor is in a delicate situation: there *seems* to be land far away, in front of the boat, but she is not entirely sure. A tiny, blurred and brownish spot has just appeared over the horizon. Given the perceptual conditions, the travel time and his ocean knowledge, the evidence thresholds do not support an alternative, inconsistent judgment. However, she does not want to rush. Everybody is tired and anxious to reach the mainland to rest and eat properly. Therefore, she is cautious. She patiently waits and the spot slowly grows a little bit. It is neither big nor clear enough. Nobody –even those with the best eyesight– could discern if it is the

promised new land or just a cloud. However, her eagerness and her desire to finish the journey gain, and she rashly reports that they are arriving mainland. However, after a few minutes, few are those who are joyful: the spot was not land at all, it was a dark and huge cloud. An unexpected storm is coming quickly. Her fellows blame her and back to work, hurrying to get ready for the storm. They have reasons to be disappointed with the lookout because she has cheated. After all, despite evidence thresholds did not support alternative, inconsistent statements, they did not support her judgement either. Tired and excited by the first impression, she “decided” to fill in what evidence lacked. In sum, —unintentionally— she has not been *objective* and her claim did not bear knowledge. Consequently, the crew must face the consequences of a false judgment.

This example shows how important objectivity is for humans. Objectivity is the mark of correct knowledge *about the world* and the requirement of observing it is almost a truism about all knowledge-seeking processes, science among them. To claim that there is land ahead, that hydrogen atoms possess one electron or that $7 + 5 = 12$ is not as cheap as claiming that I feel moody, that certain clothing is fashionable or that a joke is funny. Former cases involve something more than purely epistemic factors only accessible to me. Contemporary empirical sciences, such as physics, chemistry or biology, involve increasingly complex and entangled research methods. They require huge teams of researchers, the manipulation of big and expensive technological devices and a huge deal of theory employed both to perform these manipulations and to interpret the data obtained. Undoubtedly, the probability of making misjudgements in such an environment is higher than in the context of a quiet and sunny day in the ocean. This is the reason why scientists are extremely careful when it comes to judging. What would happen if a scientist asserts that p even when evidence is not enough to support that p ? Even worse, what happens if evidence involves a misinterpretation of the data or a misuse of the instruments? At best, she will realize the mistake. In the worst scenario, the statement would be published and a scandal would ensue.

The concept of *objectivity* is essentially epistemic: it qualifies the properties of a special kind of knowledge, our knowledge about the world. Although no definition is expected, *Knowledge* is a normative concept. It involves norms, that is, some link between specific conditions and agents’ epistemic actions which establishes a particular standard of correctness for these actions.¹ This is the case of as-

¹I will not develop a formal analysis of the concept of *norm* here. First, I leave open what kind of connection between conditions and actions is required. For instance, (151) employs an *ontological* concept of norm. According to it, the link is a property —properties are another headache for philosophers— of statements and beliefs. The main example is the minimal property of “being true”, which —modulo classical logic— divides sentences or propositions into true and false sentences or propositions. Thus, given the condition that p , “ p ” must be true. Generally, metaphysics establishes the laws determining when some object has some property. Similarly, this is the case of physical laws. Alternatively, there is a *linguistic* concept of norm. According to it, norms would be theories. Specifically, they would be conditionals, stated in some language $\mathcal{L}$ which is correctly understood by speakers. However, such understanding would already involve a “previous” understanding of the norms governing $\mathcal{L}$. As a result, the threat of circularity and paradox arises.

The norms affecting knowledge are pragmatic: they impose standards of correctness on agents’ epistemic actions. Thus, the kind of modality they involve is *deontic* modality. I will represent in this chapter duty by black modal operators “■” and “◆”. Usually, the system employed to analyze it is $\mathcal{K}_\eta$, where “ought” implies “may”: ■ $\varphi \rightarrow$ ◆ φ . See (105, pp. 47-48) for some reasonable doubts about this system. Some norms are constitutive of activities. For example, the rules of chess are constitutive of this game. However, obedience to norms is not necessary for the performance of activities *constituted* by them. A rule of football is not to touch the ball with the hand. However, if a player touches —intentionally or unintentionally— the ball with her hands it will be a conceptual

serting and judging.² Under some necessary or sufficient conditions, statements and judgements satisfy the required standards. In the best scenario, correct statements and judgements about the world are constituted as knowledge-bearers.³

So far nobody has been able to offer a reasonable characterization of the norms involved in knowledge. Gettier cases have been extremely resistant to philosophical treatment, and nobody expects a successful reductionist analysis soon. All it is agreed is that knowledge is factive: if it is known that *p*, then *p*. Consequently, knowledge is usually taken as a primitive concept. And, this transitively applies to objective knowledge. A reductionist analysis of this concept seems unreasonable.⁴ Nevertheless, objectivity hides an appealing and intuitive idea. This consists in the fact that the content of beliefs, judgements, or statements bearing knowledge is —partially at least— non-epistemic. That is, they are not about agents' epistemic responses —for example, judging, believing, asserting, reporting or answering. Instead, they are about the world. Again, consider the former example. The sailor was tired and anxious to reach the mainland. Furthermore, evidence thresholds did not support the opposite judgment that the tiny, brownish spot was a cloud. As a result, she erroneously “concluded” that the crew was arriving mainland. A few minutes later they knew that the boat was facing an unexpected storm. The looker's report was *incorrect*. This example shows that the content of her report exceeded her epistemic responses. Indeed, the error was caused because she included her responses —her belief that the spot was not a cloud, caused by her desires and expectations— among appropriate evidence required by the situation. As the usual slogan says: objective knowledge is *independent* —in some sense of “independence”— from agents (23; 24; 25; 118). I will not attempt any definition or reductionist analysis of “objectivity”. Nevertheless, I will attribute two necessary theses to the norms ruling objective knowledge.

Independence The content of beliefs, judgements, or statements involved in objective knowledge does not *depend* on epistemic responses as believing, judging, asserting, reporting, answering, etc.

Responsiveness Objective knowledge partially *depends* on the content of beliefs and theories involved in knowledge.

The idea behind the conjunction of these theses is the following: against anti-realist conceptions of knowledge, knowledge about the world exceeds knowledge about agents' states. Independence makes this explicit. Moreover, it is reasonable to expect the content involved in knowledge —the world— to be before knowledge itself. The priority direction is the opposite of the direction involved in anti-realist conceptions of knowledge. This is what *Responsiveness* expresses. *Because* the world is such that the law

mistake to conclude that she is no longer playing football. At most, what is indispensable is that agents understand the norms governing their activity. See (147).

²An assertion is a communicative action that consists of sanctioning the content of some sentence as true. A judgment is a psychological action that consists of sanctioning some propositional content as true.

³Knowledge qualifies actions and entities relating to propositional content, such as statements, judgements and beliefs.

⁴See (118) for several attempts of analysis and why they have failed. A further challenge arises from the complex etymology of “objectivity”. The word has been used in a variety of different ways, acquiring a complex entanglement of meanings. Sometimes it applies to theories, while others it applies to methods or attitudes. It is employed as a moral value too. See (23; 24; 25).

of gravitation⁵ holds, agents can gather knowledge about it. Compare this with knowledge about fashion or comedy. Some clothing is fashionable because, in the first place, people find it so. Also, jokes are funny, or disgusting, because of people's sense of humour. In both cases, truths are grounded in people's responses.

2.2.1 Independence

Both Independence and Responsiveness include a notion of *dependency* in their formulation. This notion is used very often in descriptions of what objective knowledge is. But, how should it be understood? In the case of Independence, (151) has proposed a characterization in terms of analyticity. In his book *Truth and Objectivity*, Wright develops four criteria for assessing the objectivity of the theory formed by the true sentences of a language —thus, for the objectivity of languages. Two of these criteria — the projectivist position concerning the Euthyphro contrast and Cognitive Command— explain how a language is epistemically loaded by appealing to the notion of analyticity. Simplifying things a bit, the former criterion holds that a language is not objective when the Knowability Principle (KP)

KP

$$\phi \rightarrow K\phi$$

—where “ ϕ ” is a metavariable representing sentences of the language and “ K ” is a sentential operator expressing knowledge.⁶— is analytic of the content of the language. Consequently, KP would imply —modulo factivity— that ϕ and $K\phi$ are analytically equivalent. In turn, the second criterion holds that a language is objective when the fact that KP does not hold is analytic of the language. Leaving aside the challenges of analyticity, the proposal was criticized by (148) on the basis that it imposes unreasonable constraints on metasemantic explanations⁷ of reference. Wright's criteria entail that both non-objective and objective languages fix reference by analytic factors alone. However, this is rarely the case. Remember the moral of Putnam's (108) Twin Earth thought experiment. For example, even in the case of colour —which Wright considers a case of an anti-realist discourse— empirical discoveries determine the reference of colour words and which truths about colours speakers accept. Physics has something to say about that. Consequently, Wright's criteria for objectivity or non-objectivity fail systematically.

I will not endorse the required notion of dependency in Independence with any commitment to analyticity. On the contrary, I will assume that what is required is an appropriate notion of explanation

⁵Newton's gravitation law states that a mass point is attracted by another mass point with a force proportional to the product of the masses and inversely proportional to their distance. The law is taken as an approximation to the solutions of relativistic field equations under low-speed conditions.

⁶Usually, KP is formulated as hidden a modal operator:

$$\phi \rightarrow \Diamond K\phi.$$

Nevertheless, Fitch's paradox (44; 14) shows that this formulation collapses in the former under two modest assumptions: that K distributes over conjunction and factivity. Because of this, and Wright's exposition of the issue, it does not matter which formulation is chosen.

⁷Metasemantics *explains* how words acquire its meaning —I am using here a vague notion of *meaning*. The fact that a word has a particular meaning is not a metaphysical primitive. In turn, semantics *describes* what each word means and how their meanings relate to each other.

—maybe, a metaphysical notion of explanation. If the content of the language depends on speakers’ epistemic responses this means that it is *explained* at least in part by speakers’ epistemic responses. This is consistent with the fact that metasemantic explanations do not include exclusively analytical factors. Even if the content of language involves speakers’ epistemic responses, those facts explaining how the language acquires its meaning can involve empirical matters. According to this, Independence says that the content of languages is not *explained* by the epistemic responses of speakers. Using the symbol “ $\Vdash$ ” to represent the appropriate notion of explanation, this can be expressed as follows:

$$\text{Independence: } C\phi \nVdash \phi,$$

where “ C ” represents speakers’ cognitive responses regarding the sentence ϕ .^{8,9}

There is a reasonable doubt as to whether this formulation implies that knowledge is not factive. For, suppose that conditions $C\phi$ are entangled with knowledge of ϕ . Then, Independence would entail $K\phi \nVdash \phi$. Fortunately, even if this were the case this is not equivalent to the negation of factivity. Factivity is the conditional

$$K\phi \rightarrow \phi.$$

If this were read as expressing something stronger than a material conditional, surely this would not be the same notion of explanation as the represented by “ $\Vdash$ ”. Factivity is analytically true of the notion of knowledge —i.e. $\mathcal{A}(K\phi \rightarrow \phi)$. On the contrary, the idea of dependency involves a notion of explanation that must be weaker than analyticity. Therefore, factivity and Independence are mutually consistent. Conversely, $\Vdash$ must be stronger than a plain strict conditional —i.e. $\Box(\phi \rightarrow \psi)$. The idea of dependency entails that some facts¹⁰ are more fundamental than others. Thus, $\phi \Vdash \psi$ says that ϕ is more fundamental than ψ — ψ is the case *because* of ϕ . As every explanatory relation, dependency is asymmetric.¹¹ On

⁸As relations of entailment, the notion of explanation could be generalized as holding between sets of facts and facts:

$$\Gamma \Vdash \phi,$$

where Γ is a set of facts.

⁹Rosen (121, p. 127) proposes a similar analysis of Idealism in terms of metaphysical grounding.

¹⁰I will use a minimal notion of *fact*. This ensures that facts are simply the metaphysical correlate of the concept of truth. Thus, given a sentence ϕ , I will speak interchangeably of “the truth of ϕ ” or “of the fact that ϕ ”. See 2.6 for a proof that a minimal notion of fact is conservative over a minimal notion of truth. For those who distrust metaphysical concepts, any mention of fact can be replaced by speaking of a sentence being true.

¹¹I will leave open which one is the precise relation expressed by “ $\Vdash$ ”. A candidate is the relation of *metaphysical grounding* (120). Either way, it should satisfy the following properties —where Δ and Γ represent sets of sentences expressing facts or propositions:

<i>Transitive</i>	If $\Gamma, \phi \Vdash \psi$ and $\Delta \Vdash \phi$, then $\Gamma, \Delta \Vdash \psi$
<i>Irreflexive</i>	$\phi \nVdash \phi$
<i>Asymmetric</i>	If $\phi \Vdash \psi$, then $\psi \nVdash \phi$.
<i>Not connected</i>	It is not the case that: $\Gamma, \phi \Vdash \psi$ or $\Gamma, \psi \Vdash \phi$
<i>Non-monotonic</i>	It is not the case that if $\Gamma \Vdash \phi$, then $\Gamma, \psi \Vdash \phi$

Explanation must be non-monotonic. On the contrary, every fact would partially explain every other fact. This is highly unpleasant, almost trivializing the explanatory role. For instance, assume that mental states are grounded —or supervene— on the brain’s physiology. Then, by no means they would be explained —or supervene— by the

the contrary, the paradoxes of the strict conditional show that it is not asymmetric at all. The explanation hidden in the idioms of dependency is in between analyticity and the strict conditional.

In turn, if $C\phi$ is entangled with $K\phi$, Independence is equivalent to the negation of certain formulation of KP. This thesis can be understood as something stronger than a material conditional. According to KP, facts are intrinsically epistemic: there is no gap between them and agents' knowledge of them. Therefore, KP can be understood as saying that ϕ depends on or is explained by $K\phi$: $K\phi \Vdash \phi$. A realist conception of knowledge strongly rejects this assumption. This is precisely what Independence says. Consequently, $\Vdash$ could help to identify what is the core of the division between realist and anti-realist conceptions of knowledge.

Note that Independence does not imply that the language is referentially successful. That is, it does not imply that the language is geared with terms and generalized quantifiers, as well as committed to the existence of a particular realm of objects. For example, suppose that you accept a *de re* conception of modality. Terms and quantifiers nested inside intensional contexts would be referentially innocuous. However, according to such a conception of modality, it does not reduce to analyticity or other representational concepts. *De re* modality is realist in spirit, regarding modal laws as independent from agents as natural laws. Also, see (12) for the possibility of languages with no referential apparatus at all.

Despite its intuitive appeal, it might seem that Independence faces some counterexamples. For example, does psychology contradict this thesis? After all, psychology is intended as an objective science, but its theories are precisely about human cognitive responses. This seems to imply a contradiction: psychology would not be an objective science after all. Fortunately, this is not the case. What Independence requires is that the content of those statements endorsed by any speaker does not depend on *her* epistemic responses. This is completely satisfied by psychology as well as any other human science. The knowledge gathered by psychologists is not determined by their epistemic responses. Their object of study is human beings —generally, cognitive organisms—, but how these subjects are psychologically constituted does not depend on the researchers' epistemic responses. Thus, that C-fibers are responsible for pain is by no means the case *because* some researchers judged it so. Also, that a particular subject faces some psychological problem is by no means the case *because* her therapist judged it so. Truthfully, because researchers correctly judge that p is the case, they obtain knowledge about p . But p itself —the content of their judgement— is not the case because they judge it so.

But, is not Independence too strong? After all, all attempts to analyse objectivity have failed so far (118). The efforts made to isolate those factors tied to the researchers' judgement from knowledge-gathering processes have shown hopeless. As a result, a substantial amount of studies —especially those focused on understanding how social and pragmatic factors affect science— have criticized the very idea

brain's physiology and the fact that the Earth and the Moon satisfy —approximately— Newton's gravitational law —leaving aside marvellous surprises in physics! As a result, classical semantic entailment is not a sub-relation of the required notion of explanation: $\models \not\subseteq \Vdash$.

Otherwise, I will not commit myself to a Lewisian conception of a fundamental metaphysical domain over which everything else pends. Thus, I will leave it open whether $\Vdash$ is well-founded. That is, I will leave open the question concerning the existence of infinite descendent chains of fundamentality between facts —... $\phi_{\alpha+1} \Vdash \phi_\alpha \Vdash \dots \Vdash \phi_3 \Vdash \phi_2 \Vdash \phi_1$. It is natural to extrapolate humans' usual experience about composites, such as sets or mereological fusions, concluding that there should be a fundamental domain of grounding facts which depend on nothing. However, humans' experience is fallible. There is nothing incoherent about atomless worlds.

contained in Independence. As a human activity, science and, more generally, knowledge—including mathematics— would involve an irreducible remainder of the researchers’ dispositions. For instance, mathematical axioms, as Power Set¹², would be justified not because of mathematicians’ privileged access to the set-theoretic universe, but because of pragmatic reasons involved in the practice of 18th and 19th-century that lead to contemporary set theory (35, Ch. 8). These kinds of proposals tend to reject a precise distinction between the context of discovery and the context of justification. According to them, epistemology would hopelessly involve the study of agents’ responses and it would be impossible even *in principle* to disentangle where the frontier between the world and their responses is. Therefore, is not Independence overlooking all this? I believe the answer is no. The previous considerations are orthogonal to Independence. This is a characterization of a realist conception of knowledge. Such a characterization is consistent with any challenge to humans’ cognitive abilities, even if the latter entails the unattainability of the former. Humans are far from being perfect when it comes to knowledge about the world. Their justifications of empirical statements as well as statements of formal science involve mistakes and inaccuracies—for instance, physical magnitudes can only be approached to a rational value—and several sorts of bias—gender, social or economical (31; 118). But, this is consistent with a meaningful notion of objectivity.

Moreover, Independence is a thesis about the content of agents’ epistemic actions, not about the *justification* of those actions. Thus, think again about the first example and reverse it. Imagine that indeed there was land in front of the boat but the looker confounded it with a storm. Consequently, the crew would change direction getting lost in the ocean. Even if they remained convinced forever that there was a storm far away, this would not be the case. To think the opposite would be almost to believe in magic. Hence, it is important to distinguish between the content of language and the evidence thresholds attained to theories.

2.2.2 Responsiveness

The notion of objective knowledge as correct knowledge about the world is overtly elaborated with empirical science as a reference. Thus, the notion of dependency that features in Responsiveness is usually taken as hiding a counterfactual relation between speakers and facts. As Benacerraf puts it:

We think that X could not know that p . What reasons can we offer in support of our view? If we are satisfied that X has normal inferential powers, that p is indeed true, etc., we are often thrown back on arguing that X could not have come into possession of the relevant evidence or reasons: that X’s four-dimensional space-time worm does not make the necessary (causal) contact with the grounds of the truth of the proposition for X to be in possession of evidence adequate to support the inference (if an inference was relevant). The proposition p places restrictions on what the world can be like. Our knowledge of the world, combined with our understanding of the restrictions placed by p , given by the truth conditions of p , will often tell us that a given individual could not have come into possession of

¹²A set b is a subset of another set a if and only if every element of b is also an element of a . The powerset of a —in symbols $\mathcal{P}(a)$ — is the set of all its subsets. The powerset axiom asserts that the powerset of every set exists.

evidence sufficient to come to know p , and we will thus deny his claim to the knowledge.
(7, 671-672)

According to this understanding of Responsiveness, this should be expressed as follows using the notion of explanation introduced in the previous section:

$$\text{Responsiveness } C\phi + \phi + X \Vdash K\phi,$$

where $C\phi$ expresses some counterfactual relation between agents and the facts depicted by ϕ , and “ X ” is a variable standing for some appropriate facts —this unknown prevents any simplistic definition of knowledge. X can be non-constructively thought of as the required epistemic conditions under which ϕ is known. Thus, $C\phi + \phi$ partially explains $K\phi$. Responsiveness says that when agents are appropriately related to the world by C , ϕ is true and X , then they know that ϕ . In the case of empirical knowledge, the counterfactual relation C is a causal relation between agents and the world. For example, knowledge about Newton’s law of inertia partially depends on physical objects satisfying it. Also, knowledge about continental drift partially depends on a complex geological history of changes in Earth’s lithosphere.

As can be seen, Responsiveness involves an explanation of knowledge. Therefore, the following worry will arise: does it entail some kind of solution to Gettier’s challenge? After all, this challenge calls for necessary and sufficient conditions under which agents can know that ϕ . If the antecedent of Responsiveness collapses into these conditions, requiring a viable solution to the challenge, there would be little chance of satisfying this criterion. Fortunately, I believe this is not the case. At best, Responsiveness entails a *local* solution to the challenge. On the contrary, it does not entail a *global* solution. But, this is precisely what the challenge asks for. Let $\mathcal{L}$ be the speakers’ language. A global solution to the challenge involves an explanation of the form

$$\text{For all } \phi \in \mathcal{L}: \Gamma \Vdash K\phi,$$

where Γ is a set of sentences. Indeed, this would be a definition of knowledge. On the contrary, a *local* solution to the challenge only affects a restricted collection of sentences Δ of the language:

$$\text{For all } \phi \in \Delta \subseteq \mathcal{L}: \Gamma \Vdash K\phi.$$

Therefore, even if “ X ” is solved in many cases, Responsiveness does not collapse into a solution to Gettier’s challenge. While a global solution to Gettier’s challenge —especially, an explanation of perceptual knowledge— is an excessive requirement, a local solution is precisely what the epistemology of special sciences calls for.

For instance, it is completely coherent to explain mathematical knowledge, without elaborating a solution to the epistemic difficulties involved in perception. A realist epistemology of arithmetic calls *at least* for a solution for X in the following expression

$$\text{For all } \phi \in \{\psi / PA \vdash \psi\}: C\phi + PA + X \Vdash K\phi.$$



And a realist epistemology of sets *at least* requires a solution for X in

For all $\varphi \in \{\psi / \text{ZF} \vdash \psi\}$: $C\varphi + \text{ZF} + X \Vdash K\varphi$.

Nevertheless, it is this approach that creates the biggest problems for mathematical knowledge. As I will argue, Responsiveness is a bad characterization of mathematical knowledge. It does not provide a *de facto* global solution, but because it cannot do so *in principle*. While it is appropriate as a characterization of empirical knowledge, it fails as a global explanation of knowledge.

2.3 MATHEMATICAL PLATONISM

The previous characterization of a realist conception of knowledge provides a helpful explanation of the most prominent challenge that mathematical epistemology faces: Benacerraf's challenge (7). In this section, I will explain what mathematical platonism is. Then, I will explain which are the main epistemological challenges it faces by using Independence and Responsiveness.

2.3.1 What is Mathematical Platonism?

Mathematical platonism is the philosophical position resulting from the conjunction of the following three theses: Existence, Abstractness and Independence (138, Ch. 3-4), (2, p. 9), (116, pp. 9, 74), (85).

Existence Mathematical entities exist.

Abstractness Mathematical entities are abstract.

Independence Mathematical entities are independent from agents.

In what follows, I will comment on each one of these theses.

2.3.1.1 Existence

Existence is a *metaphysical* thesis. —i.e., a thesis about what there is. Metaphysical theses are opposed to *hermeneutic* theses concerning the content of language.¹³ While the latter aims to explain what mathematical language, as employed by mathematicians when doing mathematics, is about, the former does not. Consequently, it is coherent to endorse Existence and at the same time to claim that mathematicians were not historically committed to any special realm of entities. For instance, their theories and methods could have been highly constructive, supplying their languages with anti-realistic semantics, such as intuitionistic or procedural semantics—as proposed by (41; 43). In parallel, using sophisticated metaphysical arguments, philosophers could have concluded that a weird domain of objects obeying mathematicians' theories does exist. Even more, these objects could support classical logic, and the theories about them

¹³I take the “hermeneutic” label from (2) and (102). The distinction between these two kinds of thesis resembles the Quinean distinction between *explanatory* and *hermeneutic* philosophies of mathematics. An explanatory philosophy of mathematics is prescriptive. It aims to settle what mathematics *ought to be* about. Take set-theoretic reductions of arithmetic. According to explanatory proposals of arithmetical language, arithmetical terms should be understood as referring to a particular collection of sets—for example, von Neumann ordinals. Explanations prescribe what language should be about. On the contrary, hermeneutic philosophies of mathematics aim to describe what the languages used by mathematicians are about.

would be proper extensions of mathematicians' old constructive ones. Although this picture is highly implausible, it is completely coherent with Existence as stated above. Nonetheless, it is natural to expect some connection between metaphysical and hermeneutic accounts of language. After all, the relevant entities are qualified as "mathematical" because they obey mathematicians' theories. Consequently, Existence must be understood as a statement about mathematical language: if read at face value, this language is about a distinctive realm of entities, mathematical entities.

There is a meta-ontological delicate point concerning this thesis: what does it mean "to exist"? From now on I will assume Frege's and Russell's monochronism (137, p. 9) about meta-ontological notions. Therefore, I will reduce existence to the concept expressed by the existential quantifier: a second-order concept loaded with ontological commitments. This is the relevant notion of existence for mathematical platonism. Otherwise, note that the above three theses speak about *entities*, not about *objects*. I consciously employ this notion to be neutral about the ontological status of the mathematical realm. More precisely, it should not be presupposed that mathematical platonism is about objects. For instance, *ante rem* structuralism claims that mathematics is about structures. Moreover, *ante rem* structuralism is usually regarded as a platonic philosophy of mathematics. But, are structures objects? A negative answer calls for fine-grained classifications of different types of platonism. For instance, *ante rem* structuralism, object-platonism, etc. To simplify things, I will adopt a Fregean concept of object.¹⁴ This is generous enough to encompass the notions of entity, thing, individual, and the like. As a result, the considerations in this text apply both to *ante rem* structuralism and to object-platonism.

Finally, Existence can be relativized to particular mathematical theories.

Existence_{PA} Natural numbers exist.

Existence_{ZF} Sets exist

Note that if mathematical theories *T* are framed in first-order logic and they are read at face value, the rejection of *Existence_T* implies that such theories are at least partially false.

2.3.1.2 Abstractness

Abstractness makes explicit the most remarkable metaphysical feature of mathematical objects: they are abstract. Despite the complexity of this expression and the doubts expressed by some people (2, Ch. 8), there is a simple way to characterise it. An entity is abstract when it does not exemplify spatio-temporal properties. Thus, provided that every physical entity is spatiotemporally located, abstract entities are not among the physics inventory. For instance, think about real numbers, real-valued functions, or manifolds. Such entities are essential for empirical sciences, as physical magnitudes reveal. Physical magnitudes, such as temperature, speed, intensity, or mass, are understood as total, real-valued functions defined on the domain of concrete, spatiotemporally located objects. However, neither real numbers nor real-valued

¹⁴According to this, objects are characterized semantically as possible referents of terms. The result is a broad notion of object. For instance, as soon as theoreticians refer to concepts, concepts are objects too. Nonetheless, the characterization allows for objects of different order and different kinds. This suffices to do justice to Frege's maxim: "never to lose sight of the distinction between concept and object." (46, p. x)

functions are part of this domain. Abstract objects are, loosely speaking, “out” of space and time. These objects are not just spatiotemporally isolated from us —like distant galaxies and other remote events. On the contrary, they are —so to say— causally ineffective.

Abstractness can be understood as a thesis about the semantic status of physical attributions in two different ways. First, it can be understood as claiming that statements endorsing abstract objects with physical properties are false. For example, suppose that somebody says “the number 2 is blue”. Furthermore, assume that the context disambiguates her utterance, making it clear that she is referring to an abstract object —i.e. the number two. Then, if her utterance is an assertion, this will at least be false — numbers have no colour. Alternatively, abstractness can be understood as claiming that such statements are meaningless. That is, physical attributions involving abstract objects would be ill-formed expressions breaking the rules of grammar (137, pp. 93-94), or at best, they would bear a third truth value —neither the truth nor the false. According to the first, the sentence “the number 2 is blue” is ungrammatical. Grammatical rules forbid “to be blue” to be predicated on the term “the number 2”. Although predicates are allowed to apply to terms, this type of construction could be an exception. Maybe constructions combining terms that refer to some kind of entities —abstract objects— and predicates that express a property about other, very different kinds —spatio-temporal objects— hide a categorical mistake. Thus, compositionality would obey a finer classification of terms and predicates. On the other side, these attributions may lack a determined truth value. Even if they are grammatically correct, they may be insufficient to select well-determined semantic content. Or, if they are, the world may be insufficient to make them true or false. Which one is the correct reading is not especially important for my purposes.

2.3.1.3 Independence

This is the thesis explained in section 2.2.1.

$$\text{Independence } C\phi \not\models \phi,$$

where $C\phi$ represents speakers’ cognitive responses regarding the sentence ϕ . Independence constitutes the real core of mathematical platonism. Mathematical entities and mathematical facts are epistemically independent from agents. As some counterexamples show, Existence and Abstractness are insufficient to provide a satisfactory characterization. Let’s start with Existence, and think about mathematical psychologism. Simplifying things a bit, this is the anti-realist thesis that mathematics is about mental events. According to psychologists, numbers exist as mental constructions of some type. Consequently, they would be at least temporarily located objects. This example shows that Existence is consistent with anti-realism. Thus, it is not enough to identify the realist bedrock of mathematical platonism. Also, Abstractness is not enough either. Think about fictionalism, the thesis that mathematical objects and mathematical facts are fictitious —see Chapter 3. According to this, the number 3 would be as Hamlet or Oliver Twist: they do not exist. However, the fictionalist believes there is a clear difference between these two kind of fictitious objects. While Hamlet and Oliver Twist are spatio-temporal objects, the number 3 is abstract. It does not exist. But, if it existed, it would not be a spatiotemporal object. Therefore, Abstractness is also consistent with anti-realism and is not enough to identify platonism.

Mental constructions and fictitious objects are not epistemically independent from agents. This is what constitutes both psychologism and fictionalism as anti-realist philosophies of mathematics. None of these satisfy Independence. When applied to mathematics, this says that mathematical facts, for instance, arithmetical facts, are independent of agents:

$$\text{For all } \varphi \in \{\psi / \text{PA} \vdash \psi\}, \text{ then } C\varphi \not\models \varphi,$$

where C is the corresponding epistemic context. For example, the arithmetical fact that $5 + 7 = 12$ is independent from agents. Independence is about facts. However, as stated in section 2.3.1, it is about objects: mathematical objects are independent of agents. Therefore, the form of the latter statement may give rise to some concern. Are the two formulations of Independence equivalent? Fortunately, the answer is “yes”. The independence of objects can be retrieved as the independence of identity facts. More specifically, an object α is independent of agents when the fact that this object is self-identical — $\alpha = \alpha$ — satisfies Independence.

$$\alpha \text{ is independent from agents} \equiv_{df} C(\alpha = \alpha) \models \alpha = \alpha$$

As it was explained in section 2.2.1, the notion of dependency is analyzed as an appropriate notion of explanation ($\models$) between facts —recall that this is a minimal notion of fact, see 2.6. However, the literature on the philosophy of mathematics contains alternative ways of understanding the notion of dependence. For instance, (82, Ch. 11) argues that mathematical objects satisfy Independence if this is understood as expressing *counterfactual independence*. This form of independence claims that mathematical objects may exist when agents do not —i.e. mathematical worlds uninhabited by agents are possible. If mathematical objects exist necessarily, they will be independent in this sense. However, counterfactual independence is a relatively weak notion and it was criticized by (38). Moreover, if an anti-realist concept of modality is accepted, the counterfactual independence of mathematical facts could be compatible with their inherently epistemic nature. To say that they are counterfactually independent of agents would have no metaphysical import if the space of possibilities is just an epistemic device and not a metaphysical category.

Stronger notions of independence elaborate on comparisons between mathematical and physical objects. Such comparisons demand that numbers, functions and other inhabitants of the mathematical realm must have the same status as mountains, tables or animals —see (82, Ch. 11) Linnebo3 and (93). However, I believe that the best way of understanding such analogies is not by charging mathematical platonism with Independence, but with Responsiveness.

$$\text{Responsiveness } C\varphi + \varphi + X \vdash K\varphi,$$

where $C\varphi$ expresses some counterfactual relation between agents and the facts depicted by φ . Numbers and mountains are alike not only because they are epistemically independent of agents. According to the analogy, they are similar in the way they relate to agents. In the case of mountains, $C\varphi$ would express a causal relation between them and agents. In the case of mathematical objects, it would express other, non-causal kind of relation —mathematical objects are causally ineffective. However, as mathematical

platonism was formulated, Responsiveness is not part of it. The former is only the conjunction of Existence, Abstractness and Independence. I believe that to include Responsiveness in the characterization of platonism would be going too far. It is not part of this concept. When added, the result is a very particular kind of mathematical platonism: Mathematical Platonism, with capital letters (89).

$$\text{Mathematical Platonism} =_{df} \text{mathematical platonism} + \text{Responsiveness}$$

Nevertheless, this is a bad movement for platonism. The addition of Responsiveness causes the main problems for mathematical knowledge. Fortunately, as I will explain in Chapter 5, this movement can be avoided.

2.3.2 The Epistemological Challenge

Mathematical Platonism is regarded as epistemologically problematic. The reason is that unlike spatio-temporal objects, such as books, trees or stars, it is metaphysically impossible to physically interact with mathematical entities. They cannot be perceived, they resist ostensive nomination, and they leave no trace of information in nature. It is not simply that their trail is undetectable because they are causally isolated, like distant galaxies. Technological devices, such as those employed to observe distant bodies in empirical science, do no better. Mathematical objects are non-observable *in principle*: it is and will be not possible to observe them. This is what to be abstract means. Abstractness forbids the attribution of physical properties to mathematical objects. Consequently, this feature is inconsistent with Responsiveness

$$\text{Responsiveness } C\varphi + \varphi + X \Vdash K\varphi,$$

when conditions C —which express an appropriate relation between agents and what φ says— involve a physical relation. But, human beings —let’s discard Cartesian dualism— are entirely physical. Therefore, there is pressure to accept that C must be some kind of physical, counterfactual relation. However, this is forbidden by the abstract nature of mathematical objects. And, if it is not, it is far from clear what C is and how human beings obtain knowledge about such an inaccessible portion of the world. Even more, how can mathematicians be referentially successful?

This is Benacerraf’s challenge (7). Originally, Benacerraf presented the challenge as the impossibility of satisfying at the same time two desiderata imposed on philosophical accounts of mathematics. First, any such philosophy must provide a theory of truth that matches the theory of truth of the whole language. In the words of Benacerraf, a theory of truth which implies truth conditions that are obviously conditions of their truth. According to this desideratum, there is only one notion of truth, and mathematical truth is no different from the truth of sentences about concrete objects. Both cases are part of a general theory of truth for natural language. For Benacerraf this is Tarski’s semantic definition of truth. Second, a successful philosophy of mathematics must explain how mathematical knowledge is possible. This is a reasonable desideratum that any explanation of scientific knowledge must satisfy. If we cannot guarantee that a scientific discipline provides us with knowledge, how can it be considered a science? These are the two desiderata that a satisfactory philosophy of mathematics must satisfy. “An account of mathematical truth, to be acceptable, must be consistent with the possibility of having mathematical knowledge [...].

An acceptable semantics for mathematics must fit an acceptable epistemology” (7, p. 667). Here is where the problem arises. These two desiderata do not go very well together. The most promising way of endowing mathematics with a Tarskian theory of truth is to accept platonism.¹⁵ Thus, mathematical terms are associated with mathematical, abstract objects “whose nature, as normally conceived, places them beyond the reach of the better-understood means of human cognition (e.g., sense perception and the like)” (7, pp. 667-668). As a result, the road to a reasonable epistemology seems to be blocked. The satisfaction of one desideratum leads to the breakdown of the other.

Benacerraf and many other philosophers have seen this dilemma as a challenge to mathematical Platonism. For them, if Platonism is rejected—in addition to achieving a satisfactory epistemology—there is room to find a satisfactory theory of truth. On the contrary, platonism makes any explanation of mathematical knowledge impossible. As I will argue, this follows because they accept Responsiveness as an essential feature of mathematical knowledge. When this is abandoned, there is full room to answer the challenge. Benacerraf’s challenge has the following structure. First, assume that mathematical knowledge satisfies Responsiveness. That is, knowing that ϕ consists of that ϕ is the case, being related through C to ϕ —and perhaps certain unknown factors X .

(1) Mathematical knowledge satisfies some instance of *Responsiveness*.

Second, human beings are spatiotemporal entities without special cognitive powers. Put quaintly, they do not have a dual nature such that one half is part of the realm of abstract objects.

(2) Human beings are spatiotemporal entities.

Third, mathematical objects are abstract objects. If they exist, they are outside space and time.

(3) *Abstractness*.

Moreover, since human beings are physical entities, the properties of mathematical, abstract objects could not be explained by their epistemic capacities. Remember that they are not miraculous. Hence, (2) and (3) support Independence.

(4) *Independence*.

Now, (3)—the abstract nature of mathematical objects—entails that human beings and mathematical objects cannot be physically related. Consequently, Responsiveness cannot be stated by such a relationship: the conditions C will not be physical. Besides, mathematical platonism entails (3).

(5) If mathematical platonism is true, then conditions C in *Responsiveness* do not involve a physical, counterfactual relation.

¹⁵For Benacerraf, Tarski’s definition entails mathematical platonism. However, this is not true. It only implies that the terms are associated with objects, but is silent on the nature of these objects. As I will argue below, mathematical platonism does not entail Tarski’s semantics either.

But, if there is no physical relationship between human beings—who are physical entities—and mathematical objects, what relationship can there be? Plausibly none. (2)—the fact that human beings are physical entities—, (4)—the fact that mathematical objects are independent—, and (5) plausibly entails that there is no relation between human beings and mathematical objects. Consequently, (1) has no true instances. Besides, platonism entails (4).

(6) It is plausible that if mathematical platonism is true, *Responsiveness* has no true instances.

However, it is generally accepted that human beings have mathematical knowledge.

(7) Human beings have mathematical knowledge.

Therefore, (1) and (7) entail that Responsiveness has true instances—i.e., there must be some relation that solves *C* in the “equation.” Consequently, (1), (7) and (6) entail that mathematical platonism is plausibly false.

(8) Plausibly, mathematical platonism is false.

Mathematical platonism cannot explain mathematical knowledge, and this invalidates it as an explanation of mathematics. At least one of its component thesis—Existence, Abstractness, or Independence—must be false.

Usually, the challenge has been presented by adding an additional premise: that every instance of responsiveness replaces *C* with a causal relation. For example, Benacerraf argues that the best theory of knowledge must be a causal theory (CTK). According to this, to know that ϕ , the knowers and the fact that ϕ must be causally related.

(CTK) If subjects *S* know that ϕ , then there is a causal connection between *S* and that ϕ — $C\phi$.

Then, (6) could be deleted and the argument would deductively conclude that platonism is false. However, the argument presented here does not depend on CTK—which has often been questioned. All it requires are solutions for *C* in Responsiveness. As a consequence, deductive power is lost, but a stronger challenge to platonism is obtained. Platonism cannot plausibly explain mathematical knowledge, and this is enough to get it into trouble. Otherwise, (1), (5), (6) and (7) have been formulated in terms of mathematical *knowledge*. This opens the door to possible suspicion. Someone might say that asking for an explanation of knowledge is too demanding. If this is an issue, the premises could be formulated in terms of *justified belief*. Simply reformulate Responsiveness by removing the unknown “*X*” and replacing the “*K*” operator with another operator expressing justified belief (79).

2.3.2.1 Reinforcing the Challenge

The challenge can be transformed into an even stronger challenge if in addition to asking for an explanation of knowledge—or justified belief—one asks for an explanation of the reliability of mathematicians. This is what Field (40, pp.25-30, 230-239) and Linnebo (79) do. Mathematicians are experts in their field and are primarily responsible for mathematical knowledge. Consequently, their judgements and

statements concerning this field must be reliable. But, if this is the case, we need an explanation of their reliability —as an explanation of the reliability of experts in any field is needed. Let a *reliability claim* be a conditional of the following form:

Reliability Claim

$$\forall \varphi (\text{Mathematicians accept } \varphi \rightarrow \varphi \text{ is true}),$$

where φ is a mathematical sentence, and “ $\rightarrow$ ” at least expresses a material conditional. Now, replace (1) in the above argument with the claim that mathematicians are reliable —concerning mathematics:

(1*) Mathematicians satisfy the reliability claim.

This new premise puts explanatory pressure on platonism. If the mathematicians are reliable, it must be explained why this is so. Even if we do not demand an explanation of mathematical knowledge —or justified belief—, we do need to explain what connection exists between mathematicians and mathematics.

(7*) An explanation of the reliability claim is required.

Then, the challenger uses the argument supporting (6) to plausibly deny that there is any relationship between mathematical facts and mathematicians —and, *a fortiori* to deny that mathematical platonism satisfies Responsiveness. Note that this argument does not use Responsiveness, only Abstractness and Independence —i.e. mathematical platonism.

(6*) It is plausible that if mathematical platonism is true, then there is no connection between mathematicians and mathematical facts.

But, in the absence of this connection, the challenger concludes that plausibly Platonism cannot provide any explanation of the reliability claim. Therefore, from (6*) and (7*) she concludes (8): mathematical platonism is plausibly false. This is enough to put platonism in trouble. At the very least it is enough not to claim that Platonism is true. Furthermore, this challenge is stronger than the original Benacerraf challenge. For, it does not involve either knowledge or justified belief. (7*) only asks for an explanation of the reliability of mathematicians. Furthermore, an explanation of mathematical knowledge —or justified belief— does not immediately entail an explanation of (1*). Consequently, it is possible to respond to the first challenge without responding to the second.

2.3.3 Pluralist Reactions to the Challenge

A successful account of mathematics must meet this explanatory pressure. First, it must explain how mathematical knowledge —at least it must explain how mathematical justified belief— is attained. Second, it must explain why mathematicians are reliable when it comes to mathematics. The previous challenge suggests that platonism cannot fulfil this requirement, resulting in a highly unsatisfactory approach to mathematical phenomena. There are two options in this situation: try to save platonism, or abandon it in search of a better explanation. The second option has been the most widely accepted, which has led to the rejection of some of the three theses that compose mathematical platonism. The result of this

rejection is a plethora of sceptic philosophies of mathematics. Commonly, what these philosophies reject is the existence of mathematical, abstract objects¹⁶ —i.e. they reject Existence. In doing this, they try to show one of the following two things. First, Responsiveness is not acceptable in the case of mathematics —thus, they not only reject platonism, but also the premise (1) in the above argument. Alternatively, they claim that mathematical statements are not about what they seem —mathematical objects—, but about very different matters. Then, if this shift in the content of mathematical statements is the appropriate one, it would ensure that Responsiveness has true instances in the case of mathematics.

When nothing else is added to rescue the truth of mathematical theories, the rejection of Existence leads to some kind of error-theoretic position according to which even the most elementary arithmetical statements are false. After all, even Peano-Dedekind arithmetic includes among its axioms the existence of an initial element concerning the successor operation: the number 0. For example, this is the case of fictionalism. Fictionalism considers that mathematical theories are useful but false fictions. Thus, by denying the existence of mathematical objects, fictionalism —and the other error-theoretic positions— aim to avoid Responsiveness. Mathematical knowledge would be similar to fictional knowledge. But, knowledge of fiction does not require that what the fiction describes is the case, nor does it require any connection *C* with these facts. By definition, fictions are false, and what they describe is not the case. For example, the events described in *Hamlet* are false. And to *know the plot* of this play it is not necessary that Hamlet kills Laertes, nor that the readers *know facts* similar to these. Moreover, since we have good explanations of knowledge of fiction, mathematical knowledge could be explained in the same way.

Importantly, fictionalism considers that the negation of Responsiveness entails the negation of a strong concept of mathematical objectivity. Therefore, fictionalism is a pluralistic conception of mathematics. Just as there is no single correct fiction involving empirical objects —such as Hamlet, boats or castles—, there is no single correct fiction about mathematical, abstract objects. In addition to the theories traditionally accepted by mathematicians —for example, the arithmetic we all learn in school —, there are many other deviant theories, which are acceptable from a metaphysical and epistemological point of view. There are theories where the largest natural number exists, or where addition is not commutative. The only thing that fiction must satisfy is consistency —after all, what is at stake is an explanation of classical mathematics. The objectivity of certain mathematical theories would be reduced to the fact that they have been traditionally accepted and studied by mathematicians. For example, basic arithmetic, or mathematical analysis. These would be fictions “taken for true in the canonical history of mathematics”. See Chapter 2 for more on fictionalism. Other positions which also reject Responsiveness are scientific structuralism (57) and meta-ontological deflationism (20). As fictionalism, these positions consider mathematics as a useful cognitive device used by scientist to depict the world or to achieve their epistemic goals —for instance, to obtain accurate predictions. Simplifying a little, the main difference between them and fictionalism is that they are metaontologically neutral: they do not deny or affirm that mathematical objects exist. They simply refrain from pronouncing on this point —thus, they do not accept Existence either.

¹⁶Not all sceptic philosophies of mathematics reject the existence of mathematical objects. As I said in Section 3.1.3, psychologism accepts the existence of mathematical objects when they are understood as psychological constructions.

Alternatively, other sceptical philosophies of mathematics try to save mathematical truth. To do so, they translate mathematical theories into other true theories without demanding the existence of mathematical objects. Often, these translations involve a change of language. For example, this is the case of Hellman's modal structuralism (55), which attributes modal content to mathematical statements. According to it, mathematical statements would be about what is possible or impossible. As a result, the classical quantifiers in them would not play a referential role, and mathematical statements would be true even if mathematical objects do not exist. Consequently, Responsiveness may be safe for some of those proposals involving a translation. If the facts described by the translations bear a relation C to human beings, then Responsiveness will have true instances concerning mathematics —of course, this is doubtful in the case of modal sentences.

What about the first option? Is it possible to rescue platonism from epistemological challenges? The most obvious answer on the part of platonism is to object that Benacerraf's epistemological considerations are biased from the start. Following Quinean holism, he claims that a "reasonable" epistemology must have the form of a global theory of knowledge. "An account of knowledge that *seems* to work for certain empirical propositions about medium-sized physical objects but which fails to account for more theoretical knowledge is unsatisfactory —not only because it is incomplete, but because it may be incorrect as well, even as an account of the things it seems to cover quite adequately" [*Italics in the original*] (7, p. 662). Therefore, if this is correct, and some portion of human knowledge satisfies Responsiveness —for instance, perceptual knowledge—, then mathematical knowledge will satisfy Responsiveness too:

If subjects S know that φ , then there is a solution for C in *Responsiveness* concerning $K\varphi$,

where φ is a mathematical sentence and C is the same for instances of Responsiveness involving empirical knowledge. Nevertheless, this situation assimilates mathematical knowledge to empirical knowledge. And, this is highly doubtful. As I said, Benacerraf's reasons to endorse such a strong thesis are Quinean in spirit, appealing to the "interdependence of our knowledge in different areas" (7, p. 662). But, holism can be resisted. The point is that instead of accepting holism, its consequences —if it implies an assimilation of mathematical knowledge to empirical knowledge— can be taken as evidence against it. To claim that mathematical knowledge involves the same kind of connection between human beings and mathematical facts as empirical knowledge does is unjustified. Even if it is true that empirical and formal sciences are interwoven and can not be clearly demarcated, there is room to observe differences between them. Mathematical methods and standards of correctness as well as mathematicians' practice strongly diverge from those present in physics or biology.¹⁷ Therefore, to claim that mathematical explanations of knowledge —or justified belief— must fit the standards of a global theory of knowledge, and thus Responsiveness, is biased from the start and filled with prejudice. Linnebo (79, p. 546) has dubbed this answer *the Natural Response*.

The Natural Response comes as a reasonable answer to the epistemological challenges. However, this is not enough. At best, it gives new hope for platonism to continue the search for an answer to these

¹⁷See (94, Ch. 2-3) for an analysis of the difference between the methodology of empirical science and the methodology of set theory. Also, see (35) for a broad exposition of the difference between the methodologies involved in different areas of mathematics.

challenges. It does not immediately provide us with such an answer. Indeed, what remains to be done is the most difficult: to provide an acceptable explanation of mathematical knowledge compatible with platonism. Surely, this requires the rejection of Responsiveness —the main feature of empirical knowledge. Surprisingly, when this has been done, the result has involved substantial revisions of traditional platonism. As fictionalists, advocates of mathematical realism argue that the negation of Responsiveness entails the abandonment of a strong form of mathematical objectivity. Platonism becomes a pluralistic philosophy of mathematics, adopting the form of Balaguer’s full-blooded platonism (FBP) (2) or Hamkins’ multiversism (54). Consequently, both anti-realist and realist positions that have survived the epistemological challenges share the same characteristic: essentially, they are pluralist philosophies that deny a strong notion of mathematical objectivity.

In chapters 3 and 4 I will analyse the main representatives of these two major positions. In chapter 3 I will analyse fictionalism. Then, in chapter 4 I will deal with FBP and multiversism. Using an eminently semantic conception of language, I will argue that none of these provide us with a satisfactory explanation of mathematical knowledge. Indeed, I will argue that their very formulation requires that speakers have a minimum of prior mathematical knowledge. Consequently, contrary to what they say, a minimum “amount” of mathematical objectivity must be compatible with the rejection of Responsiveness. Finally, in chapters 5, 6 and 7 I will explain how this is the case and how human beings acquire enough mathematical knowledge to elaborate serious mathematical theories.

2.4 WHAT MATHEMATICAL PLATONISM IS NOT

In the previous section, I explained what mathematical platonism is. Mathematical platonism is the conjunction of three theses: Existence, Abstractness and Independence. In this section, I will explain what mathematical platonism *is not*. Due to its non-sceptical status, platonism has been associated with more substantial metaphysical or semantic theses on several occasions. I will expose the two main groups of ideas that were charged to platonism: metaphysicalism and the semantic theory of truth. I will analyze which assumptions they rely on and which of those assumptions are doubtful. Then, I will argue that none of these assumptions are among the defining theses of platonism. As a result, it will be clear that it is innocent of many of the accusations made against it.

2.4.1 Metaphysicalism

First, mathematical platonism must not be confounded with metaphysicalism.¹⁸ *Metaphysicalism* is the conjunction of the following three theses:

Fundamentality The world’s metaphysical structure *uniquely* carves it up into its constituent parts: objects, properties and relations.

Structure The world’s metaphysical structure is independent of epistemic factors.

¹⁸I take this label from (116, p. 5). The core ideas of *Metaphysicalism* were exposed and criticized by several philosophers under different headings. For example, (112), (94), (60), (82, p. 31).

Correspondence It analytically follows from the concept of truth that an atomic sentence $P\tau_1 \dots \tau_n$ is true if and only if:

- (i) the term τ_1 refers to the object a_1 , and..., and the term τ_n refers to the object a_n ,
- (ii) the predicate P expresses the relation R ,

and

- (iii) it is the case that $Fa_1 \dots a_n$.

I have restricted Correspondence to atomic formulas for the sake of simplicity. This thesis can be extended to complex formulas in an obvious way by employing Tarski's recursive clauses. Alternatively, it can be postulated inside it that each complex formula corresponds to a unique complex fact.¹⁹

Metaphysicalism is a cousin of Russell's and Wittgenstein's Logical Atomism. It provides us with an appealing and simplistic explanation of how language's content, truth and beliefs relate to each other: the truth of sentences and the truth of our beliefs would be a function of the world's metaphysical structure. This explanation, simplistic as it is, was qualified as a metaphysical fantasy.

The metaphysical fantasy is that there is a totality of Forms, or Universals, or "properties", fixed once and for all, and that every possible meaning of a word corresponds to one of these Forms or Universals or properties. The structure of all possible thoughts is fixed in advance —fixed by the Forms. (113, p. 6)

Metaphysicalism is a dubious philosophical conception. It involves very strong metaphysical and semantic concepts. For example, the metaphysical concept of metaphysical structure. Such a structure would articulate the world in a series of basic components. Or the semantic concept of correspondence. According to this, the meaning and truth of sentences would consist in conforming to this structure. Sentences that conform correctly would be true, and those that do not would be false. None of these concepts is part of platonism. In the following I will discuss what each of the three theses composing Metaphysicalism consists of, showing that they are orthogonal to the latter. Consequently, Metaphysicalism can be abandoned as bad metaphysics and bad philosophy of language, while platonism is preserved.

2.4.1.1 Fundamentality and Structure

Fundamentality involves the idea of a *metaphysical structure* that carves the world up in a unique, specific way. According to this, the world would be composed of simple items —commonly, objects, properties,

¹⁹Otherwise, I am assuming that logical form can be read directly from surface syntax. This is not a trivial assumption. Logical forms are theoretical entities. They are extremely useful in formal semantics —for instance, enabling a precise definition of logical entailment and related concepts, such as validity, consistency, completeness and the like. However, it is doubtful that logical forms play any substantive role in communication or in truth semantic clauses for natural language (60), (115, pp. 7-8). For the sake of expository simplicity, I will accept such entities, as well as the fact that they correspond to the surface syntax.

relations, categories, etc.— arranged in complexes which compose this structure.²⁰ For example, according to Logical Atomism, objects and properties as well as objects and relations combine to form atomic facts. And, atomic facts combine between them to form complex facts. The exact components of the world and the operations between them vary depending on the particular metaphysical proposal. The important point is that platonism is not committed to any metaphysical conception of this kind.

First, metaphysical structures are not mathematical structures. They are not mathematical objects. Especially, they are not ordered tuples, as the “structures” used in model-theoretic definitions of truth. Sometimes, platonism is said to be committed to the existence of a mathematical structure. Loosely speaking, *the true* mathematical structure.

What mathematicians standardly mean when they say that a sentence is true is that it is true in the *standard model*, or the *intended structure* —or as we’ll see, the class of intended structures— for the given branch of mathematics. [Italics in the original] (2, p. 60)

For example, according to this view, set theory would be a description of *the* universe of sets. The structure where all sets live. Similarly, Peano-Dedekind arithmetic would be a description of *the* sequence of natural numbers. However, this is just an elaborate way of saying something simpler: that platonism accepts a robust form of objectivity. Among all the theories stated in set-theoretical language, there is an exclusive, maximally consistent set of sentences that provides us with knowledge about these objects: the true set theory, or the true number theory. Generally, when someone speaks about the standard or intended structure of a given mathematical theory, this is only an indirect form of referring to the set of true sentences of its language. No metaphysical concept is involved here. Platonism is not immediately committed to metaphysical structures —only set theoretic platonism is committed to mathematical structures, in the sense of ordered tuples.²¹

Indeed, the status of metaphysical structures is unclear. What does it mean that the world —whatever this is— has a structure? Presumably, metaphysical structures are not architectural entities, such as buildings or bridges —perhaps metaphysicians can be considered architects of the world? For example, if we speak about “the existence” of such structures, we will be quantifying over them. Consequently, it is reasonable to think that structures are objects. Even if quantification is devoid of ontological commitments —i.e. it does not entail existence—, what is quantified acquires the status of an object —see (15; 17) for a detailed argument. But, if this is the case, it is doubtful that metaphysical structures can perform the task they are supposed to do. Allegedly, the world’s metaphysical structure carves it up into its constituent parts. Among them are the objects inhabiting the world. Now, suppose that the metaphysical structure of the world is also an object. If it is an object, it must be part of the world too. Therefore, we are faced with a dilemma. Either the structure explains itself, or we need a second structure to explain the ‘position’ that the first structure occupies in the world. Now, the second structure will be an object too. Therefore, the second option leads us to an infinite regress. As for the first, it implies that the structure explains itself.

²⁰I will put in brackets here any criticism on the already troublesome expression “the world”, as well as the gloss that it comes with parts.

²¹It could be retorted that *ante rem* structuralism is a form of platonism committed to metaphysical structures. However, this is not synonymous with mathematical platonism. *An rem* structuralism arises when the commitment to structures is added to the latter.

Thus, the two positions imply a form of circularity. And, although not every form of circularity is pernicious, it is pernicious if the relation is explanatory. And, in this case, it is. Consequently, it is plausible that structures are not objects—and cannot be quantified—or they do not perform the explanatory role desired by the metaphysician. But, if they are not objects, what are they? And, if they do not play an indispensable explanatory role, why should they be accepted? As Isaacson puts it:

What are particular mathematical structures? Despite asserting e.g. that the natural numbers constitute a particular mathematical structure, I also say that the structure of the natural numbers (in the sense under discussion) is not a mathematical object nor indeed an object of any sort. In that sense talk of ‘the structure of the natural numbers’ is a *façon de parler*. In discussion Stewart Shapiro offered the suggestion that this construal of particular mathematical structures is a form of *if-then-ism*. [Italics in the original] (65, p. 31)

Second, none of the three theses composing mathematical platonism—Existence, Abstractness and Independence—entails Fundamentality. Fundamentality entails the existence of those objects that are part of the structure of the world. Even if existence is thought of as a second-order concept, not as a property carved up into the world, this concept takes into account the objects and properties that are part of the world’s structure. Indeed, it is plausible that Metaphysicalism supports the following connection between existence and structure: an object exists if and only if it is a constituent of the world. Consequently, if mathematical objects are part of the world, Fundamentality—modulo the previous, reasonable assumption—entails Existence. However, the opposite is not the case. Existence does not entail Fundamentality. That some kinds of objects, such as mathematical objects, exist, by no means entails a position concerning the metaphysical structure of the world. In particular, the description of the existing objects remains the same with and without metaphysical structure. Again, structures should be taken as a *façon de parler*, that is, as an abbreviation of a more fundamental idea. This idea would be that there is a unique correct description of objects and how they relate to each other.

Finally, the epistemological independence of the world’s structure—i.e. Structure—does not follow from platonism either. As has just been seen, the commitment to this metaphysical concept does not follow from platonism. So, neither does any particular claim involving it. The thesis Structure is similar to Independence—which is part of platonism. Both assert that some kind of entity and facts involving such entities are epistemically independent of agents. Besides, as in the case of Fundamentality and Existence, Structure entails the epistemological independence of the objects, properties, relations, etc. which compose the world’s structure. If this is epistemologically independent from agents, so will be the former. Consequently, if mathematical objects are part of this structure, Structure entails Independence. But, the opposite—once more—is false. Platonism is about objects, not about metaphysical structures.

2.4.1.2 Correspondence

Correspondence constitutes a substantial account of meaning and truth. Once the world’s structure has done its job and objects, properties and the like are on the table, agents will use adequate representational devices to depict them—or at least the metaphysicalist thinks so. As a result, Correspondence comes as an elaboration of the metasemantic “picture theory” advocated by Wittgenstein in the *Tractatus*. Linguistic

components are correlated with components of the structure. Terms are correlated with objects, and predicates are correlated with properties and relations. Then, the semantic content of every expression is explained by using this correlation.

Therefore, Correspondence is intended as a metaphysical elaboration of semantic definitions of truth, such as Tarski's (140) model-theoretic definition. According to semantic definitions, truth is a substantive concept which is defined using additional semantic concepts. In the case of Tarski's definition, the concepts of reference and satisfaction. Instead of satisfaction, Correspondence relies on the metaphysical notion of the world's carving. Thus, according to it, $P\tau_1 \dots \tau_n$ is true if and only if the world's metaphysical structure carves it up in such a way that $Fa_1 \dots a_n$ is the case. Semantic definitions of truth and other substantive definitions—for example, causal conceptions of truth—, have been criticized on several occasions by those who defend a *minimal* or *deflationist* theory of truth (151; 64). Every theory of truth considers that the T-scheme

$$\varphi \text{ is true iff } \varphi$$

follows analytically from this concept. According to the proponents of the minimal theory, this would be the only thing that follows analytically from the concept of truth. Consequently, a complete definition of truth reduces to the T-scheme. Truth is extensionally specified using a recursive and contextual definition—see 2.6 for some comments on minimalism. Therefore, if platonism entails Correspondence or a semantic theory of truth, the criticism posed by minimalists on the latter would transitively apply to the former.

However, mathematical platonism does not entail any particular theory of truth. As I will argue in the next section, it is consistent with the minimal theory truth, with Tarski's definition, or with some version of Correspondence. After all, platonism is a position about what kinds of objects there are—mathematical abstract objects—, not about how truth must be understood. All this may seem obvious, but Platonism has often been assimilated into semantic theories of truth. The reason is that it accepts the following two claims: mathematical objects exist and mathematical terms refer to these objects. Now, if the claim that mathematical terms refer to mathematical objects is loaded into Correspondence, this entails that mathematical theories are true. But, once more, the opposite is false. If mathematical objects exist, mathematical terms refer to these objects, and mathematical theories are true—i.e. if platonism is true—, Correspondence does not follow. The assumptions say nothing about what human beings mean when they speak about truth. Even if mathematical terms are referentially successful, it is not immediate that truth should be defined through the notion of reference. Platonism does not entail Correspondence.

The present discussion shows the following. Assume that mathematical objects exist and that mathematical terms refer to them. Then, Metaphysicalism entails platonism:

$$\text{Existence} + \text{Reference} + \text{Metaphysicalism} \rightarrow \text{platonism},$$

where “Reference” abbreviates the fact that mathematical terms refer to mathematical objects. This may suggest a close connection between both positions. However, the connection is exhausted by the previous conditional. Against appearances, the opposite conditional is false. Platonism does not entail Metaphysicalism:

platonism $\leftrightarrow$ Metaphysicalism.

Indeed, it does entail none of its three theses —Fundamentality, Structure, and Correspondence. These three theses are highly controversial. They were criticized on several occasions. Fortunately, platonism is not committed to any of them. Consequently, mathematical platonism can be accepted while these are abandoned.

2.4.2 Theories of Truth

As I said in the previous section, platonism was identified with semantic theories of truth on several occasions. Therefore, I will insist on this point here: platonism does say nothing about the content of the truth predicate or about what truth is. One of the clearest examples is Benacerraf, who endorses platonism with a commitment to Tarski's model-theoretic definition of truth.

I call the "platonistic" account that analyzes (2) ["There are at least three perfect numbers greater than 17"] as being of the form (3) ["There are at least three FG's that bear R to a"] "the standard view". [...] One of its primary advantages is that the truth definitions for individual mathematical theories thus construed will have the same recursion clauses as those employed for their less lofty empirical cousins. (7, 668-668)

Moreover, he adds: "I take it that we have only one such account: Tarski's, and that its essential feature is to define truth in terms of reference (or satisfaction)" (7, 667). Another example of this tendency is Dummett, who identifies platonism with classical mathematics, and both with the semantic conception:

On a platonistic interpretation of a mathematical theory, the central notion is that of truth: a grasp of the meaning of a sentence of the language of the theory consists in a knowledge of what it is for that sentence to be true. [...] On the theory of meaning which underlies platonism, an individual's grasp of the meaning of such a sentence consists in his knowledge of what the condition is which has to obtain for the sentence to be true, even though the condition is one which he cannot, in general, recognise as obtaining when it does obtain. (30, 14-15)

Dummett uses this identification to build up two arguments against Tarski's definition and, transitively, against Platonism.

2.4.2.1 The Infinite Regress Challenge

The first argument is the Infinite Regress Challenge (IRC). Let's assume platonism. Also, assume that Tarski's semantic definition is also an explanation of speakers' linguistic competence —i.e. an explanation of their understanding of language. Although Dummett does not do so, this is tantamount to accepting Davidson's (26; 27) truth-conditional semantics.

(4.i) Accept Davidson's truth-conditional semantics.

According to this, knowledge of a sentence's meaning consists of the speakers' ability to state its truth condition.²²

(4.ii) Understanding a sentence is tantamount to understanding its truth condition.

Then, he notes that the truth condition of a given sentence involves the same expressive resources as the sentence.

(4.iii) A sentence and its truth condition involve the same expressive resources.

At this point, he notes that (4.ii) and (4.iii) entail an infinite regress. The explanation of how speakers understand the meaning of a sentence reduces to their understanding of its truth condition. But, this truth condition involves the same expressive resources as the sentence. Therefore, the *explanandum* features in the *explanans* —i.e. we use what we want to explain in the explanation. Thus, speakers' understanding of a sentence's truth conditions needs explanation too. But, (4.ii) reduces this explanation to the speakers' understanding of the truth condition of this truth condition. And, (4.iii) requires to repeat this movement again. Consequently, an infinite regress arises

(4.iv) An infinite regress arises.

Dummett uses this to accuse Tarski's semantic definition —i.e. Davidson's truth-conditional semantics— and platonism of being unable to explain the linguistic competence of speakers. Since speakers understand mathematical language, platonism must be false.

2.4.2.2 The Language Acquisition Challenge

The second argument is Dummett's famous Language Acquisition Challenge (LAC) (66). He opposes it against Tarski's definition —i.e. Davidson's truth-conditional semantics— and platonism's ability to explain speakers' linguistic competence. Namely, he claims they do not meet the following maxim, which every such explanation must satisfy:

Dummett's Maxim Use exhaustively determines meaning. (30, p. 14)

The argument goes as follows. Assume platonism. Thus —modulo his identification— assume Tarski's definition of truth —i.e. Davidson's truth-conditional semantics. However, Tarski's semantics equips mathematical language with evident-transcendent truth conditions. There are sentences such that speakers do not know if their truth condition is satisfied or not.

(4.a) Mathematical language has evident-transcendent truth conditions.

To take a classic example, nobody knows if CH is true or false. Now, take one of these sentences φ and —without loss of generality— assume it is true. The assertion of a sentence at least involves the ability to recognize that it is true. Therefore, this sentence will not be assertible.

²²For very different reasons, this is true of the minimalist theory of truth, which considers that " φ is true" is analytically equivalent to φ .

(4.b) ϕ is true and not assertible.

In turn, Dummett's maxim entails that truth is revealed in use. Therefore, if a sentence is true, it must be assertible. That is, all the instances of the following scheme are true:

(4.c) If sentence ψ is true, then ψ is assertible.

However, (4.b) provides a counterexample to (4.c). Consequently, Dummett's maxim and Tarski's semantics are inconsistent. From this Dummett concludes the falsity of the latter. Thus, platonism is false.

On the contrary, he takes as characteristic of the practice of assertion a version of the *KK*-principle:

If ϕ is assertible, then it is assertible that ϕ is assertible.

This means that assertion satisfies his maxim: if a sentence is assertible, speakers recognize it is assertible. Instead of truth conditions, Dummett proposes that meaning should be explained by providing the language with assertibility conditions. Verificationism-friendly philosophers tend to dismiss truth-conditional semantics because it does not pass LAC.

2.4.2.3 Avoiding the Challenges

To insist once more, Platonism does not imply Tarski's definition of truth or any other definition. It is a metaphysical thesis concerning mathematical objects, not a semantic thesis about truth or any other related concepts. Therefore, this simple fact safeguards platonism from criticism posed by IRC and LAC.

Nevertheless, I will show why this criticism does not work against Tarski's semantic definition of truth either. As a result, there will be no doubt that platonism is safe. First, both challenges conclude that semantic conceptions of truth override speakers' linguistic competence. Here is the first problem. Despite the intuitive connection between truth and meaning, Tarski's truth definition does not entail a theory of speakers' linguistic competence. The reduction of the theory of meaning—an intensional notion—to the theory of truth was obtained by Davidson after imposing a collection of specific constraints on any satisfactory theory of meaning.

The concept of truth played no ostensible role in stating our original problem. That problem, upon refinement, led to the view that an adequate theory of meaning must characterize a predicate meeting certain conditions. It was in the nature of a discovery that such a predicate would apply exactly to the true sentences. (26, 310)

Therefore, I take Dummett's arguments as arguments against Davidson's truth-conditional semantics—an extensional approach to meaning—, instead as arguments strictly against Tarski's definition of truth.

Now, I accept Dummett's conclusion: Davidson's truth-conditional semantics—i.e. Tarski's semantic definition of truth—provides a *partial* explanation of speakers' linguistic competence. The point is that this is not evidence against Davidson's theory. The task of this theory is not to perform the explanatory functions demanded by Dummett. So, if it fails on this point, this is something that cannot be held against it.

Let's start with IRC. The regress affects a particular aspect of speakers' linguistic competence: how speakers acquire their competence. There is a difference between explaining *what understanding consists of* and the requirement of explaining *how understanding is achieved*. Therefore, both IRC and LAC lead to the same problem: speakers' acquisition of their linguistic competence. When I say that Davidson's truth-conditional semantics is a partial explanation of speakers' linguistic competence is because this theory does not explain how speakers *acquire* it²³ However, this is so because the theory is not designed to explain how speakers learn to speak. Davidson's semantics states that to know the meaning of a sentence reduces to knowing its truth condition. This is a *partial* —but perfectly adequate— explanation of speakers' competence. And the theory seeks nothing more than this. Furthermore, it can be supplemented with additional explanations. For instance, by adding a use theory of meaning able to deal with language learning. Truth-conditional semantics does not forbid this kind of extension.

Let's turn to LAC. This argument relies on Dummett's maxim: use exhaustively determines meaning. This imposes the condition that truth does not exceed assertion —i.e. it entails (4.c). But, this is —modulo the T-schema— to accept the following version of KP:

$$\varphi \rightarrow A\varphi,$$

where the operator “*K*” —expressing knowledge— is substituted by the operator “*A*”, that expresses assertion. This says that whenever a sentence is true it is assertible, and therefore recognisably true. From this, Dummett concludes that the only way of explaining speakers' linguistic competence is to accept this sort of connection between meaning and epistemic factors. Once this connection is broken —i.e. truth and assertion take different ways—, language learning is not explainable. However, there are a variety of problems with this line of thought.

First, there is a difference between learning a language and knowing the collection of its true sentences —the extension of the truth predicate. Davidson's semantics does not collapse this difference. Truth conditional semantics explains meaning by using truth, but it does not reduce knowledge of the former —i.e. linguistic competence— to knowledge of the latter. What it prescribes is that to understand a sentence is to understand its truth condition: speakers must know which is the truth condition associated with each sentence. But, this is tantamount to requiring knowledge about the consequences of the T-schema. At most, this semantics accepts the previous version of KP to this theory. Now, as far as the T-schema generates a recursively, axiomatized theory, there is no problem in meeting this requirement.

On the contrary, Dummett's maxim does collapse the aforementioned difference between learning a language and knowing the extension of the truth predicate. However, this is an unreasonable demand. Knowledge of the consequences of the T-schema is enough to explain language learning when it is combined —for example— with a use theory of meaning. Consider the following toy model —it is not a serious attempt to explain how speakers acquire their linguistic competence. Let's assume the division between primitive and non-primitive vocabulary for simplicity. At the beginning, a use theory of meaning could explain how speakers learn to use the primitive vocabulary, and how some words —as proper names— are linked to some items in the environment. Think about kids naming some objects. Doing this, the theory will explain speakers' understanding of some simple sentences —i.e. how these sentences

²³Horwich (64) offers the same criticism.

become associated with a particular truth condition—and their understanding of the sentences’ components as simultaneous events. Since Frege and Quine it is clear that sentences are the minimal empirical unity of semantics. At this point, Davidson’s semantics enters the picture. The compositional nature of truth conditions explains how more complex sentences are linked to their corresponding truth conditions. Thus, this rather simplistic—and unrealistic—model shows how Davidson’s semantics is consistent with a plausible theory that explains language learning. Furthermore, since this semantics entails the negation of Dummett’s thesis, the model shows that this thesis must not be imposed as an axiom on explanations of linguistic competence. Truth-conditional semantics has room for answering LAC.

This is enough to dismiss Dummett’s criticism of Davidson’s semantics. However, one question remains to be answered. To what extent is Dummett’s maxim reasonable? To the point that it entails some version of KP, it is at odds with those discourses satisfying Independence. Indeed, if KP is analyzed as the negation of the latter, it would be falsified in those languages. Consequently, the question reduces to this dilemma: does a language satisfy Independence or not? I believe plenty of discourses do—the language of physics, perception, or “pure” mathematics are just a few candidates.²⁴

2.5 VARIETIES OF INDEPENDENCE

In this section, I will say a bit more about Independence. This is the thesis that the content of sentences is not metaphysically explained by appealing to epistemic factors

$$\text{Independence } C\phi \not\models \phi,$$

where C are speakers’ cognitive responses. According to it, the language is not covertly about agents’ epistemological responses. For example, suppose that perceptual discourse has this feature. Then, when somebody says “the table is blue”, her statement is about an object which is not metaphysically explained by her epistemic responses—the table. The statement is not about her sense impressions or her judgments. Consequently, a statement such as “the table is blue” is not synonymous with “it seems that the table is blue” or “I judge that the table is blue”. Epistemic operators come as an extra component when the discourse satisfies Independence.

As I have argued in section 2.3.2, Responsiveness causes Benacerraf’s challenge. The platonist’s most immediate response is to reject this thesis. Mathematical knowledge does not have the same standards as empirical knowledge. Therefore, to demand it is unreasonable. However, the rejection of Responsiveness does not solve the challenge. At most, it opens the door for platonism to find a solution, i.e., a reasonable epistemology of mathematics. But, this task is difficult because platonism entails Independence.

Independence opens a gap between agent’s knowledge-gathering activities and the world—i.e. language’s content. According to Independence, mathematical facts are not explained by the

²⁴For criticism on Dummett’s thesis and KP based on a simple condition see (145; 146). The condition is that predicates of language have a *fadeout*. This is a numerically indexed series of epistemic contexts such that at n it is known that not p , at $m > n + r$, it is known that p and at k such that $n < k < m$ it is not known that p or not p . According to Williamson’s arguments, this feature entails the failure of some instances of KP, and thus the failure of Dummett’s maxim. Moreover, according to him, almost all discourses have fadeouts. Therefore, Dummett’s maxim turns out to be almost globally implausible for linguistic phenomena.

epistemological answers of mathematicians. Thus, human beings do not have a perfect track of them. This characteristic is shared by the empirical sciences. If there is a world beyond us, there is a risk of falling short of knowing it. But, the situation of mathematical objects is even worse. According to Abstractness, these objects are abstract—they are outside space and time. Human beings cannot be physically related to them. Consequently, mathematical knowledge appears as a kind of mystery. If Platonism is true, how is it possible to be aware of such strange objects? Now, the gap proves particularly difficult to close.

In the following sections, I will explain two different dimensions of this gap: an *extensional* and an *intensional* form of independence. The first is based on the epistemic imperfection of agents. The gap between them and the world appears more clearly when there are unknown facts. The second refers to the precariousness of knowledge. Even if science reaches a hypothetical stage of perfection, this would not erase all the steps taken previously, nor would it prevent the possibility of regression. Additionally, 2.6 elaborates on the relation between Independence and objective disagreements. Then, 2.6 points to the hyperintensional nature of the gap.

2.5.1 Metaphors of Independence

Before getting down into business, I would like to say a few words warning against a too demanding way of understanding the idea that mathematical objects are independent from agents—see section 2.3.1.3. It is usual to find in the literature a more substantial gloss of Independence. According to this reading, this hides an analogy between mathematical abstract objects and concrete objects, such as tables, trees or mountains.

Platonism, as a philosophy of mathematics, is founded on a simile: the comparison between the apprehension of mathematical truth to the perception of physical objects, and thus of mathematical reality to the physical universe. For the Platonist, mathematical statements are true or false independently of our knowledge of their truth-values.[p. 202] (30)

The analogy is a natural assumption of those accounts of mathematics that are Quinean in spirit, such as Putnam's (109, p. 65) quasi-empirical conception of mathematics or Benacerraf's holism (7). More examples are found in (93). As a result of this analogy, mathematical platonism—which contains Independence—is, again, endorsed with additional, suspicious theoretical commitments.

As I said in section 2.3.1.3, the analogy is not a qualification made on Independence. Instead, it is better understood as the conjunction of Independence and Responsiveness. According to those who advocate the metaphor, mathematical theories have the same *objective status* as empirical theories. But, as I argued in section 2.2, the objectivity of empirical theories consists in the conjunction of Independence and Responsiveness. Consequently, the physical analogy endorses the mathematical discourse with both thesis.

However, the characterization of mathematical platonism elaborated in section 2.3.1 does not contain Responsiveness. By no means platonism is committed to comparisons between mathematical objectivity and the objectivity of other fields of knowledge. This second thesis—Responsiveness—is an extra

assumption. Although it is consistent with platonism, it is not among its consequences. The result of adding Responsiveness is Mathematical Platonism—with capital letters:

$$\text{Mathematical Platonism} =_{df} \text{mathematical platonism} + \text{Responsiveness}.$$

As I have argued in section A, Responsiveness is precisely what provokes Benacerraf’s challenge. As I will argue in Chapters 5-7, once it is abandoned, there is room to meet the challenge. Thus, I agree with those authors who reject the analogy as a feature of abstract objects. For instance, Linnebo (82, p. 189) claims that his abstract objects are only counterfactually independent. Then, he adds: “It is doubtful, however, that this counterfactual independence suffices for my view to qualify as mathematical Platonism. The independence thesis is often flexed out by comparing mathematical objects with physical objects.” (p. 190)

2.5.2 Extensional Independence

A sentence is *extensionally independent* when it or its negation is true, but speakers do not recognize that this is the case.

Extensional Independence A sentence φ is *extensionally independent* if and only if φ or $\neg\varphi$ is true, but there is no epistemic context C such that $C\varphi$ or $C\neg\varphi$.

When this occurs, there is a gap between the content of language and speakers’ epistemic responses. Truth is not coextensive with knowledge, belief, assertion and other epistemic operators. For example, think about the Milky Way. There are numerous estimates of how many stars there are in our galaxy. However, the exact number of stars is unknown. Allegedly, there is an exact number n of stars in the Milky Way. Therefore, the sentence “The Milky Way has n stars” is extensionally independent, since it is true, but we cannot recognize that this is the case. Extensional independence corresponds to Dummett’s Semantic Realism (30)—i.e. it corresponds to the negation of Dummett’s thesis, see section 2.4.2.2. According to him, evident-transcendent truth conditions are a direct route to a world standing beyond us. “Realism in Dummett’s sense is thus one way of laying the essential semantic groundwork for the idea that our thought aspires to reflect a reality whose character is entirely independent of us and our cognitive operations” (151, p. 4). When there is a gap between the world and epistemology, then we can be sure that the former does not reduce to the latter.

Classical mathematics is full of examples of extensional independence. Elementary mathematical theories, such as Peano-Dedekind arithmetic (PA), are highly incomplete. Let “ $\vdash$ ” abbreviate deducibility. If PA is consistent, then there are many arithmetical sentences φ such that $PA \not\vdash \varphi$ and $PA \not\vdash \neg\varphi$. The most iconic cases are Gödel’s sentence—a toy sentence explicitly designed to show PA’s incompleteness—and Con(PA). More natural examples are Goodstein and Kruskal’s theorems. Therefore, if our mathematical knowledge were restricted to PA’s consequences, and PA were true and consistent, all those sentences would be extensionally independent.

It is doubtful that our mathematical knowledge is so weak. Most cases used to show PA's limitations exemplify a *benign* form of incompleteness (68, p. 36). Specifically, all the previous cases are Π_1^0 -sentences in Levy's hierarchy.²⁵ These sentences are true if and only if adding them to PA results in a consistent theory.²⁶

For all Π_1^0 -sentences φ : φ if and only if $\text{Con}(\text{PA} \cup \{\varphi\})$.

Therefore, assuming knowledge of PA's consistency, we know that $\text{PA} \cup \{\varphi\}$ is consistent as soon as φ is shown to be logically independent from PA. Surprisingly, Π_1^0 -sentences offer themselves as true once they are shown to exceed PA. This is the case of the famous Gödel's sentence and $\text{Con}(\text{PA})$. The real exciting cases of incompleteness become at the level of Σ_1^2 -sentences such as CH, which so far remain undecided under natural axioms.

2.5.3 Counterfactual Independence

A sentence is *counterfactually independent* when it is possible that it or its negation is true, while speakers do not recognize that this is the case. Therefore, a sentence is counterfactually independent when it is possible that it is extensionally independent. This simple concept aims to capture a strong idea. Imagine an ideal epistemic situation where no sentence was unknown —i.e. where extensional independence had no counterexamples. Then the facts described by the language would still be metaphysically independent of any epistemic factor. Although in this situation there is no extensional gap, there is still a modal gap between epistemology and the world.

Counterfactual Independence A sentence φ is *counterfactually independent* if and only if $\Diamond(\varphi$ is extensionally independent).

Consider the Sun. The Sun existed long before there were living beings. Now, consider the ideal situation where physics has achieved a perfect description of the universe. In this situation, the sentence “The Sun is a star” is not a counterexample to extensional independence —the Sun is a star, and speakers know that. However, it would be counterfactually independent. Because, before living beings existed, the Sun

²⁵Levy's hierarchy is a classification of formulas defined as follows.

A sentence φ is Σ_n^β if and only if $\varphi \equiv \exists x_1 \forall x_2 \dots Qx_n \psi$,

where Q is $\exists$ if n is odd, Q is $\forall$ if n is even, and quantifiers and terms are of order equal or less than β .

A sentence φ is Π_n^β if and only if it is the negation of a Σ_n^β -sentence.

All sentences of classical languages of a given order are equivalent to a Σ_n^β or a Π_n^β -sentence. Here, n -order languages are interpreted as quantifying over a particular numerical system. First-order languages —as the language of PA— quantify over objects, which are elements of order 0. Second-order languages quantify also over reals —i.e. subsets of naturals. Third-order languages quantify also over sets of reals, etc. Then, the order in Levy's classification indicates which entities are quantified over. For example, Gödel's sentence and $\text{Con}(\text{PA})$ are Π_1^0 -sentences —they have a universal quantifier and all their terms and quantifiers have order 0. In turn, CH is a Σ_1^2 -sentence.

²⁶See (129, p. 80) for a proof.

was already a star, but nobody knew it. From the previous characterization, it immediately follows that extensional independence entails counterfactual independence. However, the opposite is not true.

Counterfactual independence is a highly plausible case of Independence. Even if the whole physical history of the universe, its past, present and future, were known —and there were no single case of extensional independence—, the world would not become part of the mind —to put it quaintly. Realism and objectivity would not break down as senseless notions from a past, imperfect era. Just as they do not break down if our present knowledge is limited to the complete knowledge of the room in front of me at this moment. Just as it would not be the same if we limited our present knowledge to the complete knowledge of the room in front of me at this moment. As in this simple example, fiction presents us with a case of knowledge fixation, and there was a time when knowledge was not so perfect.

Counterfactual independence can be modelled in a simple way using possible world semantics. The content of sentences can be modelled as a set of metaphysical possible worlds. Then, if a sentence is counterfactually independent, its corresponding set of possible worlds will not be a subset of those corresponding to sentences involving epistemic operators. As a result, “ ϕ is true” is not necessarily equivalent to “ $C\phi$ ”, where C is some epistemic context.

Both extensional independence and counterfactual independence are cases of Independence. Consequently, platonism implies that mathematics is independent in these two senses. If platonism is true, not every mathematical sentence is recognizable as true. Moreover, even if the ideal situation were mathematical ignorance disappeared, the present imperfect situation would constitute a case of counterfactual independence. Independence, together with the failure of Responsiveness, is what causes the epistemic concerns affecting mathematical platonism. For, the first entails a gap between mathematical facts and our knowledge-gathering practices. In turn, the second leaves us at a loss as to how to bridge this gap. The proposals I will discuss in chapters 3 and 4 deny that there is a gap to be closed —fictionalism, full-blooded platonism, and multiversism deny that a substitute for Responsiveness has to be found. This is so because they equate Responsiveness and mathematical objectivity, giving up on both. On the contrary, in chapters 5 to 7 I will argue that, if there is a gap, —despite the difficulties— it can be closed. This is possible because mathematical objectivity and Responsiveness are not equivalent.

2.6 CONCLUSION

In this chapter, I have analyzed mathematical platonism and the main challenge it faces, Benacerraf’s challenge. Mathematical platonism is the philosophical thesis that abstract objects exist independently from us. According to it, mathematics is the science that studies those objects. In turn, Benacerraf’s challenge claims that if platonism is true, then mathematical knowledge cannot be explained. The challenge works if one accepts the requirement that mathematical knowledge meets the standards of empirical knowledge. In particular, Responsiveness: the thesis that objective knowledge involves a strong, counterfactual relation between agents and the world. However, this requirement is not reasonable. Mathematical knowledge does not obey the same rules as the natural sciences. If we reject Responsiveness there is room to explain mathematical knowledge and to save platonism.

Additionally, I explained what platonism is not: it does not involve a thesis about the metaphysical

structure of the world or a theory of truth. Finally, I have explained what means that some objects are independent of us. Specifically, I have distinguished between the two main forms of independence: extensional independence and counterfactual independence.

APPENDIX 2.A

Since Tarski's seminal (140), any reasonable theory of truth is required to imply all instances of the T-schema

φ is true if and only if φ .

According to it, a formula φ is true if and only φ . It supplies each sentence with a condition to know if it is or not in the extension of the truth predicate. The *minimal theory* of truth (MT) (64), (150) imposes an additional —minimal— requirement: ask for nothing more than the T-Schema. According to minimalists, the T-schema is a contextual definition of the truth predicate. It is all that is demanded by the speakers' use of this predicate. The latter comes with an opportune referential apparatus that allows them to economically make a collection of assertions in a single move —as in “all theorems of PA are true”—, as well as to assert something when homophonic resources are not available —as in “whatever Austin said is true”. In consequence, the content of the truth predicate is completely explained by the T-schema. It is analytically true, and nothing more is.

Here I will show that a minimal theory of facts (TF), which captures the basic features of this concept, is equivalent to MT. Consequently, a minimal concept of fact does not add anything over and above the latter. Further metaphysical elaborations of this concept can give rise to a more substantive concept of fact, which goes beyond the discourse on truth —understood in a minimal sense as well. I will show that TF is coextensional —in the sense explained below— to MT. This is enough to ensure that the use of “is a fact” collapses into the use of “is true”. Minimal facts do not add new explanations concerning minimal truth —explanations which are lost by MT.

Let $\mathcal{L}$ be a language and let “ $Sen_{\mathcal{L}}(\varphi)$ ” abbreviate “ φ is a sentence of $\mathcal{L}$ ”. Let T be a monadic predicate obeying MT. Let $MT_{\mathcal{L}}$ be the theory that results from restricting the instances of MT to $\mathcal{L}$. Let $\Gamma = \{\ulcorner \varphi \urcorner / Sen_{\mathcal{L}}(\varphi)\}$ be a set of names for the formulas of $\mathcal{L}$. Then, the language of $MT_{\mathcal{L}}$ is $\mathcal{L}^* = \mathcal{L} \cup \{T\} \cup \Gamma$. Now, let's characterize the language and axioms of TF. To provide false sentences with content, the theory must distinguish between positive and negative facts, or mere possible facts and plain facts. To keep things as simple as possible and avoid an obvious objection —namely, that the notion of fact is explained through modal notions— I will differentiate between Russellian propositions and facts. Thus, let Δ be a set of individual constants denoting facts —“ p ”, “ q ”,... will be employed as meta-variables for these terms. Then, “ $R_{\mathcal{L}}(p)$ ” abbreviates “ p is a Russellian proposition”, and “ $F_{\mathcal{L}}(p)$ ” abbreviates “ p is a fact”. Finally, let's introduce a binary predicate “ C ” holding in $\{\varphi / Sen_{\mathcal{L}}(\varphi)\} \times \mathcal{L}$: “ $C(\varphi, p)$ ” stands for²⁷ “ φ corresponds to p ”. The language of $TF_{\mathcal{L}}$ is $\mathcal{L}^+ = \mathcal{L}^* \cup \Delta \cup \{R_{\mathcal{L}}, F_{\mathcal{L}}, C\}$.

²⁷The notion of *correspondence* can be understood in terms of structure. However, this is not necessary.

Note that $\mathcal{L}^+$ is an extension of $\mathcal{L}^*$, the language of $\text{MT}_{\mathcal{L}}$. Both theories include names for formulas²⁸ in $\text{Sen}_{\mathcal{L}}$ and a truth predicate T . Also, given that the notion of well-formed formula and —presumably— the notion of complex fact are recursive, “ $C(\varphi, p)$ ” is recursive too. Finally, the axioms of $\text{TF}_{\mathcal{L}}$ are all the instances of the following:

- i For all $\ulcorner \varphi \urcorner$, there is a $p, R_{\mathcal{L}}(p)$ such that $C(\ulcorner \varphi \urcorner, p)$
- ii For all $\ulcorner \varphi \urcorner, p, q$: if $R_{\mathcal{L}}(p)$ and $R_{\mathcal{L}}(q)$, then if $C(\ulcorner \varphi \urcorner, p)$ and $C(\ulcorner \varphi \urcorner, q)$, then $p = q$
- iii For all $p, R_{\mathcal{L}}(p)$
- iv For all $\ulcorner \varphi \urcorner, p$: if $R_{\mathcal{L}}(p)$ and $C(\ulcorner \varphi \urcorner, p)$, then: $F_{\mathcal{L}}(p)$ iff φ
- v For all $\ulcorner \varphi \urcorner$: $T(\ulcorner \varphi \urcorner)$ iff there is some p such that $R_{\mathcal{L}}(p)$, $C(\ulcorner \varphi \urcorner, p)$, and $F_{\mathcal{L}}(p)$.

Axioms i-v prove all consequences of $\text{MT}_{\mathcal{L}}$. This is an essential requirement, provided that $\text{TF}_{\mathcal{L}}$ aims to explain the behaviour of the truth predicate in $\mathcal{L}$. Now I can state in what sense $\text{MT}_{\mathcal{L}}$ and $\text{TF}_{\mathcal{L}}$ are coextensional. First, $\text{TF}_{\mathcal{L}}$ cannot prove anything new about the truth in $\mathcal{L}$: it is conservative over $\text{MT}_{\mathcal{L}}$.

Theorem 1. *$\text{TF}_{\mathcal{L}}$ is conservative over $\text{MT}_{\mathcal{L}}$.*

Proof. I will employ completeness to prove $\{\psi / \text{Sen}(\mathcal{L}^*)(\psi) \wedge \text{TF}_{\mathcal{L}} \vdash \psi\} \subseteq \{\psi / \text{MT}_{\mathcal{L}} \vdash \psi\}$ for first order $\mathcal{L}$. The result can be extended to higher-order languages by induction on the complexity of formulas. Let $\mathcal{M}$ be a model of $\text{MT}_{\mathcal{L}}$. $\mathcal{M}$ can be extended to a model $\mathcal{M}^+$ of $\text{TF}_{\mathcal{L}}$: its domain $\mathbf{M}^+$ would be $\mathbf{M} \cup \mathbf{R}_{\mathcal{L}}$, and predicates $F_{\mathcal{L}}$ and C would be interpreted in $\mathcal{P}(\mathbf{R}_{\mathcal{L}})$ and $\mathbf{M} \times \mathbf{R}_{\mathcal{L}}$ according to Axioms ii and iii. Note that all $\mathcal{M}^+$ is elementarily equivalent to $\mathcal{M}$ relatively to $\text{Sen}(\mathcal{L}^*)$. Assume $\varphi \in \text{Sen}(\mathcal{L}^*)$ and $\text{TF}_{\mathcal{L}} \vdash \varphi$. By correctness, for all $\mathcal{M}^+ \models \text{TF}_{\mathcal{L}}$, $\mathcal{M}^+ \models \varphi$. By elementary equivalence, for all $\mathcal{M} \models \text{MT}_{\mathcal{L}}$, $\mathcal{M} \models \varphi$. Finally, by completeness, $\text{MT}_{\mathcal{L}} \vdash \varphi$. $\square$

This result, interesting as it is, should not be a major surprise. After all, $\text{TF}_{\mathcal{L}}$ offers a metaphysical explanation of truth, it is not a revisionist theory. The use of T provides the empirical data, while $\text{TF}_{\mathcal{L}}$ and $\text{MT}_{\mathcal{L}}$ try to explain data. Therefore, does Theorem 1 pose any problem on $\text{TF}_{\mathcal{L}}$? Moreover, despite $\text{TF}_{\mathcal{L}}$ does not prove anything new in $\mathcal{L}^*$, it incorporates a whole theory about new entities: facts. Arguably, it is a valuable theory telling us some untold story about such entities and their relation with truth and $\mathcal{L}$. The problem for $\text{TF}_{\mathcal{L}}$ is that this is not the case. Everything it proves can be replicated by $\text{MT}_{\mathcal{L}}$ and Theorem 1 can be strengthened. To keep things as simple as possible, let’s introduce another axiom:

- vi For all $p, R_{\mathcal{L}}(p)$, there is a $\ulcorner \varphi \urcorner$, such that $C(p, \ulcorner \varphi \urcorner)$.

Axiom vi simplifies the proof of the following theorem. It could be eliminated by working with equivalence classes of equivalent Russellian facts.

Theorem 2. *For all γ , $\text{Sen}_{\mathcal{L}^+}(\gamma)$ there is a ψ , $\text{Sen}_{\mathcal{L}^*}(\psi)$ such that $\text{TF}_{\mathcal{L}} \vdash \gamma \leftrightarrow \psi$.*

²⁸MT can be formulated, modulo Gödelian codifications, as a sub-theory of arithmetic. This will be not assumed here.

Proof. First of all, let's define a translation τ from $\mathcal{L}^+$ into $\mathcal{L}^*$. Let $h(\ulcorner \varphi \urcorner) = \varphi$ be a function from Γ into $\varphi / \text{Sen}_{\mathcal{L}}(\varphi)$ —i.e. a function from the set of names of formulas into the set of formulas. Axiom vi ensures that, for all terms p there is a φ , $\text{Sen}_{\mathcal{L}}(\varphi)$ such that $C(\ulcorner \varphi \urcorner, p)$. τ is defined recursively as follows:

- 1 For $\text{Sen}_{\mathcal{L}}(\varphi)$: $\tau(\varphi) = \varphi$
- 2 $\tau(\ulcorner \varphi \urcorner) = \ulcorner \varphi \urcorner$
- 3 $\tau(p) = \ulcorner \varphi \urcorner$ (such that $C(\ulcorner \varphi \urcorner, p)$)
- 4 $\tau(x) = x$
- 5 $\tau(t_1 = t_2) = \tau(t_1) = \tau(t_2)$
- 6 $\tau(T(\ulcorner \varphi \urcorner)) = T(\ulcorner \varphi \urcorner)$
- 7 $\tau(R_{\mathcal{L}}(p)) = \tau(p) = \tau(p)$
- 8 $\tau(F_{\mathcal{L}}(p)) = T(\tau(p))$
- 9 $\tau(C(\ulcorner \varphi \urcorner, p)) = \varphi \leftrightarrow h \circ \tau(p)$
- 10 τ is extended over connectives and quantifiers as expected. In particular: $\tau(\exists x \psi(x)) = \exists x \tau(\psi(x))$

The axioms i and vi guarantee—modulo ZF—the existence of a one-to-one mapping from $R_{\mathcal{L}}$ onto $\text{Sen}(\mathcal{L})$. Therefore, 9 could be simplified as follows: $\tau(C(\ulcorner \varphi \urcorner, p)) = \varphi = \tau(p)$. Now, I will prove $\text{TF}_{\mathcal{L}} \vdash \gamma \leftrightarrow \tau(\gamma)$, for $\text{Sen}_{\mathcal{L}^+}(\gamma)$, by induction on the complexity of formulas.

I Atomic γ . Cases 5, 6 and 7 are trivial.

Case 8. Assume $\text{TF}_{\mathcal{L}} \vdash F_{\mathcal{L}}(p)$. By Axiom vi, $\text{TF}_{\mathcal{L}} \vdash$ there is some $\varphi \in \text{Sen}(\mathcal{L})$ such that $C(\varphi, p)$. Thus, $\tau(p) = \ulcorner \varphi \urcorner$. By Axiom v, $\text{TF}_{\mathcal{L}} \vdash T(\ulcorner \varphi \urcorner)$. Assume $\text{TF}_{\mathcal{L}} \vdash T(\ulcorner \varphi \urcorner)$, such that $\tau(p) = \ulcorner \varphi \urcorner$. By Axiom v, $\text{TF}_{\mathcal{L}} \vdash F(q)$, for some $q, C(\ulcorner \varphi \urcorner, q)$. By Axiom ii, $\text{TF}_{\mathcal{L}} \vdash p = q$. Therefore, $\text{TF}_{\mathcal{L}} \vdash F_{\mathcal{L}}(p)$.

Case 9. Assume $\text{TF}_{\mathcal{L}} \vdash C(\ulcorner \varphi \urcorner, p)$. Then: $h(\ulcorner \varphi \urcorner) = h \circ \tau(p) = \varphi$. So, trivially: $\text{TF}_{\mathcal{L}} \vdash h(\ulcorner \varphi \urcorner) \leftrightarrow h \circ \tau(p)$. Now, assume $\text{TF}_{\mathcal{L}} \vdash h(\ulcorner \varphi \urcorner) \leftrightarrow h \circ \tau(p)$. By the definition of τ : $\text{TF}_{\mathcal{L}} \vdash C(\ulcorner \varphi \urcorner, p)$.

II Case 10. The proof for complex formulas proceeds as expected through induction. In particular, assume $\text{TF}_{\mathcal{L}} \vdash \exists x \psi(x)$. Instantiating: $\text{TF}_{\mathcal{L}} \vdash \psi(t)$. By the induction step, $\text{TF}_{\mathcal{L}} \vdash \tau(\psi(t))$. Introducing the quantifier: $\text{TF}_{\mathcal{L}} \vdash \exists x \tau(\psi(x))$. The converse case is similar.

□

Corollary 1. For all γ , $\text{Sen}_{\mathcal{L}^+}(\gamma)$ there is a ψ , $\text{Sen}_{\mathcal{L}^*}(\psi)$ such that: if $\text{TF}_{\mathcal{L}} \vdash \gamma$, then $\text{MT}_{\mathcal{L}} \vdash \psi$.

Proof. Follows from Theorems 1 and 2. □

Corollary 2. $TF_{\mathcal{L}}$ is consistent relatively to $MT_{\mathcal{L}}$.

$TF_{\mathcal{L}}$ and $MT_{\mathcal{L}}$ are coextensive in the strong sense of Theorem 2. Despite talking about facts, everything that $TF_{\mathcal{L}}$ implies is equivalent to something told by $MT_{\mathcal{L}}$. $TF_{\mathcal{L}}$ does show nothing new about truth.

APPENDIX 2.B.

Objectivity is related to substantive disagreements and speakers' convergence on truth. The issue has been addressed originally in (144), who proposed that speakers' convergence in their assertions is a distinctive feature of objective discourses. For example, physicians tend to agree on their assertions as well as on which theories are correct about the world. Moreover, when two physicians disagree over some issue, at least they agree that one and only one of them is right. On the contrary, speakers' opinions about fashion or comedy tend to diverge to a much greater extent. Usually, what some person finds fashionable or funny, another find disgusting or offensive. Speakers do not converge on comedy or fashion. Moreover, speakers do not try to convince those who have different opinions about such matters by appealing to correct theories. Nevertheless, things are not so simple. Even though the picture is naively correct, it is far from clear that physicians have the required resources to always determine which of two disputants is right — especially since quantum mechanics appeared. In turn, speakers who belong to the same population share largely the same ideas concerning fashion and comedy.

To avoid these problems, (151) shifted the attention from speakers' empirical behaviour to the rules underlying disagreements in objective contexts. According to him, speakers converge on truth because of a deeper reason. Namely, because the rule governing their assertions is representational: speakers aim to depict a portion of the world. Consequently, where a discourse aims to depict some facts, disagreements cannot be solved at random. "Where we deal in a purely cognitive way with objective matters, the opinions which we form are in no sense optional or variable as a function of permissible idiosyncrasy, but are commanded to us" (p. 146). His analysis of the phenomena supplies us with an explanation in terms of language's analytic features. The resulting criteria for objectivity are too weak and were criticized by (148).

In turn, I will explain that cognitive failures in objective disagreements as well as speakers' convergence on truth in objective contexts are explained by the conjunction of three simple items:

- (i) The K-rule,
- (ii) Factivity of knowledge,

and

- (iii) The Law of non-contradiction.

To deal with several speakers I will treat "K" not as a sentential operator, but as a binary predicate: thus, " $K(x, \varphi)$ " means that x knows that φ is the case. Also, I will employ a binary predicate "A" for

assertion: “ $A(x, \varphi)$ ” means that x asserts “ φ ”. Now, assume that the rule governing assertion in objective contexts is the K-rule —see (148) for a defence of the K-rule:²⁹

K-Rule

$$\forall x \blacksquare (A(x, \varphi) \rightarrow K(x, \varphi)),$$

where the black square “ $\blacksquare$ ” expresses deontic necessity. The K-rule says that all speakers ought to assert a sentence only when they know that it is true. Also, assume that knowledge is factive:

Factivity

$$\forall x (K(x, \varphi) \rightarrow \varphi).$$

If speakers know that φ , then φ is the case.

A *disagreement* between two speakers x and y is as a situation where x asserts φ and y asserts $\neg\varphi$. This case can be generalized to situations where x asserts that φ , y asserts that ψ , but φ and ψ together entail a contradiction. For the sake of simplicity, I will restrict the exposition to the previous case.

Disagreement x and y disagree on φ if and only if $A(x, \varphi)$ and $\neg A(y, \varphi)$, or $\neg A(x, \varphi)$ and $A(y, \varphi)$.

A speaker x exemplifies a *cognitive shortcoming* just in case it asserts φ , but she does not know that φ .

Cognitive Shortcoming x exemplifies a *cognitive shortcoming* if and only if $A(x, \varphi)$, but not $K(x, \varphi)$.

Thus, those speakers who do not obey the K-rule are cognitively faulty.

Convergence on truth can be explained as a consequence of the fact that if φ is the case, no speaker is sanctioned to assert “not φ ”

$$\text{Convergence } \varphi \rightarrow \forall x \neg K(x, \neg\varphi)$$

Convergence entails that, in the idealized situation where all the speakers obey the K-rule, all of them assert the same sentences, and only true sentences. Now, I will prove that the K-rule, the factivity of knowledge and the Law of non-contradiction (NC) entail Convergence.

Theorem 3. *K-Rule + Factivity + NC* $\vdash$ *Convergence*

Proof. Assume the K-rule, Factivity and NC. Suppose that φ is the case. For *reductio*, suppose that for some speaker a , it happens that $K(a, \neg\varphi)$. By Factivity, $\neg\varphi$ is the case. So, φ and $\neg\varphi$ is the case, which contradicts NC. $\square$

Consequently, if the language of a community of speakers satisfies the K-rule, Factivity, and NC, and they obey this rule, then they will converge on the sentences asserted—which will be true sentences. The language of empirical sciences as well as the language of perception are candidates to satisfy the requirements. Consequently, this would explain why speakers exemplify a high degree of convergence in these areas.

²⁹Any rule involving factive epistemic conditions C would do.

These three factors also entail that disagreements are objective. When a discourse obeys the K-rule, the Law of non-contradiction, and knowledge is factive, then one of two disputants must be faulty when a disagreement arises.

Theorem 4. *Assume that speakers a and b disagree on φ . Then: $L\text{-Rule} + \text{Factivity} + \text{NC} \vdash a$ exemplifies a cognitive shortcoming or b does.*

Proof. Assume that speakers a and b disagree on φ . By definition, $A(a, \varphi)$ and $A(b, \neg\varphi)$. Also, assume the K-Rule, Factivity and NC. Suppose —without loss of generality— that φ is the case. For *reductio* suppose that $K(b, \neg\varphi)$. Factivity entails that $\neg\varphi$ is the case. Therefore, φ and $\neg\varphi$, what contradicts NC. Consequently, $A(b, \neg\varphi)$, but not $K(b, \neg\varphi)$. By definition, b exemplifies a cognitive shortcoming. The proof is analogous if φ is not the case. $\square$

When a discourse obeys the K-rule, NC, and knowledge is factive, then no contradictory assertions are sanctioned by the rules of the language. The fundamental reason is that the factivity of knowledge and the Law of non-contradiction entail that knowledge is consistent, in the sense that

Consistency-K

$$\forall x(K(x, \varphi) \rightarrow \forall y \neg K(y, \neg\varphi)).$$

This is the main feature of those discourses we usually tend to qualify as objective. The languages of physics, mathematics, logic, or perception are ruled by consistent epistemic conditions, such as knowledge. These command us which one of the horns in a dilemma is the correct one —if any.

Does this mean that anti-realistic discourses are ruled by epistemic conditions allowing massive inconsistencies as well as choosing at will what is true and what is false? Not *necessarily*. Anti-realist discourses are consistent with the Law of non-contradiction. What happens is that, usually, anti-realist discourses allow —even if not massively— at least some contradictions. This is the case of fashion. Nothing forbids that some clothing is fashionable and disgusting at the same time. Because for some clothing to be fashionable all that is required is that some people find it fashionable. Equally, if it is disgusting. As a result, anti-realist discourses of this kind should be ruled by paraconsistent logic.

On the contrary, it is harder to endorse languages holding cases of *extensional* or *counterfactual* independence with the failure of the Law of non-contradiction. This would be tantamount to endorsing the world with metaphysical or physical contradictions. I do not claim that metaphysical contradictions are senseless —for instance, quantum mechanics is a candidate for hosting contradictory situations. However, they come with a cost and we need to be extremely careful when assessing if it fits the metaphysical budget. To claim that mathematics or empirical science is contradictory would entail massive revisions in the whole body of human knowledge.

Finally, note that the rejection of the Law of non-contradiction does not entail that *individual* speakers are systematically inconsistent. In particular, the negation of Consistency-K is consistent with the following one:

Consistency-Individuals

$$\forall x(K(x, \varphi) \rightarrow \neg K(x, \neg\varphi)).$$

Therefore, while two speakers are allowed by the failure of the Law of non-contradiction to sustain contradictory judgements about fashion or comedy, each one would not be allowed to make inconsistent claims. Nobody could be K -related both to φ and $\neg\varphi$. For instance, while $K(x, \varphi)$ and $K(y, \neg\varphi)$, this does not mean that both speakers x and y are allowed to assert that φ and $\neg\varphi$.

APPENDIX 2.C.

Margin for error principles are used by *epistemicism* to capture the characteristic features of knowledge in vague contexts. According to advocates of epistemicism, they describe the epistemic situation underlying the *sorites paradox*. This puzzle affects predicates such as “heap”, “bald”, “red”, or “tall”, which have unclear or fuzzy boundaries of application. For all these predicates, there is no sharp boundary between the objects which satisfy them and those which do not. For example, one grain of sand does not make a heap. In turn, ten thousand grains will do. The problem arises because there are numbers n greater than one, but smaller than ten thousand such that is not clear if n grains of sand make a heap or not. The epistemicist solution claims that there is a definite answer for all these cases. What happens is that we do not or cannot know what the answer is. Thus, for all numbers n , n grains of sand are a heap, or n grains of sand are not a heap, even if nobody can give a definite answer.

Margin for error principles are conditionals of the following form:

if x and y differ incrementally with respect to φ , and $K(Px)$, then Py ,

where “ K ” is an operator expressing knowledge, P is a vague predicate, and φ is a feature whose variation affects the satisfaction of P . For example, if x is formed by n grains of sand, and y is formed by $n + 1$ grains, and it is known that x is a heap, then y is also a heap. Epistemicists regard these principles as a characterization of *inexact* knowledge. They claim that knowledge in vague contexts must be “surrounded” by the satisfaction of the predicate in question —on the contrary, it would not be knowledge. If speakers know that an object satisfies the vague predicate this is because the closer cases also satisfy it. Therefore, they provide knowledge with a safety area surrounding it —this is why it is knowledge.

Williamson (145; 146) has shown that margin for error principles are inconsistent with the knowability principle KP — $\varphi \rightarrow K\varphi$. This result can be reconstructed in modal logic. First, let’s formalize KP as the following strict conditional

KP

$$\Box(\varphi \rightarrow K\varphi).$$

Indeed, it is natural to read KP as entailing a strict conditional of this form. Its standard formulation only involves the material conditional. However, KP is usually taken as the characteristic feature of anti-realist conceptions of knowledge. Consequently, it is reasonable for advocates of these positions to read KP as an analytical truth of anti-realist knowledge or as a metaphysical principle of anti-realism. In both cases, KP will contain an operator strong enough to imply the previous strict conditional. Similarly, margin for error principles can be formalized as a strict conditional too:

$$\Box(\Diamond K\varphi \rightarrow \varphi).$$

This says that, if in some proximate —possible— situation φ is known, then φ must be the case. As the classical case, ME establish a safety area for knowledge. Now, it can be proved that KP and ME are mutually inconsistent in normal modal logic.

Theorem 5. *Let $\vdash$ be probability in any normal modal logic. Assume $\Diamond\varphi$ and $\Diamond\neg\varphi$. Then: $ME \vdash \neg KP$.*

Proof. The proof can be recovered in tableaux form. Let R be the accessibility relation between worlds. Assume $\Diamond\varphi$ and $\Diamond\neg\varphi$ (in w_0). Assume ME:

$$\vdash \Box(\Diamond K\psi \rightarrow \psi), (\text{in } w_0).$$

Now, for *reductio* assume KP:

$$\vdash \Box(\psi \rightarrow K\psi), (\text{in } w_0).$$

By elimination of the diamond, suppose that φ , (in w_1 , such that $w_0 R w_1$). By KP, we have that $\varphi \rightarrow K\varphi$, (w_1). Thus, $K\varphi$, (in w_1), and $\Diamond K\varphi$, (in w_0). But, by ME, φ , (in w_0). Now, by elimination of the diamond again, assume $\neg\varphi$, (in w_2 , such that $w_0 R w_2$). By KP, $\neg\varphi \rightarrow K(\neg\varphi)$, (in w_2). Thus, $K(\neg\varphi)$, (in w_2), and $\Diamond K(\neg\varphi)$, (in w_0). But, again by ME, we have $\neg\varphi$, (in w_0). Hence, $\varphi \wedge \neg\varphi$, (in w_0): a contradiction. $\square$

This little result tells us something important about Independence. As in the case of counterfactual independence, suppose that modality in KP and ME is metaphysical modality —see section 2.5.3. Independence —both in the form of contextual or counterfactual independence— establishes a gap between the world and knowledge. Then, in the case of contingent sentences, ME shows that a certain margin of safety is necessary to know the world. The open gap introduces an epistemic risk: the risk of not knowing how things are. Therefore, bridging that gap comes at a cost. And, the margin established by ME assures us that it is worth paying it. However, what happens if the sentences are necessarily true? For instance, this could be the case of mathematical sentences. In this case, the theorem does not follow. There is no risk of failure! If one thing is always true, we cannot miss. We will always hit the target, and the margin set by ME is irrelevant. At this point, the following question arises: if mathematics is necessary, does Independence imply no epistemic difficulties? If this were so, then Benacerraf's challenge would not really be a problem. Mathematical knowledge would be free as soon as mathematical language is understood. For example, if the axioms of arithmetic are necessarily true, and speakers understand what “ $5 + 7 = 12$ ” means, then they will know that this is necessarily true —therefore, true. However, even if this picture is correct, it does not eliminate the explanatory pressure put on platonism by the epistemological challenges. It only shows that the gap open by Independence must be a *hyperintensional* gap —i.e. the concept of *explanation* contained in this thesis is a hyperintensional concept. Some instances of the gap can be analyzed by the modal idioms —for example, cases of extensional and counterfactual independence. Nevertheless, this may not be all that it contains. The explanation of mathematical knowledge introduced in Chapter 5 will be hyperintensional in nature.

3 Anti-realist Pluralism

Fictionalism

3.1 INTRODUCTION

In this chapter, I will analyze fictionalism. Fictionalism accepts mathematical language at face value: mathematical theories are about mathematical, abstract objects. However, it denies the existence of these objects. According to fictionalism, mathematical theories are fictions developed by mathematicians throughout history for various reasons. Consequently, mathematics would not be a *science* of objects. No more than other fictions, such as *Hamlet*, *Oliver Twist*, or *War and Peace*. Instead of truthfully describing the mathematical realm, mathematicians create stories and study what happens in them. Thus, fictionalism reduces mathematical knowledge to logical knowledge: knowledge about the consistency and consequences of those fictions.

I will challenge fictionalism for three reasons. First, I will argue that under two reasonable assumptions, fictionalism cannot explain mathematical reliability. Then, I will try to do better, arguing that even if these assumptions fail, fictionalism cannot explain mathematical reliability. In consequence, it cannot solve Benacerraf's reinforced challenge. Second, I will distinguish between two types of fictionalism: *full fictionalism* and *algebraic fictionalism*. The former claims that mathematical language is a meaningful language about abstract objects. In turn, the latter does not attribute any content to mathematical language. I will argue that if full fictionalism can explain how mathematical vocabulary acquires its content and how mathematicians acquire their linguistic competence, then so can platonism. Since the failure of platonism is the best argument for fictionalism, this is fatal for it. Finally, I will argue that algebraic fictionalism is in no better position. The reduction of mathematical knowledge to logical knowledge presupposes the linguistic competence of mathematicians. But, if algebraic fictionalism tries to explain this, then it reduces to full fictionalism.

3.2 FICTIONALISM

Fictionalism is the philosophical account of mathematics that results from the rejection of *Existence*. According to fictionalists, mathematical objects do not exist. Consequently, fictionalism is a form of nominalism and a sceptic philosophical position. In turn, fictionalists agree with platonists in taking mathematics at face value. For them, the sentence "2 is prime" aims to be about an object, the number

two. It asserts that this object has the property of being prime. As a result, fictionalism entails that mathematical theories are not true—they are false, or neither true nor false. Even weak mathematical theories, such as Peano-Dedekind arithmetic, entail existential claims—for example, that there is a prime number between 1 and 3. Thus, mathematics would be similar to fiction. Think about the famous play *Hamlet*. Hamlet—the main character of the tragedy, not Shakespeare’s son—did not exist. Therefore, claims about him are not true—they are false, or neither true nor false. For example, it is not true that Hamlet killed Laertes. Nor is that Ophelia jumped into the river. None of these events have taken place. Nevertheless, the analogy only goes this far. Fictionalism does not entail any further connection between fiction and mathematical theories. In particular, it does not deny important disanalogies between them.

Fictionalism offers a promising account of mathematics. If it is correct, it would be in a position to explain mathematical knowledge. This would simply be explained using the best theories of knowledge about fiction at our disposal. More specifically, fictionalism denies that mathematical knowledge satisfies Responsiveness

$$\text{Responsiveness } C\phi + \phi + X \Vdash K\phi,$$

where $C\phi$ expresses some counterfactual relation between agents and the facts depicted by ϕ , and “ X ” is a variable standing for some appropriate facts—this unknown prevents any simplistic definition of knowledge. This is the thesis that knowledge of some facts is—partially—explained by being C -related to these facts. For example, to know that the Sun is a star, Responsiveness requires that: (i) the Sun is a star, (ii) we are causally related to this fact, and maybe (iii) some unknown factors X . Usually, empirical knowledge is regarded as a paradigmatic case of Responsiveness. On the contrary, knowledge about the plot of a novel or a play does not obey the same demands. If these are fictions many of the sentences they contain will be not true—i.e. what the plot depicts is not the case. This immediately invalidates Responsiveness. In turn, fictionalism solves X in the following statement

$$X \Vdash K\phi$$

with the best explanation of knowledge about mathematical fiction.

In Chapter 2 I argued that Responsiveness is the main cause of Benacerraf’s challenge and its reinforced version. By failing to find satisfactory instances of this scheme, platonism gets into trouble. Fictionalism avoids the same challenge by rejecting Responsiveness. However, this does not constitute the final response. Like all accounts of mathematics, fictionalism faces an important explanatory pressure. Even if mathematical theories are fictions, fictionalism must explain the knowledge provided by them. At this point, it comes with a rather simple idea. Knowledge about a mathematical, fictitious sentence ϕ reduces to two things: (i) knowing that ϕ follows from a given mathematical, fictitious theory Γ , and (ii) knowing that Γ is consistent:

$$K(\Gamma \vdash \phi + \Gamma \text{ is consistent}) =_{df} K_F \phi,$$

where K_F is “fictionalist knowledge.” In this way, fictionalism reduces mathematical knowledge to logical knowledge—knowledge about logical consequence and consistency. As a result, as long as mathematicians are good at knowing logical facts, then the fictionalist can explain mathematical knowledge. And, fictionalists claim that mathematicians are very good at logic.

Benacerraf's first problem is avoided on this account since the idea that a good mathematical theory must successfully refer to any objects, let alone a unique system of objects, is abandoned. Asking which object, precisely, is the number 2 is analogous to asking who, precisely, is Hamlet. On a plausible view of fictionalist discourse, the name "Hamlet" really refers to no one, and so it is wrongheaded to ask, outside of the context of the fiction, who Hamlet is. (76, p. 125)

Mathematical knowledge does not concern bare statements such as "there are no naturals x, y, z, n such that $n > 2$ and $x^n + y^n = z^n$ ". On the contrary, mathematical practice is about generating fictions. In the case of arithmetic, a fiction about numbers. In the case of set theory, a fiction about sets. Thus, mathematical knowledge is knowledge about entailment, i.e., about what follows and what does not from the corresponding generative assumptions of these "fictions" —the axioms of the theory.

Strictly speaking, fictionalism denies Abstractness and Independence too.

Abstractness Mathematical objects are abstract.

Independence $C\varphi \not\models \varphi$,

where C are agents' epistemic responses concerning φ . Abstractness claims that mathematical objects are abstract. In turn, independence claims that mathematical facts are not explained by epistemic factors. Thus, both sentences involve reference to mathematical objects, which do not exist. Consequently, they are not true. However, fictionalism does not deny the weaker version resulting from prefixing them with the condition that mathematical objects exist. It accepts that, if mathematical objects existed, they would be abstract objects. Similarly, if mathematical objects existed, they would be independent of agents' epistemic responses. The failure of these two theses shows the real strength of fictionalism. If there are no abstract objects, the need to explain how speakers relate to them —a mysterious relation— disappears. Even more, if there are no mathematical objects —of whatever kind— the gap between the world and mathematicians vanishes. The explanation of mathematical knowledge does not consist in explaining how this gulf is bridged. Thus, the task becomes simpler.

As a result of its scepticism, fictionalism rejects a strong notion of mathematical objectivity. There is no such thing as *the* true mathematical theory. The only relevant objectivity for mathematicians is logical objectivity. This dictates that there are many consistent mathematical theories, all equally acceptable from a metaphysical and semantic point of view. Fictionalism is inherently plural. Sometimes, this has been used against it —after all, mathematical practice harbours a certain degree of objectivity. However, fictionalists have a chance of getting back a weaker form of objectivity. They introduce the predicate "to be true in the fiction", such that a sentence is true in a given fiction just in case it follows from the axioms of the theory. For example, "there are infinitely many primes" is true in the fiction PA. Similarly, "there are inaccessible cardinals" is true in the fiction ZF+Con(ZF). Once this is done, the fictionalist can understand plain truth as truth in the *canonical story* of mathematics, i.e., the story developed through centuries by mathematicians. This form of objectivity will be much weaker than the objectivity accepted by platonism. For example, assume that the canonical story about sets is ZF and is consistent. Then, this

story is highly incomplete. It does not prove nor disprove CH or $V = L$. None of these sentences will be true in the canonical story of sets. Hence, no one of them will be true *simpliciter*.¹

Fictionalism has been criticized on several occasions —see (2, p. 98) and (5, pp. 8-15) for a survey of the criticism posed on fictionalism. The Indispensability Challenge and the accusation that inadvertently reintroduces abstract objects —sentence types— (2, pp. 99) and (5, p. 15) are among the most famous criticisms. Here, I will leave aside all these worries. Instead, I will pursue a different route of attack, opposing two challenges to it. First, I will argue that fictionalism does not meet the reinforced version of Benacerraf's challenge. Fictionalism offers a very effective solution to its original version. By treating mathematical theories as fictions, it achieves a nominalistically acceptable epistemology —if logical knowledge is nominalistically acceptable. However, this is not enough to explain the reliability of mathematicians. Under extremely weak assumptions, this picture entails that mathematicians make systematic mistakes. This makes them highly unreliable.

Second, using an eminently semantic conception of language, I will argue that fictionalism cannot explain how mathematical language acquires its meaning, nor how mathematicians understand it. The core idea of the argument is that language is not a syntactic or grammatical category. Indeed, grammatical categories are at the same time semantic categories and vice-versa. For example, consider proper names. An object is a proper name because it attempts to perform a semantic function: pointing to an object, among other things. The name "Aristotle", among other things, points to the ancient Greek philosopher Aristotle. If it did not perform this function, it would not be a name. It would just be a sound, a spot on the screen, or some other object. The same applies to predicates. An object is a predicate because it attempts to perform a semantic function: to apply with truth to other objects, among other things. Thus, using the semantic nature of language, I will argue that fictionalism either

- (1) Cannot explain how the content of mathematical language is fixed, nor how mathematicians understand it,

or

- (2) Its explanation allows platonism to solve Benacerraf's challenge.

(2) is fatal for fictionalism, since platonism's inability to respond to Benacerraf's challenge is the main argument in favour of it. In turn, (1) is the Language Acquisition Challenge (LAC) of Chapter 2 —see Chapter 2, 2.4.2.2. Now, this challenge invalidates the very formulation of fictionalism. This reduces mathematical knowledge to logical knowledge. However, logic is also a linguistic phenomenon. Consequently, if fictionalism fails to explain how mathematicians understand mathematical language, its explanation of mathematical language will fail too. In some sense, this argument tells us that fictionalism is in no better position than platonism concerning mathematical knowledge —it does not make it less mysterious.

From this I will conclude that an alternative explanation of the understanding of mathematical language is required. At the same time, this must be an explanation of mathematical knowledge. Therefore, what we need is an explanation of linguistic competence in a language that comes equipped with certain

¹ See (3; 4) for an attempt to solve this problem.

mathematical truths —a minimal true mathematical theory. This will be sufficient to solve the above two challenges, as we as to formulate meaningful philosophical propositions, such as fictionalism. At the same time, it will provide a justification for a certain degree of mathematical objectivity stronger than that proposed by fictionalism. Mathematical objectivity cannot be defined from the plural realm of fictions. On the contrary, once we have a minimal notion of mathematical objectivity and truth, we will be able to open up to the mathematical multiverse.

3.3 THE RELIABILITY CHALLENGE

Under two natural assumptions, fictionalism does not explain mathematical assertion. Therefore, even if it provides an acceptable solution to Bencacerraf’s challenge, it does not explain the reliability of mathematicians —who would systematically make false assertions. The first assumption is that the rule of assertion to be the knowledge rule:²

K-Rule

$$\blacksquare(A\varphi \rightarrow K\varphi),$$

where “A” is an operator expressing assertibility, “K” is an operator expressing knowledge, and the black square $\blacksquare$ expresses deontic modality. The K-Rule says that a sentence must be asserted only if it is known that it is true (147). The second assumption is that mathematicians do assert some mathematical sentences taken at face value. For example, simple arithmetical sentences such as “ $2 + 2 = 4$ ” or “it is not the case that $7 + 6 = 14$ ”. It is doubtful whether mathematicians make assertions in complex areas of mathematics. For example, in algebraic disciplines, such as abelian group theory, mathematicians explicitly study the consequences of some collection of statements. In this case, it is reasonable to conclude that mathematicians do not assert the property of associativity. Instead, they would assert a conditional. Namely, the conditional “if the statements under study are true, then the property of associativity holds”. Algebraic areas of mathematics have a conditional nature (35, Ch. 6). However, the practice underlying arithmetic differs notoriously from the practice underlying other fields of mathematics. Indeed, it’s natural to accept that mathematicians make simple arithmetical assertions. In this case, they do not only study the consequences of Peano-Dedekind axioms. Arithmetic existed long before these axioms did. Consequently, I believe the second assumption is also reasonable.

Now, fictionalists regard arithmetical sentences as not true. Therefore, by factivity — $K\varphi \rightarrow \varphi$ —, knowledge about arithmetical sentences is not attainable. Finally, by the K-Rule, arithmetical sentences are not assertible. If the K-rule is the rule governing assertion, and if mathematicians assert arithmetical sentences, then they make untrue assertions. Besides, they make untrue assertions on a systematic basis. If mathematicians assert that $2 + 2 = 4$, then they will also assert that $2 + 3 = 5$, that $2 + 2 \neq 5$, etc. Mathematicians will be disposed to assert infinitely many untrue sentences. Under two weak and natural assumptions —the K-rule and the assumption that mathematicians assert some arithmetical sentences—,

²The notion of *knowledge* is highly problematic. However, all arguments in this section depend on a single factor: the factivity of knowledge. Therefore, any rule based on a factive condition —for instance, justified belief— would work. Nevertheless, I will use the K-rule for simplicity.

fictionalism entails that mathematicians are highly unreliable professionals. To be clear, fictionalists do not reject that mathematicians make assertions. What they deny —under the previous assumptions— is that mathematical sentences taken at face value are assertible. Consequently, fictionalism turns out to be an error-theoretic account of mathematical knowledge, endorsing mathematicians with massive errors.

3.3.1 Paraphrase Functions

Is there any way out of the trouble? I will explore three plausible routes which could lead the fictionalist to safe terrain. First, fictionalism is not the unique form of anti-realist anti-platonism that denies Existence. Deductivism, formalism, and conventionalism are just a few examples. The main difference between fictionalism and these positions is that fictionalism takes mathematical statements at face value. On the contrary, many of the alternative forms of anti-realist anti-platonism use paraphrase functions to save mathematical truths. That is, they use a translation preserving entailment from the language of mathematics to another nominalistically kosher language. Take the case of deductivism. Deductivism translates a mathematical sentence φ as a conditional of the form

$$\bigwedge \Gamma \rightarrow \varphi,$$

where $\bigwedge \Gamma$ is the conjunction of the axioms required to prove φ .³ Note that the falsity —or the lack of truth value— of $\bigwedge \Gamma$ and φ is compatible with the truth of the whole conditional. Similarly, *formalism* paraphrases mathematical sentences as claims about what is provable in particular formal systems.⁴ Different accounts propose different translations. The important thing is that, according to them, mathematical sentences taken at face value “represent” or “suggest” (2, p. 100) their corresponding translation. Paraphrase functions preserve the truth of mathematics and account for mathematical assertion at the cost of breaking the connection between ontology and semantics. While true, mathematics is no longer the science of an abstract realm of objects.

Could fictionalism appeal to paraphrase functions to explain the reliability of mathematicians? Deductivism, formalism, and other forms of anti-realist anti-platonism do not differ metaphysically or semantically from fictionalism. All of them deny the existence of mathematical objects and respect logical entailment. Again, consider deductivism. As deductivism, fictionalism accepts that if φ follows from Γ , then the conditional $\bigwedge \Gamma \rightarrow \varphi$ is logically true. Generally, given a mathematical sentence, the translation proposed by some anti-realist philosophy of mathematics that denies Existence and accepts classical logic is true if and only if is true for fictionalism. Therefore, the latter could adopt paraphrase functions to explain mathematical assertion. The strategy is the following. When mathematicians assert a mathematical sentence, they are asserting a true sentence. However, they are not asserting it at face value. On the contrary, they use such a sentence to represent another covert one: the sentence assigned by the translation.

³Usually, deductivism requires finitely or at least recursively axiomatized mathematical theories. This is not a huge loss. All theories known by humans have one of these characteristics.

⁴For instance, “ $2 + 5 = 7$ ” would be translated as the sentence claiming that “ $2 + 5 = 7$ ” is provable in PA, Q or another formal system —note that this sentence is not a conditional. There is a close connection between deductivism and formalism —let’s ignore that formalism requires a formal framework. By completeness and the deduction theorem, $\bigwedge \Gamma \rightarrow \varphi$ is valid if and only if φ is provable from Γ .

But, this is enough to explain the reliability of the mathematicians. Benacerraf's strengthened challenge only asks for an explanation of why mathematicians tend to assert true sentences. It says nothing about what the semantic content of those sentences should be. Therefore, translations could help the fictionalist to solve the problem.

This strategy faces two problems. First, fictionalism would lose its *raison d'être*. The difference between it and the remaining alternatives to platonism rests, precisely, in the use of paraphrase functions. Unlike them, fictionalism does not appeal to any surrogate truths to interpret mathematical theories. It simply accepts that they are not true. In consequence, if fictionalism is supplemented with additional paraphrases it will collapse in a different philosophical position.

Second, nominalistically kosher paraphrase functions are too weak to recover classical mathematics. Rayo (117) has shown that there is no uncontroversial paraphrase of arithmetic satisfying a minimal list of nominalistic requirements.⁵ An *uncontroversial paraphrase function* is a paraphrase function —i.e. a function preserving entailment— that satisfies the following three constraints:

- (i) The logical resources of the output-language do not exceed n -order quantification, for finite n ,
- (ii) The output-language does not contain intensional operators,
- and
- (iii) It preserves truth-values even when the the output-language has a finite domain.

It should be clear why any translation neglecting (i)-(iii) is controversial. The first constraint (i) forbids translations based on infinitary output-languages. These translations are highly controversial because infinitary languages, i.e., languages with an infinite vocabulary, are linguistically controversial. After all, human languages are finite in nature. Nobody knows what an infinitary language exactly looks like and most philosophers reject them as useful pieces of mathematics in model theory. The second constraint (ii) blocks controversial *subtraction operations*. For example, (98) proposed to translate a mathematical sentence φ as

⁵Rayo also introduces the concept of a nominalistically acceptable paraphrase. A nominalistically acceptable paraphrase τ is a paraphrase function such that:

- i τ assigns sentences with the same truth conditions as “ $\exists!_n x F$ ’s” to sentences of the form “the number of the F ’s = n ”
- ii τ preserves entailment
- and
- iii τ assigns sentences with trivial truth conditions —that is, true regardless of what the world is like— to true sentences of pure arithmetic and sentences with impossible truth conditions —false regardless of what the world is like— to false sentences of pure arithmetic.

Requirement (iii) seems to entail the necessity of mathematics: only translations employing necessarily true sentences are considered. This could include deductivist and formalist translations. Thus, if φ follows from Γ , then $\bigwedge \Gamma \rightarrow \varphi$ is necessarily true. However, not all translations give necessarily true outcomes. To solve this problem, the world in (iii) can be understood as concerning only empirical matters.

φ , except for the part about mathematics.

Subtraction operators are problematic, since they assume that we can clearly delimit the mathematical and non-mathematical content of a sentence. In general, these operations involve the ability of splitting the semantic content of sentences into clearly distinct parts —other examples are (55; 29; 153). However, this is highly controversial. It is far from clear that such a division can always be made according to the subject matter at the time. For example, it is controversial whether we can distinguish the mathematical from the non-mathematical content of a sentence, the empirical from the non-empirical, or the actual from the possible. It cannot be shown that every translation based on subtraction operations is expressible only using modal operators. However, most of them do —see (29; 153). Consequently, the second restriction —if not sufficient— is at least necessary to avoid controversial translations of this kind. Finally, the third constraint (iii) ensures that paraphrase functions do not involve controversial metaphysical assumptions. Specifically, the assumption that there are infinitely many objects. Without this constraint the truth value of the translations could depend on highly controversial assumptions about the cardinality of the domain of objects.

As soon as it is clear that (i)-(iii) ensure that a translation is not controversial —at least they reduce significantly its controversial status—, the moment of discourage arrives. There is no paraphrase function for PA's language satisfying (i)-(iii) that respects the truth of PA's axioms. The proof is simple. By (i) and (ii), the output-language $\mathcal{L}$ only has a finite lexicon and quantifiers of finite type. By (iii), the translation preserves truth even though $\mathcal{L}$ only has a finite domain. Now, the collection of truths of languages of finite order, with a finite lexicon, and a finite domain is decidable. Besides, by definition, paraphrase functions are recursive. Therefore, the set of arithmetical truths would be decidable: a contradiction.

Fictionalism can still accept translations of finitistic portions of PA satisfying (i)-(iii). However, this is insufficient to account for mathematical assertion —assuming the K-rule and that mathematicians assert some arithmetical sentences. If mathematicians assert at face value that $2 + 2 = 4$, then surely they will assert that there is no bigger natural number. However, finitism entails that such a number exists. Therefore, a finitistic translation is not sufficient to explain the reliability of the mathematicians —at least not beyond doubt. Fictionalism cannot rely on paraphrases —satisfying or not (i)-(iii)— to explain mathematical assertion. At least not without controversy. This is enough to dismiss the first option to solve Benacerraf's reinforced challenge. Little will be done if you try to move from a contentious situation —the failure of explaining mathematical assertion— to another, equally disputed one.

3.3.2 Hermeneutic Fictionalism

A second option is to rely on the distinction between *hermeneutic* and *revolutionary* fictionalism introduced by (18) [see also (5, p. 12)]. According to hermeneutic fictionalism, mathematicians indeed have the intention of creating a fiction. When they do mathematics they *do not intend* to refer to and describe a particular realm of objects. Mathematicians talk is considered by themselves as a form of fictional talk. For example, when they say that 3 is a prime number they would not intend —unsuccessfully— to refer to and describe a particular object. Rather, they would be like Shakespeare. When he wrote *Hamlet*, and

imagined that Hamlet's father is murdered with ear's poison, he was not trying to refer to somebody who fits the description "being Hamlet's father" —the main character of the play, not Shakespeare's son.

When a play or a film is performed, it is reasonable to think that the actors do assert at face value what the fiction says. For example, when the actor playing Horatio in *Hamlet* says

Two nights together had these gentlemen,
Marcellus and Bernardo, on their watch,
In the dead vast and middle of the night,
Been thus encounter'd. A figure like your father

he is asserting at face value that Marcellus and Bernardo have seen the ghost of Hamlet's father. His assertion will be false, and the actor will know it. But, this is not a problem. Indeed, that is what being a good actor is all about, knowing how to state with conviction false things. In turn, creators and consumers of fiction do not literally affirm what fiction says. Instead, they make claims about the fiction. They say things like that *the characters* of Marcellus and Bernardo are frightened when they see the ghost of Hamlet's father. Or that the death of Ophelia's *character* is a dramatic moment in the play. The qualifier "characters" reveals that they refer to the work as a literary creation. They do not claim that anyone is frightened, nor that anyone is dead. This leads us to differentiate between two levels of discourse concerning fiction: the *textual* and the *metatextual* level (47).

Textual Assertion at face value of what the fiction says

Metatextual Assertion about the fiction as a creation

The textual level is the level of the fictional plot. It consists of the use of the language in which the work is written, and the use of information exclusively contained in the work. For example, at the textual level of *Hamlet* there is no information on whether there are kangaroos in Australia. Thus, actors are situated at the textual level of fiction. In turn, the metatextual level is the level of people that relates to the work as a creation. Creators and consumers are situated at the metatextual level. Traditionally, no self-references are found in fictions, which recognise themselves as a creation. Therefore, in many cases the textual and the metatextual levels are perfectly differentiated —although this is not always the case.

Hermeneutic fictionalism attempts to solve the challenge posed by mathematical assertion by appealing to this division. Mathematicians —as creators and consumers— would be at the metatextual level of mathematical fiction. An immediate consequence is that mathematicians rarely assert any mathematical sentence. Usually, creators do not believe their own creations. Also, consumers, when they are engaged in the plot of a fictional novel, movie or a theatre play, they do not take it —no matter how good and realistic it is— as something that is really happening. For example, they are aware that the actor performing Hamlet behaves *as if* he were slaying Claudius. Nobody calls the police when the curtain falls to report a murder —indeed, this would be a comically awkward situation. This, precisely, is the cornerstone of imagination and fictional mood: it allows us to cross the limits of the factual.

Similarly, mathematicians would never assert that 3 is prime. They would be perfectly aware that arithmetic, or the mathematical theory in question, is a fiction. Consequently, they would only assert things about it at the metatextual level. That is, they would make claims about what happens in the fic-

tion of arithmetic, in the fiction of analysis, set theory, etc. However, they would not state mathematical sentences verbatim at the textual level. As a result, fictionalism comes with a perfectly good explanation of mathematical assertion as well as mathematicians' reliability. As the creators of these fictions, mathematicians know to a great extent what is going on: mathematical knowledge would be explained through a theory of humans' creation. Furthermore, mathematical assertions at the textual level would be marginal among mathematicians, and would not affect their reliability.

In light of all this, hermeneutic fictionalism seems a promising solution to the assertion challenge. So, can we rely on it to explain mathematical assertions and the reliability of mathematicians? Unfortunately, this is doubtful. Hermeneutic fictionalism is a highly implausible interpretation of mathematical practice. Indeed, no one takes it seriously (5). Specifically, it rejects the second assumption of the challenge. Namely, that mathematicians make some arithmetical assertions at face value. This assumption is placed at the textual level of mathematics. It says that mathematicians make assertions at the textual level of arithmetic—for instance, they assert at face value that $2 + 2 = 4$. Now, if we compare hermeneutic fictionalism with this assumption, the latter comes out better. Empirical evidence supports that mathematicians indeed make such simple assertions at face value on a systematic basis. On the contrary, there is no strong empirical evidence supporting hermeneutic fictionalism. Mathematicians do not behave like writers or artists. And their creations—the mathematical theories—do not have the same characteristics either. Mathematical theories are much more rigid than fictional creations, which undergo continuous elaboration and reworking by creators and followers. Likewise, mathematical knowledge is much more rigid than the creative process of fiction (38). Consequently, hermeneutic fictionalism cannot be accepted at the cost of denying the second premise. What fictionalism needs is a solution compatible with this.

3.3.3 Revolutionary Fictionalism

What about the alternative *revolutionary* fictionalism? Contrary to hermeneutic fictionalism, the revolutionary claims that mathematicians do assert mathematical sentences taken at face value. Therefore, mathematicians are at the textual level of the mathematical fiction—not at the metatextual level. In this sense, they would be more actors than creators. However, this brings us back to the challenge of explaining mathematicians' reliability. If the above two assumptions—that mathematical language obeys the K-Rule and that mathematicians make arithmetical assertions at the textual level—are accepted, then mathematicians would make untrue assertions on a systematic basis. Revolutionary fictionalism entails that mathematicians are highly unreliable.

Which are the options for the revolutionary? If she is unable to meet the challenge, it will be at a disadvantage. Let me explore two options—I will consider the second in the next section. As I said, any solution should accept the modest assumption that mathematicians assert at the textual level simple arithmetical sentences taken at face value. However, what is the situation concerning the other assumption? Is the revolutionary inevitably committed to the K-Rule? First, let's assume momentarily that she is and see what happens. The second assumption claims that the K-Rule is the rule governing mathematical *assertion*. Therefore, what revolutionary fictionalism has to do is to accept this rule while explaining why mathematicians are reliable. At this point, an immediate solution is to deny that mathematicians

speech acts are assertions. Instead of asserting, their practice would be one of *acceptance*. The distinction between asserting and accepting would allow mathematicians to accept sentences without believing or asserting them. For example, suppose that mathematical sentences are acceptable if and only if they follow from the canonical story of mathematics —(5, p. 14). Then, mathematicians will *accept* that $2 + 2 = 4$. And they accept that $2 + 3 \neq 4$. But in no case will they *assert* these sentences at the textual level. Thus, they could accept untrue sentences without asserting untrue sentences. Their speech acts would only be assertions in appearance —they would be “acceptances”. This move would allow the revolutionary to reconcile two things. On one side, they can accept the K-rule. On the other, they prevent mathematicians from systematically making mistakes. These would rarely assert anything at the textual level of the mathematical fiction.

This solution is problematic for two reasons. First, the fictionalist must explain what the linguistic or epistemic practice of acceptance consists of. It is not enough to say that this practice exists. She must describe what its rules are and explain how it differs from other normative practices, such as asserting or judging. This is a theoretical task that no one has done so far. Until then, this will not be an acceptable solution. Second, in this context “to accept” seems to be synonymous to “to believe” or “to judge positively” (62; 18). Agents accept that φ when they believe or judge positively that φ is the case. For example, mathematicians accept that $2 + 2 = 4$ when they believe that this is the case. Also, they do not accept that $2 + 2 = 5$ because they do not believe this is the case. But, if this is so, accepting would not constitute a different practice —contrary to the practice of giving and accepting things. Moreover, by the K-rule, mathematicians must assert φ only if they know that φ . Therefore, they must assert φ only if they believe that φ —knowledge entails believing. And, if the equivalence between believing and accepting is correct, mathematicians assert φ only if they accept that φ . Consequently, if believing and accepting were really synonymous, assertion would be the linguistic correlate of acceptance. No difference would be found between them. Thus, there are serious reasons to believe that acceptance is not a normative practice in itself.

The fictionalist could reply that the expression “to accept” is used by the theoretician to describe a convention. In this case, the convention of point out a few mathematical sentences among the myriad of them —those sentences that follow from the canonical story of mathematics. Conventions are different from constitutive rules, as the rules governing speech acts or games (147). The latter are necessary to perform the acts ruled by them. More precisely, the rule that one must φ to perform an act is constitutive of the act if and only if it is necessary that one must φ it in order to perform the act.⁶ Take the case of chess. Necessarily, the rule that the knight moves in L must be observed to be playing chess. If the knight moved four squares at a time, it would be a different game. On the contrary, conventions are arbitrary. They are not “essential” to any activity, and can be replaced by others —more or less— at will. Using the notion of convention, revolutionary fictionalism can argue that mathematicians “accept by convention” certain mathematical sentences. Those sentences belonging to the canonical story of mathematics. Thus, they could accept by convention untrue sentences, without believing or asserting at the textual level these sentences.

⁶Demanding that φ is necessarily satisfied will be excessive. This would imply that playing a game is incompatible with breaking the rules.

This solution is as problematic as the previous one. The main problem is the collapse of revolutionary fictionalism into other forms of anti-realism. For example, it could collapse into conventionalism — which is very problematic (110). In general, it will collapse into a position very similar to hermeneutic fictionalism. This is because the appeal to conventions reintroduces a level distinction very similar to the one made by the latter. Mathematicians will not assert sentences accepted by convention. That is, they will not make assertions at the *textual level of the convention*. Thus, they will not make untrue assertions, solving the challenge posed by the reliability of mathematicians. Instead, they will assert that this or that sentence is accepted by convention. But, this is similar to claim that mathematicians move at the *metatextual level of convention*. However, as I argued in the previous section, this level distinction does not allow to explain the reliability of mathematics.

The real problem of appealing to different constitutive norms or conventions is that the second assumption of the challenge is inadvertently denied: that mathematicians do assert some simple arithmetical sentences on a systematic basis. For these kind of solutions to work, mathematicians cannot assert any mathematical sentence. They are allowed to do something else with them: what is commanded by the alternative rules or by the conventions. But, they cannot make mathematical assertions on a systematic basis. On the contrary, both solutions would fail to solve the challenge posed by mathematical reliability, because *they say nothing about mathematical assertion*. Consequently, they do not prevent mathematicians from making untrue assertions on a systematic basis if they make mathematical assertions. As I said in the previous section, this is more plausible than the opposite, and the revolutionary fictionalist should accept it. Consequently, no one of these solutions is helpful.

3.3.4 Rethinking Assertion

The only solution left is to deny the first assumption of the challenge, i.e., to deny the K-Rule

$$\blacksquare(A\varphi \rightarrow K\varphi).$$

If assertion obeys a non-factive, weaker condition than knowledge —which, by factivity entails truth—, it would be safe to assert untrue sentences as well as to assert sentences without knowing if they are true. There are two options: to deny that the K-Rule is a rule governing mathematical discourse or to deny that the K-Rule is the unique rule governing assertion.

Let's start with the former case. Assume the K-Rule is not among the rules governing mathematical discourse. Moreover, remember that the K-Rule is a constitutive rule —as any rule governing speech acts (147). By assumption, it is the unique constitutive rule of assertion. Now, there are two options again: or assertion is governed by the K-Rule in other discourses, or no discourse is. Take the first option, and suppose —without loss of generality— that assertions about ordinary objects, such as trees, houses or mountains, are ruled by the K-Rule, as it seems to be the case. Then, because of the constitutive nature of the rules governing speech acts, mathematicians simply would assert nothing. The rules governing mathematical discourse are constitutive. But, the K-Rule —the unique rule governing assertion— is not among them. Consequently, mathematical discourse will not be endowed with the practice of making

assertions. Mathematicians would not make untrue assertions because they would not make assertions at all.

This is simply false. Conditionalization and *Modus Ponens* are sufficient conditions for the language to allow assertion. Thus, to deny that mathematical discourse is not one of assertion is to deny mathematics these two phenomena. But, they are essential to mathematical practice. Just think of the use of *Modus Ponens* in classical and non-classical mathematical proofs. This is more than enough to reject the first option. Even worst, think about the practice of making questions. No one will deny that mathematicians, as well as usual speakers, ask questions about mathematical matters. Often, such questions are what guide their research. Is $V = L$ true? Does every polynomial with complex coefficients and higher degree than zero has complex roots? Any discourse holding the practice of asking questions, also holds the practice of answering them: this is the practice of making assertions. Even if mathematicians are unreliable, this is so precisely because they make false *assertions*.

Alternatively, the revolutionary fictionalist can claim that no discourse is governed by the K-Rule. Thus, assertion would be governed by other rule. But, which one? An in-depth discussion of this issue would take me too far away from my objectives. Let me point out that it is sufficient that the K-Rule governs assertion in some discourse for it to govern assertion in all discourses. Again, this follows from the constitutive nature of rules governing speech acts. Furthermore, it is highly plausible that the K-Rule governs assertion in some discourse. For example, assertion concerning ordinary objects. When people claim —sincerely— things such as “the tree is taller than the house” or “the table is made of wood”, this is because they at least consider that they know that this is the way things are —i.e. they consider that they know that the tree is taller than the house or that the table is made of wood. Note that this is consistent with the fact that these people do not really know how the things are. All that K-Rule requires is that they consider that they do know. Conversely, if people is sincere and do not know these things they would output rather different reports. For instance, they will say something as “I do not know how tall the tree is” or “this table seems to be made of wood, but I am not sure”. When people do not know something and act sincerely —i.e. in accordance with the rules—, they make it explicit in their statements. Therefore, a weaker rule than the K-Rule seems implausible. This rule would involve a *non-factive* condition weaker than knowledge —for example, belief. I.e., it would be of the form

$$\blacksquare(A\varphi \rightarrow C\varphi),$$

where C is some non-factive condition of assertion: it is not the case that $C\varphi \rightarrow \varphi$. In turn, a stronger rule —resulting from a stronger condition C than knowledge— will be as fatal as the K-Rule for revolutionary fictionalism. For, in this case C will be factive. Therefore, it will forbid the assertibility of untrue sentences. Moreover, the rule will govern assertibility in all discourses —including mathematics. Consequently, if the second assumption is also accepted —mathematicians assert simple arithmetical sentences—, the reliability challenge arises again. See (147) for a detailed defence of the K-Rule against weaker and stronger alternatives. This is enough to put some pressure against this option. It is highly implausible that no discourse is governed by the K-Rule. Consequently, is highly plausible that assertion is governed by the K-Rule in any discourse whatsoever —including mathematics.

Finally, there is one option left —the second horn of the first dilemma. Revolutionary fictionalism could defend that assertion, as most games, is governed by several rules. When talking about ordinary objects, speakers would obey one of those rules. When talking about sets, functions and numbers, mathematicians would follow another, different rule. Furthermore, the latter should involve non-factive conditions C for warranted assertibility —see the previous paragraph. Such a rule would make safe the assertion of falsities and sentences lacking a definite truth value. For instance, “ $2 + 2 = 4$ ” could be untrue while $C(2 + 2 = 4)$. The condition of assertion C does not entail truth —in this case it does not entail that $2 + 2 = 4$. Thus, untrue sentences could satisfy it without contradiction.

Again, I cannot stop here to provide a careful analysis of this option. Nevertheless, I also find it highly implausible. Let me briefly state two reasons why I think so. First, non-factive rules of assertion for mathematics meet the same criticism exposed in the previous paragraph. The comments made on the discourse of concrete objects apply in this case too. Mathematicians assert that $2 + 2 = 4$ because—at least—the consider that they know that $2 + 2 = 4$. Similarly for other arithmetical sentences. When they consider that they do not know something, mathematicians do not assert that this is the case. For example, take CH. Some set theorists *believe* this is a false statement. Moreover, some of them claim that they want to prove that this is so. However, nobody dares to assert that CH is true or is false. Thus, belief does not seem enough for mathematical assertion. This speaks against rules weaker than the K-Rule —i.e. a rule resulting from replacing C by a non-factive condition. On the other side, stronger, factive conditions are useless to solve the challenge posed by mathematical assertion. Consequently, it is highly plausible that mathematical assertion is governed by a rule at least as strong as the K-Rule. But, if this is so, revolutionary fictionalism cannot explain mathematical assertion. Second, if assertion is governed by several rules, all of them will apply in all discourses allowing for assertion. For example, they will rule assertion about ordinary objects as well as about mathematical objects. For, suppose that an activity is governed by several rules. Also, suppose this activity is present in different discourses. Then these rules will have an effect on all of these discourses. On the contrary, it is doubtful whether this activity is the same activity across discourses. After all, its identity depends on the constitutive rules. So, suppose that in one discourse no activity is governed by the same rules as the activities present in another discourse. In this case we are forced to conclude that the former is not present in the latter. From this it follows that if mathematical assertion involves a non factive condition, this will also apply to the discourse about ordinary objects. However, as I argued in the previous paragraph, this is implausible.

In conclusion, the rejection of the K-Rule as the unique rule governing assertion is highly problematic. The consequences of this rejection spread beyond mathematical language, reaching other discourses that fictionalism wishes to remain untouched. If mathematical language is as any other language which allows for assertion, any change made in the former will affect the latter.

3.3.5 The Unreliability of Fictionalist Mathematicians

It follows from all the previous arguments that fictionalism must *plausibly* accept the two assumptions of the challenge: that the K-Rule is the unique rule governing assertion and that mathematicians make simple, arithmetical assertions. Moreover, if these two assumptions are true, then fictionalism entails that

mathematicians make untrue, forbidden assertions on a systematic basis. Consequently, if fictionalism is true it is highly plausible that mathematicians are highly unreliable professionals. This is enough to argue against it. Any position that plausibly makes mathematicians unreliable professionals—concerning mathematics—is a bad explanation of mathematics. In this sense, fictionalism is not so different from platonism. Like this, it cannot explain Benacerraf’s—reinforced—challenge.

However, this result can be improved: fictionalism cannot explain the reliability of mathematicians *in principle*. The solution of Benacerraf’s reinforced challenge requires solving X in the following template.⁷

(3.i) Fictionalism can explain why mathematicians are highly reliable when gathering knowledge about X

(3.ii) If (3.i) is the case, then fictionalism can explain why mathematicians accept a mathematical sentences ϕ only if ϕ conforms to X .

Then, (3.i) and (3.ii) entail that

(3.iii) Fictionalism can explain why mathematicians accept mathematical sentence ϕ only if ϕ conforms to X .

Moreover,

(3.iv) According to fictionalism, all mathematical sentences ϕ conforming to X are true

Therefore, from (3.iii) and (3.iv),

(3.v) Fictionalism can explain why mathematicians accept mathematical sentences ϕ only if ϕ is true.

Here “to accept a sentence” means “to be disposed to assert a sentence.” Now, by definition, fictionalism does entail no true instances of (3.iv). According to fictionalism, any solution of X that satisfies (3.i) is not truth-conducive. It says that mathematicians are good at two things: knowing which fictions are consistent and which sentences follow from them. Thus,

$$X = \text{Consistency} + \text{Logical entailment}.$$

However, consistency and logical entailment are not criteria for the truth of sentences—the fact that a sentence is consistent, or the fact that it follows from a consistent theory, is not enough for its truth. Consequently, the following instance of (3.iv) is false:

(3.iv*) According to fictionalism, all sentences ϕ that are consistent or follow from a consistent theory are true.

⁷I take the form of this argument from (2, pp. 51-52).

Furthermore, fictionalism does not entail any other solution for X . Consequently, it does not solve Benacerraf's reinforced challenge. And, *this is so even if the K-Rule is denied*.

For the latter, just note that the K-Rule is not necessary for the challenge. The fact that mathematicians are reliable is not equivalent to the fact that they obey a rule of assertion based on a factive condition—such as the K-Rule. Even if this is the case, they may break the rule when they realise that it does not lead them to the truth. In this situation the rules of language are at odds with the rules of epistemology—breaking the former is an epistemological virtue. But, this is a coherent situation. For example, imagine that for asserting something it is enough to believe it. Believe is not factive. Thus, mathematicians would have to be vigilant to know when their beliefs are not truth-conducive. As a result, in this situation it makes even more sense to require an explanation of mathematical reliability. Concerning the second assumption, its status is more dubious. If mathematicians do not make assertions at face value at the textual level of mathematics, the challenge runs the risk of being trivialized. Consider again the above template. The absence of mathematical assertions makes the material conditional in (3.iii), (3.iv) and (3.v) vacuously true. Mathematicians make assertions only if they conform to X because—in this situation—they do not make assertions at all. Then, any solution for X makes the material conditionals true. Nevertheless, (3.iii) and (3.v) ask for an *explanation* of the conditionals. If this requires intensional or hyperintensional resources, then making the conditionals true will not be enough. An explanation of mathematicians' silence still will be required. Therefore, the absence of mathematical assertions does not trivialize the challenge either.

In sum, the failure of the K-Rule does not trivialize Benacerraf's challenge. Moreover, if explanations require something more than making material conditionals true, then the absence of mathematical assertions does not trivialize the challenge either. In any event, fictionalism cannot provide *in principle* a satisfactory solution for X . Consequently, fictionalism does not solve *in principle* Benacerraf's reinforced challenge.

3.4 FULL FICTIONALISM AND THE LANGUAGE ACQUISITION CHALLENGE

Fictionalism denies the existence of mathematical objects. At the same time, it agrees with platonism in taking mathematical language at face value. According to fictionalism, sentences such as “2 is prime” or “the rationals are dense” aim to be about some objects, the number two and the rational numbers. Specifically, mathematical terms such as “2” aim to refer to mathematical, abstract objects. Consequently, according to it most mathematical theories will contain untrue sentences. Even the language of weak theories, such as Peano-Dedekind arithmetic, is endowed with a referential apparatus involving quantifiers and terms. However, if mathematical objects do not exist, mathematical sentences aiming to quantify or refer to them will be untrue—they will be false, or neither true nor false. Even worse, most mathematical theories entail existential claims falsified by fictionalism. For instance, the claim that the number two exists.

Supposedly, the failure of the referential apparatus of mathematical language does not imply that it lacks content. After all, fictions use terms that refer to no object. But, people manage to read, understand and enjoy them. For instance, take—once more—the case of the term “Hamlet” as it is used in *Hamlet*.

This term does not refer to a danish person who discovered that his father was murdered with ear's poison. Hamlet did not exist. Despite this, people have read and enjoyed this play many times. They understand untrue sentences made at the textual or metatextual levels of fiction. For example, they understand the sentence “Hamlet slayed his aunt Claudius”, or the sentence “Horatio has seen the ghost of Hamlet’s father”. Verbs such as “to slay” or “to see” are part of common English and express well-known actions. Thus, they can be used to attribute actions to fictional characters—even if we do fail in doing this. The language of fiction is not an empty frame.

Fictionalism claims that the same is true of mathematical language. Mathematical vocabulary expresses operations and relations—such as addition, exponentiation, or set formation—which are familiar to us. And, mathematicians try to attribute them—unsuccessfully—to some objects. Usually, this is taken as an advantage of fictionalism over alternative versions of anti-realist anti-platonism.

Moreover, it is worth pointing out here that even if it were the case that mathematical practice ran counter to fictionalism as well as the other versions of anti-realism, we would still have reason to favour fictionalism over these other views, because it would still be true that fictionalism provides a standard semantics for the language of mathematics. In other words, whereas non-fictionalistic versions of anti-realism provide non-standard views of the truth conditions of mathematical sentences, fictionalism provides a standard view here. Thus, even if it provides a non-standard view of whether or not these truth conditions are actually satisfied, fictionalism jibes with mathematical practice more than other versions of anti-realism do. (2, pp. 103-104)

In this section, I will challenge this assertion. Despite appearances, if mathematical language is taken at face value, fictionalism cannot explain how it acquires its content, nor how speakers understand it. Call the version of fictionalism endorsing mathematical language with standard truth conditions—i.e. platonistic truth conditions—*full fictionalism*. I will argue that full fictionalism is an untenable account of mathematics. It fails to meet the Language Acquisition Challenge (LAC) exposed in Chapter 2, 2.4.2.2.

The argument relies on the similarity between full fictionalism and platonism concerning the content of mathematical language. If mathematical language is taken at face value, i.e., as speaking about abstract objects, then both are faced with the same explanatory pressure: they have to explain how the language acquires its content, and how speakers acquire their competence. Moreover, any fictionalist explanation will be acceptable to platonism, which will use it to dispel Benacerraf’s challenges. Thus, if full fictionalism is true, then platonism dispels Benacerraf’s challenge. This is fatal for fictionalism, because platonism’s epistemological problems are the main argument in its favour.

To understand the situation better, think of Charles Dickens’ *Oliver Twist*. This the story of an orphan growing up on the tough streets of 19th century London. The novel is a stark depiction of the harsh living conditions of the city’s inhabitants. Also, it is a fiction. The events it is about did not really happen. Strictly speaking, most of the plot is untrue. Nevertheless, this is not a problem for readers. People, houses, animals, etc, are mentioned in story, but these words are part of English. That is, Dickens wrote in a *perfectly meaningful language*. There are perfectly good empirical explanations of how people acquire concepts such as *person*, *house* or *animal*. However, the case of mathematics is “a bit” different. Most of

the vocabulary of *Oliver Twist* is empirically grounded. It acquires its content through complex physical relations with the environment —people, houses, animals, etc. Conversely, people is not physically related to mathematical facts. Then, the fictionalist owes us an explanation of how *mathematese* acquires its content. To say that mathematics is a fiction does solve the LAC. As a result, full fictionalism is not in a better position than platonism.

3.4.1 The Challenge

Let P be an n -ary predicate. Also, let D be a collection of objects.⁸ Let's say that P is a meaningful predicate on the collection D if and only if it expresses some intensional entity —a concept, property, relation, category, function, etc.— that *can* hold on D , or P *can* be used to describe the objects in D .

P is meaningful on $D \equiv_{df}$ (a) P expresses some intensional entity I which can hold on D or (b) P can be used to describe the objects in D .

The side (a) of the disjunction means that I 's extension is a subcollection of D , if the objects in D exist. Concerning the side (b), if a predicate P can be used to describe the objects in D , then P feature in sentences whose truth condition involve the objects in D —this fact will be important below.

If a predicate P can be used to describe the objects in D , then P feature in sentences whose truth condition involve the objects in D .

This is enough to assign P the desired semantic function without appealing to intensional entities. I believe that (a) and (b) taken together cover the semantic function attributed to predicates by the main semantic theories on the table.⁹ If this is correct, (a) and (b) are at least a good approximation to an uncommitted gloss of what it means to be meaningful.

Now, assume full fictionalism. I.e., the idea that mathematical language has platonistic content. Full fictionalism entails that mathematical predicates are meaningful on some collection of mathematical objects. For example, the arithmetical binary predicate “to be less than or equal to” ($\leq$) is meaningful on the natural numbers. It expresses a relation —here, an intensional entity— that can hold on the natural numbers or it can be used to describe the naturals. Let P be one of those predicates. In what follows I will argue for the *plausibility* of the following conditional:

- (*) If P is a mathematical predicate meaningful on D , then: or platonism about the objects in D solves Benacerraf's challenge or some objects in D exist (or have existed).

Full fictionalism entails that mathematical vocabulary —including predicates such as P — is meaningful on some collection of mathematical objects D . From this and (*) it follows that if full fictionalism

⁸This way of expressing things implies that D is a plurality of objects. This could be avoided by speaking in a slightly more complicated way. We could say that D is a collection such that any element in D is an object. For simplicity, I will use the plural vocabulary. However, note that this is avoidable. The argument is not restricted to pluralities of objects.

⁹See (130) for a survey of the main semantic and metasemantic theories available.

concerning the objects in D is true, then platonism —about the objects in D — solves Benacerraf’s challenge or some mathematical objects in D exist (or have existed). By definition of fictionalism, the second option is no available. Therefore

If full fictionalism about the objects in D is true, then platonism about the objects in D solves Benacerraf’s challenge.

This is enough to cancel the main argument in favour of full fictionalism: the failure of platonism to solve the epistemological challenges. The argument for (*) will be not deductive. Hence, I will only claim plausibility for this conclusion.

Two comments are in order. First, the fact that a predicate P is meaningful on D is compatible with the fact that no object in D exists (or has existed). This is guaranteed by the “can” in its definition. This only requires that P expresses an intensional entity which *can* apply to the objects in D . It does not require that P applies to them as a matter of fact. Similarly, it only requires that P *can* be used to describe these objects. This ensures that to be meaningful on D does not rule out fictionalism concerning the objects in D by fiat. For instance, some predicates in *The Neverending Story* are meaningful on the collection of dragons. But, dragons never existed.

Second, (*) only involves predicates. However, the language of mathematics contains other kinds of expressions, such as operators expressing functions.¹⁰ Therefore, this conditional could seem unreasonably limited. If the argument does not work for other types of expression, the fictionalist might object that his position still works for them. Could fictionalism be the correct position concerning the objects affected by the other types of expression? The answer is negative for two reasons. First, the collection D on which a mathematical predicate P is meaningful is the same collection of objects where the other vocabulary is interpreted. For instance, the arithmetical operators “+” and “×” express total functions defined on the collection of natural numbers —or they are used to describe these objects. But, this is the same collection of objects on which the predicate “to be less than or equal to” is meaningful —according to the above definition. Consequently, if (*) works for some mathematical predicate P meaningful on D , then fictionalism about the objects in D will be in trouble no matter what happens with other kinds of expression used to talk about these objects. Second, the conditional (*) can be formulated in terms of other expressions. Again, take the case of operators. Let f be an n -ary operator. We start by defining the notion of an operator being meaningful on some collection of objects:

f is meaningful on $D \equiv_{df}$ (a) f expresses some intensional entity F which can hold on D or (b) f can be used describe the objects in D .

This definition has exactly the same form as before. Then, we state (*) using operations:

(*)’ If f is a mathematical operator meaningful on D , then: or platonism about the objects in D solves Benacerraf’s challenge or some objects in D exist (or have existed).

¹⁰By operator I mean a symbol expressing a mathematical operation. For example, the arithmetic operators “+” or “×”. To test the argument, let’s grant that these symbols not only form complex terms, but also express or describe functions.

Consequently, as soon as full fictionalism claims that some portion of mathematical language has standard content, this fact will turn against it. Moreover, this is so no matter the vocabulary used by the language.

3.4.2 Structure of the Argument

The argument has the following form. Let P be a mathematical predicate meaningful on some collection D of mathematical objects. First, I will argue that:

- (4.i) If (b) P can be used to describe the objects in D , then *It is plausible* that extensional approaches to semantics entail that platonism about the objects in D solves Benacerraf's challenge, or that some objects in D exist (or have existed).

Then, I will argue for its complementary:

- (4.ii) If (a) P expresses some intensional entity I which can hold on D , then *it is plausible* that intensional approaches to semantics entail that platonism about the objects in D solves Benacerraf's challenge.

This is enough for concluding that (*) is plausible. For, P express an intensional entity or not. In the second case, P will still be used to describe objects, but we are left only with extensional explanations of this phenomena. Thus, if this is the case, (4.i) entails that:

- (4.iii) If (b) P can be used to describe the objects in D , then *it is plausible* that some objects in D exist (or have existed).

In turn, if P expresses an intensional entity, this will be explained by means of intensional semantics. Therefore, (4.ii) entails that:

- (4.iv) If (a) P expresses some intensional entity I which can hold on D , then *it is plausible* that platonism about the objects in D solves Benacerraf's challenge.

Finally, (4.iii) and (4.iv) entail that:

- (4.v) If (a) or (b), then *it is plausible* that platonism about the objects in D solves Benacerraf's challenge, or some objects in D exist (or have existed).

But, by definition of "being meaningful on a collection D ", (4.v) is equivalent to (*). And, by distribution of plausibility over the material conditional, we have proved that (*) is plausible.

Clearly, the argument must not be taken as a deductive argument. Specifically, it only offers plausible support to (4.iii) and (4.v). One reason is that labels such as "extensional semantics" or "intensional semantics" are informal and too vague to be taken as part of a deductive argument. They constitute a clumsy classification of the main approaches in contemporary semantics. Below, I will stop and consider the main representatives of each of these two strands. Yet, there is no precise definition or description

corresponding to each semantic theory.¹¹ Therefore, I doubt that the argument can become a deductive argument. However, the plausibility of (*) is enough to put some pressure on full fictionalism.

3.4.3 Extensional Semantics

In this section I will argue for (4.i):

- (4.i) If (b) *P* can be used to describe the objects in *D*, then *It is plausible* that extensional approaches to semantics entail that platonism about the objects in *D* solves Benacerraf's challenge, or that some objects in *D* exist (or have existed).

Suppose that (b) —i.e. *P* can be used to describe the objects in *D*. This fact poses on us explanatory pressure: to explain how *P* acquires its meaning and how speakers acquire its competence concerning *P*. Let's see how the main *extensional approaches* to semantics explain both facts. Indeed, given the nominalistic preferences of fictionalism, this should favour this kind of approach over intensional alternatives. Here, I will only consider the following three theories of meaning: Davidson's (26; 27) truth-conditional semantics, causal theories of meaning (107; 72), and a use theory of meaning (63).

3.4.3.1 Truth-conditional Semantics

Davidson's truth-conditional semantics (26; 27) reduces the theory of meaning to the instances of the T-Schema

$$\varphi \text{ is true iff } \varphi.$$

According to it, the instances of the T-Schema partially explain speakers' linguistic competence, and this is enough to provide an account of the meaning of language.

Intuitively —and simplifying things—, the meaning of a sentence is what determines under which conditions the sentence is true or false. However, what are meanings exactly? The concept of *meaning* is an intensional concept. And, as such, it is troublesome. If meanings are objects “attached” to sentences, then they will be abstract objects —i.e., they will be out of spacetime. And, as in the case of mathematical objects, these are not accepted by everyone. Indeed, they come with a great deal of epistemological worries. To avoid this, truth-conditional semantics proposes to examine the explanatory role played by meanings, in order to perform it without meanings. Now, meanings are mainly used for two things. First, as I said, they determine the truth conditions of sentences. For instance, the meaning of the sentence “cats are mammals” determines that this is true if and only if indeed cats are mammals. On the contrary, if it were synonymous to the sentence “water is H₂O,” the fact that cats are mammals would not affect its truth value. Second, meaning explains the linguistic behaviour of speakers. These use the words in a very specific way depending on what the latter mean. For instance, speakers assert “this is a plant” when they are in front of a plant —and not in front of a book— precisely because of what “plant” means.

¹¹For instance, take possible world semantics. The most salient property of that theory is the use of possible worlds to define intensions. However, it is far from clear which other features such a theory must have. Indeed, there are a variety of different theories using possible worlds to achieve different objectives.

Consequently, if there is an explanation of sentence's truth conditions and speakers' linguistic behaviour that does not use the concept of meaning, this would avoid the difficulties posed by intensional objects. Thus, it would be preferable.

Davidson found that truth conditions are enough to explain speakers linguistic competence. Speakers assert sentences when their truth condition holds, and avoid to do it if this is not the case. Therefore, a theory of truth would be enough to account for the meaning of language in a safe way. And, the best theory of truth is that generated by the T-Schema.¹² As a result, these simple theory contains all the information required to pick what is important when speaking about meanings.

Let's go back to the argument. By assumption, the predicate P can be used to describe the objects in D . And, this entails that P features in some sentence whose truth condition involves these objects. That is,

$$P\tau_1 \dots \tau_n \text{ iff } \varphi(\alpha_1 \dots \alpha_n),$$

where $\alpha_1 \dots \alpha_n$ belong to D . Now, does this tell us something about these objects? No —aside the fact that they feature in the truth condition of " $P\tau_1 \dots \tau_n$." The T-Schema says nothing about the metaphysical status of the truth condition $\varphi(\alpha_1 \dots \alpha_n)$. More specifically, it does not entail that it obtains, or that some of the α_i exists. It only says that " $P\tau_1 \dots \tau_n$ " when this condition obtains. Therefore, this does not support what I want to prove —that some objects in D exist (or have existed).

Fortunately, this situation is not fatal for the argument. Indeed, it is to be expected. Truth conditional semantics is a *semantic* theory of meaning. A semantic theory of meaning describes what is the meaning of expressions and describes how the meaning of complex expressions results from the meaning of its components. Thus, semantic theories answer *what*-questions: what is the meaning of expressions? This kind of theory is descriptive. But, as I announced, I will articulate the argument around the explanatory pressure posed by LAC. This demands a different sort of theory, a *metasemantic* theory of meaning. Metasemantic theories explain how words acquire their meaning. The fact that words are endowed with meaning is not a metaphysical primitive. For instance, take the expression "to be a plant." The fact that this is a predicate that applies to objects which are plants is not among the primitive constituents of the world —if there are such things. Consequently, it must be explained. Thus, metasemantic theories answer *how*-questions: how do the expressions of language acquire their meaning and how do speakers acquire their linguistic competence? Truth-conditional semantics is just a semantic theory of meaning. It only offers a partial explanation of speakers' linguistic competence —it does not explain how speakers acquire their competence, or how sentences acquire their truth condition. Consequently, this theory must be supplemented with a metasemantic theory of meaning that meets the explanatory pressure posed by LAC. Here is where the causal theory and the use theory of meaning enter the picture. These are the main non-intensional candidates. In the following I will briefly outline each of them and show how they lead to the desired conclusion —i.e., they entail that some objects in D must exist (or have existed).

¹²Indeed, Davidson considers Tarski's truth conditions to be the best theory of truth. This adds to the T-Schema two additional semantic concepts: the truth condition of atomic sentences and quantified sentences is given in terms of reference and satisfaction. However, this is not required for the purposes of this sections. Thus, I will limit the exposition to the T-Schema.

3.4.3.2 Direct Theories of Meaning

Causal or direct theories of meaning explain how expressions acquire their meaning by appealing to strong counterfactual relations. They establish a link between events and speakers' uses of words. Take the case of names. The causal theory of reference (72) explains how speakers successfully refer by means of terms. Simplifying things a bit, the theory tells that speakers achieve reference when they are part of a causal chain of successful uses of a term. The chain has two salient features. First, each use of the term is causally connected with the previous uses. The most simple case is when a person experiences another person using the term successfully. This causal connection allows the former to also use it in right way. Second, the chain must have an initial element. This is an event consisting in the fixation of the term's referent—the “baptism” of this object. Note that this event must be causally effective. Similarly, in the case of predicates (107), reference to a collection of objects or a relation is explained by a causal relation between speakers and facts involving the elements of the collection or the relation. The connection be extended to complex expressions too. As a result, sentences relate to their truth condition through such counterfactual connections.

Nevertheless, the causal theory will not work in the case of mathematics. Speakers are not causally related to mathematical facts. The reason is that mathematical objects are abstract objects. They are “out” of space and time, and a causal connection with facts involving them is ruled out immediately. Indeed, it is highly doubtful that any form of counterfactual connection between mathematical facts and agents can be found. For this, we have to make sense of the idea of varying mathematical facts. But, this cannot be done if mathematical truths are necessary—in a metaphysical or conceptual sense. And there is reason to believe that this is the case.¹³ Besides, the abstract nature of mathematical objects and the physical status of persons make it difficult to think what kind of counterfactual relationship could be established directly between them.¹⁴ This kind of connection requires a strong coordination between the *relata*. And, the different metaphysical realms that people and abstract objects inhabit restrict the factors that could lead to such coordination.

Consequently, it is highly implausible that direct theories of meaning are effective for the purposes of full fictionalism and platonism. It is highly implausible to find a direct counterfactual relationship between mathematical facts and people. Thus, it is highly implausible that full fictionalism—and platonism—can use this sort of theories to explain how mathematical language acquire its meaning, or how speakers acquire their linguistic competence—i.e. to dispense with LAC. Besides, note that direct theories of meaning entail that the objects featuring in the truth conditions of sentences exist or have existed. Otherwise, no causal or counterfactual relationship could have been established between them and sentences. All this leave us with a unique option: use theories of meaning.

¹³See (74) for a compelling argument showing that, if mathematical sentences are true then they are necessarily true. Also, see (115; 116) and (82) for a metaphysical account of mathematical objects that makes mathematical facts necessary. In turn, (2) attributes a contingent status to mathematical objects.

¹⁴In Chapter 5 I will advocate Linnebo's (81; 82) and Rayo's (116) explanation of mathematical language. This is based on Stalnaker's (131) account of the limitations of possible-world semantics. It involves a counterfactual relation between the facts leading to the acquisition of mathematical language and speakers. But, this does not entail a counterfactual relation between mathematical facts and persons.

3.4.3.3 Use Theories of Meaning

There are several use theories of meaning (149; 13; 62). Each one of them differs from the others on important and sensitive points. To explain all these theories and their differences in detail would be too vast and general a task. It would go far beyond the scope of this text. Instead, I will explain what I consider to be the main characteristic shared by all of them. Namely, they attempt to explain how words acquire their meaning and how speakers acquire their linguistic competence using the concept of *use*. Then, assuming that intensional objects are ruled out, I will explain how this implies that, if the predicate P is meaningful on D , then some of the objects in D exist (or have existed).

Use theories of meaning explain both how words acquire their meaning and how speakers acquire their linguistic competence by telling some story about how speakers learn to use words. They are metase-mantic theories, in the sense explained above. Thus, *use* is the core concept of such theories. There are many ways of understanding this concept. *Use* may refer to a psychological action or a mental state. For example, the set of mental states that lead to the utterance of certain sounds. Alternatively, it can be understood from a behaviorist perspective, which would only take into account observable behaviour. However, the main use theories of meaning (149; 13) understand *use* as a category about communities. In this case, regardless of psychological or individual factors, the concept of *use* would be relative to a community of speakers —i.e. it would analytically entail concepts related to such an entity as a community. In general, if usage is understood as a concept relative to certain parameters, there are multiple solutions for these: community or context are two clear examples. Another important point is the status of meaning, and how use theories relate to a semantic theory. Here, there are two broad options. First, to reduce semantic theories to a theory of use—in a similar vein to truth-conditional semantics. Second, to accept meanings as *sui generis* entities and explain how they are attached to words through speakers use of these.

However it is understood, use refers to a normative action —i.e. an action subject to rules. Specifically, the concept of *use* entails a distinction between *right* and *wrong* uses of something. For example, consider the use that cycling attributes to bicycles. If someone were to throw a bicycle off a cliff, she would not be using it correctly according to the rules of this sport. Use theories of meaning attempt to explain how speakers acquire the ability to discern when the use of language conforms to a particular norm. The distinction between correct and incorrect uses is all I need for my argument.

Suppose that P is a predicate that can be used to describe the objects in the collection D . This entails that P features in sentences whose truth condition involves the objects in D . Without loss of generality, we can assume that these are atomic sentences:

$$P\tau_1 \dots \tau_n \text{ iff } \varphi(\alpha_1 \dots \alpha_n),$$

where $\alpha_1 \dots \alpha_n$ belong to D . Now, the question arises: how do use theories of meaning explain the assignment of this truth condition to the sentence “ $P\tau_1 \dots \tau_n$ ”? These will say something like the following. Speakers acquire the ability to identify what is the rule governing the use of cases —preferences, inkblots, etc.— of the predicate P . The rule says that the predicate P applies to certain objects $\alpha_1 \dots \alpha_n$ if and only if the assertibility condition of the resulting sentence “ $P\tau_1 \dots \tau_n$ ” is satisfied. And, the assertibility condition

at least implies that the truth condition $\varphi(\alpha_1 \dots \alpha_n)$ obtains. This ensures the correct connection between the sentence and its truth condition.

Sentences are connected to their truth condition through the rules governing the use of their component expressions.

Speakers can use a predicate P to describe some objects in D only if they have the ability to identify the rule governing the use of P . Then, the next question is: how do speakers acquire the ability to identify what is the rule governing the use of P ?

Since Wittgenstein's (149) and Kripke's follow-up this is a hot case of debate in the philosophy of language. My intention is not to enter into this debate. I do not want to elaborate or defend a precise theory of rule learning. Fortunately, what I need is something much more modest. Whatever the explanation of rule-learning, in the case of language it will involve *training*. Of course, once speakers are linguistically competent, the language can convey information about the rules of certain simple activities. For example, think of how the rules of a game like chess are learned. Usually, beginners are told what the rules are just before they start to play. However, the situation concerning language is not so simple. Before understanding the language, speakers cannot understand the information conveyed by it—indeed, here is where the complication of the matter lies. Therefore, the rules of language must be learned through other, non-linguistic channels. The point is that, whatever the correct theory of language rule learning is, it will imply that *some speaker has used the words according to the rules*—i.e., some speaker has used the words correctly.

Now, I am in a position to assert the crucial step: for someone to have used the predicate P correctly, some of the objects in D exist or must have existed. Recall that, by assumption, intensional entities are ruled out. The rule governing P states that P can be asserted of some objects $\alpha_1, \dots, \alpha_n$ when the assertibility condition of the sentence " $P\tau_1 \dots \tau_n$ " obtains. But, the assertibility condition at least entails that the truth condition $\varphi(\alpha_1 \dots \alpha_n)$ obtains too. To assert correctly the predicate P of some objects $\alpha_1, \dots, \alpha_n$ is to apply it with truth of those objects.

If P is correctly asserted of some objects $\alpha_1 \dots \alpha_n$, then the condition $\varphi(\alpha_1 \dots \alpha_n)$ obtains.

From this conditional and the fact that some speaker has used P correctly, it follows that $\varphi(\alpha_1 \dots \alpha_n)$ has obtained. Consequently, the objects $\alpha_1, \dots, \alpha_n$ in D exist or have existed. Indeed, this conclusion is quite intuitive. Again, take the case of bikes. Suppose that a community of speakers have been physically isolated from bikes—note this is weaker than requiring that bikes never existed. Moreover, suppose that intensional entities—such as modal operators, concepts, properties, etc.—are discarded as an epistemic source of beliefs and knowledge. In this situation, all connection between bikes and speakers has been severed. Do they have the opportunity to learn to ride a bike? No. They have never perceived a bike. There is no physical channel that transmits information from the latter to the community. Nor can they draw on a strong notion of possibility to take charge of *possible* bicycles. Under these conditions, the speakers could not even get an idea of what a bike is. Moreover, suppose that the drift of the universe led them to come across one. Then they would not know how to use it correctly according to the rules of cycling—not immediately at least. No more than a child who has seen many bikes, but has never

ridden one. Knowing how to ride a bike correctly —according to some rule— implies, firstly, that bikes exist or have existed. Secondly, that bikes have been used successfully once. Knowing how to use language correctly implies, firstly, that its components exist or have existed: words and the objects in the environment. Secondly, that words have been used properly.

3.4.3.4 Objections from Fiction

Somebody could object to this picture taking fiction —precisely— as a counterexample. Fictions exist and speakers learn how to use predicates to describe fictitious objects *that never existed*. Dragons, unicorns, distant planets and all sorts of creatures and mysterious places inhabit hundred of tales and people's imagination. Indeed, fictions —told and written— have always been an important part of learning a language. Similarly, Hamlet did not exist, but Shakespeare elaborated a precise *description* of him and his concerns. Does not all this prove that speakers learn to use predicates that have never been truthfully asserted of the objects they describe? I believe the answer is negative. This shows something more subtle. Take the case of dragons. People learn what “lizard”, “scales”, “wings” or “horn” mean. All of them are words that denote familiar objects in their environment. Then, using these words, speakers form complex expressions. For example, descriptions which talk about “non-existent objects”, such as dragons. Generally, once speakers are competent in the use of expressions which apply to objects in their environment, they can combine them in all sorts of forms to speak about fictitious situations. Witches, dragons, unicorns, and other mysterious creatures are grasped using *descriptions*. Similarly, fictional characters as Hamlet or Oliver Twist are grasped using *definite descriptions*. Descriptions as well as definite descriptions involve meaningful predicates used to depict common things, such as “to be winged”, “to be a reptile” or “to have horns”. Assume this situation is correct. Also, assume that abstract objects do not exist, and never existed. Can speakers form descriptions of them from the vocabulary used to depict non-abstract objects? I hardly think so. Abstractness —Chapter 2, 2.3.1— entails that physical properties cannot be attributed with truth to abstract objects. They are not winged, nor have horns, or are close to marvellous distant galaxies. Therefore, I believe it is implausible that any linguistic or psychological operation alone can form expressions that capture abstract objects from expressions about concrete objects —at least if abstract objects do not “collaborate in the equation.”

Nevertheless, for the sake of the argument, suppose this can be done. That is, suppose that abstract objects never existed, but speakers have expressive resources that allow them to grasp what they are. This is what full fictionalism demands. Also, let's leave aside sceptic arguments, such as Putnam's (111) model-theoretic argument or permutation arguments, against the sufficiency of linguistic and logical resources in fixing meaning. If speakers have such strong resources, then platonism can use them to explain mathematical reference and —plausibly— to dispense with Benacerraf's epistemological challenge. Indeed, this is what descriptive theories of reference propose —also, it can be read as a strengthening of Lewis' (77) reference magnetism explanation of reference. Suppose that speakers have meaningful descriptions of abstract objects. Moreover, suppose that abstract objects exist. Then, the platonist achieves reference to the latter because they are the objects that satisfy the description. Furthermore, if the use of such expressive resources is compatible with full fictionalism, surely it will be with platonism. For,

any metaphysical or semantic thesis concerning non-abstract objects compatible with fictionalism, will also be compatible with platonism. The former differs from the latter only in that it rejects Existence. Therefore, if full fictionalism can use linguistic resources to meet the explanatory pressure posed by LAC, then platonism can meet this explanatory pressure too. As a result, it is highly plausible that it can solve Benacerraf's epistemological challenge.

If intensional theories of meaning are rejected, then the direct —or causal— theory of meaning, the use theory of meaning, and the descriptive theory of reference are the main candidates to dispense with LAC: i.e., to explain how language acquires its meaning, and how speakers acquire their competence. But, given a mathematical predicate P that can be used to describe the mathematical objects in D , I have shown how the descriptive theory of reference allows platonism concerning the objects in D to solve Benacerraf's challenge. Also, I have shown how the use theory of meaning entails that some objects in D exist (or have existed). Moreover, the direct theory of meaning is plausibly useless for dealing with abstract objects. All this is enough to conclude (4.i):

(4.i)

If (b) P can be used to describe the objects in D , then *it is plausible* that extensional approaches to semantics entail that platonism about the objects in D solves Benacerraf's challenge, or that some objects in D exist (or have existed).

Consequently, the first step of the argument is achieved.

3.4.4 Intensional Semantics

In this section I will argue for (4.ii):

(4.ii) If (a) P expresses some intensional entity I which can hold on D , then *it is plausible* that intensional approaches to semantics entail that platonism about the objects in D solves Benacerraf's challenge.

Intensional theories of meaning explain meaning in terms of intensional objects, such as properties, relations, concepts, or propositions. The mark of identity of intensionality is that it does not allow for substitution *salva veritate*. In turn, this is the main feature of extensionality. Extensional contexts preserve the extension under substitution of expressions sharing the same extension. For example, the connectives of n -order classical logic are paradigmatic cases of extensional operators. Let H be an n -ary operator applying to sequences of expressions $e_1, \dots, e_n$ of the language.

Extensional operator H is an *extensional* operator if and only if for all expressions $e_{1_1}, \dots, e_{1_n}$ and $e_{2_1}, \dots, e_{2_n}$ such that each pair e_{1_i}, e_{2_i} , $0 < i \leq n$, has the same extension, then $H(e_{1_1} \dots e_{1_n})$ and $H(e_{2_1} \dots e_{2_n})$ also has the same extension.

For instance, if two sentences ϕ and ψ have the same truth value, then $\neg\phi$ and $\neg\psi$ —where “ $\neg$ ” is the negation of operator of first-order logic— also have the same truth value. On the contrary, intensional operators do not satisfy these conditions. That is, even if the e_{1_i}, e_{2_i} have the same extension, the extension

of $H(e_1 \dots e_{1_n})$ can be different to the extension of $H(e_2 \dots e_{2_n})$. For instance, consider contexts generated by believing.¹⁵ The terms “Brutus” and “Julius Caesar’s murderer” have the same extension, both refer to the same person: Brutus. Thus, the sentences “Brutus murdered Julius Caesar” and “Julius Caesar’s murderer murdered Julius Caesar” also have the same truth value —both are true. It is reasonable to expect a person who has only been told that Julius Caesar was murdered, but does not know about Brutus, to believe that Julius Caesar’s murdered Julius Caesar. However, this same person could not believe that Brutus murdered Julius Caesar —she lacks this piece of information. However, if this is so, how can the difference be explained? If there is no semantic difference between the expressions e_{1_i} , e_{2_i} —they have the same extension—, how can $H(e_1 \dots e_{1_n})$ and $H(e_2 \dots e_{2_n})$ differ in extension?

Intensional approaches to meaning explain this phenomenon by endorsing fine-grained semantic distinctions. While extensional approaches take the expressions e_{1_i} , e_{2_i} as being semantically equal —they have the same extension—, intensional approaches believe that the extensional ones are missing part of the whole story about the meaning of expressions. Specifically, they attribute additional semantic differences to the expressions e_{1_i} , e_{2_i} . Now, these share the same extension. Hence, intensional approaches have to expand semantics by adding additional semantic categories —intensional categories, such as properties, relations, concepts or propositions. Then, they associate entities $I_1 \dots I_n$ belonging to some of these categories with each one of the e_i . As a result, they are able to explain why $H(e_1 \dots e_{1_n})$ differs in meaning from $H(e_2 \dots e_{2_n})$: they differ because $I_{1_i} \neq I_{2_i}$ for some $0 < i \leq n$. Intensional operators have into account *intensions* —the objects falling under the new semantic categories.

Intensional operator H is an intensional operator if and only if for all expressions $e_{1_1}, \dots, e_{1_n}$ and $e_{2_1}, \dots, e_{2_n}$ such that each pair e_{1_i}, e_{2_i} , $0 < i \leq n$, has the same intension, then $H(e_{1_1} \dots e_{1_n})$ and $H(e_{2_1} \dots e_{2_n})$ also has the same intension.

Again, take the case of Brutus and Julius Caesar. Speakers can believe that Brutus $\neq$ Julius Caesar’s murderer because the intension associated with the name “Brutus” is different from the intension associated with the name “Julius Caesar’s murderer”. This difference transmits to the intensions attributed to the sentences $\varphi \equiv$ “Brutus murdered Julius Cesar” and $\psi \equiv$ “Julius Caesar’s murdered murdered Julius Caesar”, explaining the difference between believing that φ and believing that ψ .

Given the limitations of extensional semantics, it is worth asking whether intensional entities can be of help to full fictionalism. Can full fictionalism resort to the use of intentional entities to deal with LAC? I believe it cannot. Despite its limitations, I will use possible worlds semantics through the argument for simplicity. Possible world semantics analyzes intensions in terms of possible worlds and the extensions corresponding to them. According to this approach, the intension associated with an expression is identified by the extension of at all relevant possible worlds. Let I be an intensional entity and f a binary function assigning the extension of I at every relevant possible world w . Then,

$$I_1 = I_2 \text{ if and only if for all possible worlds } w: f(I_1, w) = f(I_2, w).$$

¹⁵Many of the phenomena related to belief can be treated as intentional phenomena. However, many others cannot. Beliefs are now considered to be hyperintentional phenomena.

Of course, possible world semantics owes us a metaphysical explanation of what possible worlds are. However, I will use possible worlds just as an heuristic device to clarify the argument on some points. Nothing of what I will say depends on the use of this formal tool.

3.4.4.1 Fictionalism and Intensions

Before arguing for the negative result, I want to say a few words about the fit between fictionalism and intensions. An immediate worry arises: intensions are abstract entities. Then, the use of intensional semantics is in tension with fictionalism. For, if the latter rejects the existence of mathematical objects to avoid the problems posed by their abstract nature, it is to be expected that fictionalism rejects the existence of *all* abstract objects. As a result, intensional theories of meaning would be as false as mathematical theories. For the sake of the argument, I will consider that intensions are available to the fictionalist. Moreover, it is perfectly coherent to be fictionalist about one kind of object, but not about others. For example, about mathematical objects, but not about concepts or propositions.

At last, I am in position to argue for (4.ii). Assume that P is a mathematical predicate expressing some intensional entity I which can hold on D . LAC requires an explanation of how speakers relate to intensions. There are two options. First, to elaborate an explanation of how speakers relate to I without involving other objects—let’s say, an explanation of the “direct” relation between speakers and I . Second, to elaborate an explanation that involves other objects than I . In what follows I will consider both options.

3.4.4.2 Direct Grasp

Suppose full fictionalism about the objects in D has at its disposal an explanation of the direct relation that speakers maintain with I —i.e., an explanation that only involves the speakers and this intension. An extremely simple—and bad— explanation is that speakers have a sixth sense that allows them to grasp these entities. Obviously, if this were so, the epistemology of modal concepts would not be a major philosophical problem.

Unfortunately, a direct explanation is not of help for full fictionalism. If this can solve the LAC using such an explanation, so can the platonist. Then, platonism will explain mathematical reference in the same way as it would do it using a descriptive theory of reference. First, it will explain how speakers grasp the intensions associated with the other expressions of mathematical language. Plausibly, if speakers directly grasp I , they will do the same with the other intensions expressed by mathematical vocabulary. Second, the platonist will stipulate that mathematical terms refer to the objects which satisfy the descriptions coached in the language. Finally, if platonism can explain mathematical reference, it is highly plausible that it can dispense with Benacerraf’s problem too.

3.4.4.3 Indirect Grasp

Let’s see what happens with the second option. Assume that objects other than I are required to explain how speakers grasp the latter. We have to consider three cases: (i) the grasp of I involves the objects it can describe—i.e., objects in its extension $f(I, w)$, for some possible world w ; (ii) the grasp of I does

not involve objects in the extension of I ; (iii) I is grasped through analytic relations to other intensional entities.

First, suppose the grasp of I is explained using some objects it can describe. Here, there are two options: these objects exist (or have existed) or they never existed. Suppose they never existed. If these objects never existed, then no counterfactual connection could have been established between them and speakers. Consequently, they can hardly be of help to indirectly grasp the intension I . What explanatory role can be played by entities with which the speakers have had no relationship? Simply, they play no role in agents' epistemological and linguistic actions. At this point, the notion of *non-existent* objects could come to the rescue. But, this would be in vain. Even if such entities are accepted in our metaphysical inventory, this does not change their status concerning agents: if some objects never existed, then they are not counterfactually related to agents and should not play any explanatory role concerning epistemology and language. On the contrary, if agents somehow were related to objects that never existed, this must be an epistemological or linguistic non-counterfactual relation. Therefore, the explanatory relation is the other way around. Agents would understand what non-existent objects are because they grasp enough concepts to describe them. For example, we can understand what Pegasus is because we have been told how Pegasus looks like. Similarly, we know what dragons are because there are tales and movies where dragons are depicted in great detail. Agents are related to Pegasus or dragons —if they are accepted as non-existent objects— because they grasp the concepts employed to describe them. Moreover, the category of *non-existent object* is metaphysically controversial. And, those explanations that use it will also be highly controversial —alternative explanations which dispense with them should be given preference. This is enough to discard non-existent objects.

Now, suppose the grasp of I is explained using some objects it can describe —i.e., some objects in $f(I, w)$, for some possible world w . Moreover, suppose these objects exist or have existed. The problem with this option is the same problem I opposed to the direct grasp of intensions. If fictionalism can explain the indirect grasp of I through some objects that exist or have existed, then so can platonism. Platonism accepts all the existential assumptions accepted by fictionalism. Unlike fictionalism, Platonism accepts all its existential commitments. Thus, if an explanation is acceptable to the former, it will be acceptable to the latter. Besides, using the indirect grasp of intensional objects, platonism can explain mathematical reference —it stipulates reference to mathematical objects using the appropriate definite descriptions— and dispense with Benacerraf's epistemological challenge. Consequently, the first option (i) leads to the desired conclusion. The second option (ii) faces the same problem.

What about the third option (iii)? Can agents' grasp of I be explained by means of some analytical entailment from other intensions $I_1 \dots I_n$? Suppose so. Then, agents' grasp of some I_i , $0 < i \leq n$, must be explained through non-analytic relations. On the contrary, speakers' understanding of I will be explained by means of an infinite descending chain of intensional entities. But this kind of situation is unacceptable in explanations of linguistic competence. It is highly implausible that in order to explain how agents grasp —for example— a concept, they need to make use of infinitely many concepts.¹⁶ Now, agents' grasp of I_i is direct or indirect. And this leads us back to the problem of previous explanations: if fictionalism can explain how agents —directly or indirectly— grasp I_i , then so can platonism.

¹⁶See Chapter 4, 4.5.1 for an argument against this situation.

The use of intentional entities puts the fictionalist in a bind, and not only because of their metaphysical status. However, this should not be a surprise. If full fictionalism about such entities explains the content of mathematical language and meets the challenge posed by LAC, then a stronger metaphysical position may also be able to do so. For example, platonism. This is sufficient to conclude (4.ii):

- (4.ii) If (a) P expresses some intensional entity I which can hold on D , then *it is plausible* that intensional approaches to semantics entail that platonism about the objects in D solves Benacerraf's challenge.

Finally, (4.i) and (4.ii) are enough to prove the desired conclusion:

- (*) If P is a mathematical predicate meaningful on D , then: or platonism about the objects in D solves Benacerraf's challenge or some objects in D exist (or have existed).

This is enough to pose full fictionalism in serious trouble. Full fictionalism entails that mathematical predicates P are meaningful on some collection of mathematical objects D . But, this and (*) entail that or platonism about the objects in D solves Benacerraf's challenge or some objects in D exist (or have existed). By assumption, the second case is forbidden. Therefore,

If full fictionalism about the objects in D is true, then platonism about the objects in D solves Benacerraf's challenge.

Nevertheless, the main arguments for fictionalism are the epistemological challenges opposed to platonism. Consequently, full fictionalism entails its own misfortune. If it is accepted, the main argument in its favour vanishes.

3.5 ALGEBRAIC FICTIONALISM AND THE CONSISTENCY CHALLENGE

The failure of full fictionalism does not immediately entail the failure of the fictionalist program. Fictionalism is not committed with any particular semantics. It just claims that mathematical objects do not exist and mathematical theories incorporate untrue sentences. It does not need to endow mathematical language with a standard semantics, nor with a platonic content. All it says is that mathematicians write untrue stories. These stories are summarized in the axioms of different canonical theories. Then, they devote all their efforts to find out what follows from these stories as well as to continue writing them.

Fictionalists view the basic assumptions of a pure mathematical theory (its axioms, or perhaps if we wish to follow Balaguer, its "full conception" of its objects) not as truths, but as generative of a fiction. Mathematical practice then becomes the practice of developing these mathematical fictions, by working out what does and does not follow from their generative assumptions. (76, 125)

Call this version of fictionalism *algebraic fictionalism*.¹⁷

¹⁷I take this label from (76) and (5), who distinguishes between formalist and non-formalist versions of platonism.

Fictionalism abandons the idea of strong objectivity. Unlike platonism, it denies the existence of a unique, true mathematical theory. And, mathematical theories need not be complete —indeed, incompleteness does not pose any pressure on fictionalism. Mathematics does not tell a closed story. Fictionalism poses a unique constraint on *classical* theories: consistency. It reduces knowledge about classical mathematics to knowledge about consistent fictions.

Since (classically, at least) inconsistent assumptions imply everything, only consistent fictions are mathematically interesting. So, as with other algebraic views, for fictionalism what makes a theory mathematically acceptable is consistency, and what makes a mathematical utterance appropriate in the context of mathematical theorizing is that it is a consequence of the assumptions of the theory in which it is put forward. (76, 125)

Therefore, knowledge about a mathematical sentence φ reduces to two things: (i) knowing that φ follows from a given mathematical, fictitious theory Γ , and (ii) knowing that Γ is consistent:

$$K(\Gamma \vdash \varphi + \Gamma \text{ is consistent}) =_{df} K_F \varphi.$$

In this way, fictionalism exchanges abstract objects for consistency. The question is: is this a good exchange for algebraic fictionalism? In what follows I will argue that the answer is negative.

(**) Algebraic fictionalism lacks an adequate concept of consistency to explain mathematical knowledge.

The argument goes as follows. There are two main concepts of consistency. On one side, the concept of consistency commonly used in model theory and proof theory is a mathematical concept. Although this is perfectly useful, for fictionalism it is only part of an untrue fiction. As a result, it cannot be used for the benefit of its epistemological explanations. On the other side, algebraic fictionalism resorts to a primitive concept of consistency. In this case, the problem is that this is a semantic concept. However, algebraic fictionalism does not attribute any content to mathematical language —this is the main difference between it and full fictionalism. The use of a primitive concept of consistency turns algebraic fictionalism into full fictionalism. But, as I argued in the previous section, this is an implausible account of mathematics. Therefore, algebraic fictionalism cannot make use of any concept of consistency. As a result, its explanation of mathematical knowledge fails.

Furthermore, full fictionalism and algebraic fictionalism exhaust fictionalism. Therefore, if both are in trouble, fictionalism as such is also in trouble. The arguments contained in the previous section and in this section are fatal to this philosophical position.

3.5.1 Mathematical Notions of Consistency

Consistency poses a problem on the fictionalist's reduction of mathematical knowledge (2, p. 99), (76, 126). Let's start with platonism. Platonism has no problem in dealing with consistency and other modal notions, such as possibility. Following usual model-theoretic definitions it can postulate a reductive account according to which for a collection of sentences to be consistent just is for it to have a model

—a model in classical first order logic is a non-empty set embedded with a collection of relations and operations.

Model-theoretic Consistency A set of sentences Γ is *model-theoretically consistent* if and only if there is a model M such that for all $\varphi \in \Gamma$, M is a model of φ ($M \models \varphi$),

where “ $M \models \varphi$ ” is defined as usual in terms of Tarski’s satisfaction relation. Alternatively, the platonist can use a proof-theoretical definition. According to this, a set of sentences is consistent just in case it does not entail a contradiction. Indeed, in classical logic a contradiction entails every sentence. Therefore, it is enough to require that the theory does not prove some sentence.

Proof-theoretical Consistency A set of formulas Γ is proof-theoretically consistent if and only if there is a sentence ψ such that $\Gamma \not\vdash \psi$.

By correctness and completeness, both definitions are equivalent in classical first-order logic.

The concepts of *model* and *proof in a formal system* are mathematical concepts. A model is a —non-empty— set embedded with a collection of relations and operations. In turn, a —finitary— proof is a number-theoretic recursive function. Therefore, the platonist can employ them to provide a perfectly good analysis of consistency. Consequently, knowledge about consistency as well as logical knowledge become entangled with mathematical knowledge. On the contrary, no reductive analysis of this kind is available for the sceptic. Fictionalism denies the existence of models and functions. So, any of the previous definitions would be false by fiat. Take the model-theoretic definition. According to the fictionalist, no set of sentences Γ has a model because no model exists. Generally, no set of sentences Γ would be consistent because all attributions of consistency would involve claims about non-existent objects. This is a highly unpleasant conclusion. We have strong intuitions about the consistency of—at least— some theories. For instance theories with a finite domain. Theories with finite domains are decidable, and PA proves their consistency. If the definition of consistency entails that no theory is consistent, this will be a bad definition. Besides, fictionalism reduces mathematical knowledge to knowledge about consistency. Then, if fictionalism is true, the model-theoretic and the proof-theoretic definitions of consistency entail that there is no mathematical knowledge. Certainly, algebraic fictionalism needs an alternative to explain consistency attributions.

3.5.2 Consistency as a Fiction

One option is to maintain the model-theoretic and proof theoretic definitions of consistency, while understanding consistency attributions as a case of applied mathematics. According to this, consistency would be just one piece of a cognitively useful fiction. It would be an important device to describe some empirical facts: the behaviour of mathematicians concerning a collection of statements. After all, there is no difference between this and how mathematical theories are used when doing physics or economy. Consistency attributions are untrue, let’s grant this. Also, they are valuable cognitive resources that—allegedly— do not add anything metaphysically relevant to the world. Therefore, to explain these attributions all we need is to explain how the mathematical theories explaining the concept of consistency

—model theory and arithmetic— are applied to describe and explain the mathematicians’ behaviour. Specifically, the process by means of which they accept and study a theory. Once this is done —either if mathematics are dispensable or they are nominalistically indispensable—, it will become clear what mathematicians are exactly doing when they recognize a theory as consistent.

This simple idea accords with fictionalism’s spirit: to deny the existence of mathematical objects, but not of mathematics. Mathematics can be used for many things, among them to explain mathematicians labour. Then, the fictionalist can resort on this particular application of mathematics to state her philosophical position concerning mathematical knowledge. There is nothing circular here. Furthermore, this should not be objected by saying that the fictionalist owes us —on pain of pragmatic incoherence— a non-mathematical depiction of her philosophical position. That is, a depiction that does not involve any mathematical vocabulary —in this case, the mathematical expression “consistency”. Indispensable applications of mathematics are compatible with nominalism (2, pp. 96-97).

However, this simple solution indeed leads to circularity. According to fictionalism, consistency is intended both as a criterion for theory choice and mathematical knowledge. Mathematicians know which theories are consistent, which theories are not, and decide to study the former. Then, according to the previous proposal, mathematicians’ choices and mathematical knowledge are explained by applying a particular theory T in which consistency is defined —model-theory or arithmetic. Consequently, the explanation of theory choice and mathematical knowledge is relative to the previous choice of T —which already involves consistency. Now, the problem is: how do we explain the choice and knowledge concerning T ? First, it is not self-explainable. No mathematical theory including Peano-Dedekind arithmetic as a subtheory proves its own consistency —if it is consistent. But, T includes arithmetic or model theory as a subtheory. Consequently, T does not prove its own consistency —if it is consistent. It will not be possible to explain the choice and knowledge provided by T by means of T itself. Thus, another theory T^* is required to explain the mathematical knowledge about T . To speak about consistency, T^* must include Peano-Dedekind arithmetic. Consequently, mathematical knowledge concerning T^* cannot be explained in T^* . This situation generates an infinite regress. In this case, the infinite regress is unacceptable. It is highly implausible that knowledge of the consistency of one theory T is equivalent to the knowledge of the consistency of infinitely many theories.

Moreover, the model-theoretic and the proof-theoretic definitions of consistency entail that consistency is a relative concept. Different mathematical theories entail different facts about which theories are consistent. For example, $ZF + \text{Mod}(ZF)$ —the sentence claiming that ZF has a model— entails the consistency of ZF , while ZF remains silent on this. However, facts about mathematical knowledge should not be relative to any mathematical theory. Consequently, the idea that consistency is a mathematical fiction is unworkable for fictionalism.

3.5.3 Primitive Consistency and the Semantic Nature of Language

Fictionalism cannot explain logical knowledge by means of mathematical theories. Indeed, the explanatory direction goes in the opposite way: using an appropriate notion of consistency, the fictionalist explains mathematical knowledge —i.e., knowledge of consistent fictions. Thus, the model-theoretic and

proof-theoretic definitions of consistency are useless for fictionalism's epistemological purposes. To explain mathematical knowledge through consistency, fictionalism needs a more robust and realist conception of the latter. Similarly, note that it cannot define consistency through modality either. For instance, it cannot stipulate that a theory is consistent just in case it is possible that all its sentences are true at the same time. There are two reasons for this. First, possibility, as well as consistency, is in need of explanation. Second, the fictionalist will be suspicious about some kinds of modality. More specifically, she will oppose any conception of *de re* modality because it would commit her with necessary properties or with *possibilia*—such entities do not fit very well in the nominalist's metaphysical inventory. Nevertheless, given the relation between consistency and possibility, the fictionalist could use the latter as long as it does not imply too strong metaphysical commitments. In sum, the fictionalist needs a non-reductive account of consistency and possibility.

Many authors (71; 39; 55; 2; 75), have advocated *intuitive* and *primitive* notions of consistency and logical possibility. Neither of these two concepts would be a mathematical concept—precisely because they are primitive, that is, they do not admit definition. To accept them is tantamount to accepting brute facts concerning consistency and modality.

So long as the logical notions are not defined in model-theoretic (or proof-theoretic) terms, but are instead considered as modal primitives, there is no principled Benacerrafian objection to our having knowledge of logical consistency, consequence, and validity. I have suggested that there are reasons even for the platonist to resist attempts to reduce these modal notions to non-modal alternatives. (75)

Fictionalists such as Field (39) and Leng (75) have used this kind of primitive concepts to secure fictionalism—i.e., to explain mathematical knowledge in terms of consistency or possibility. Suppose that speakers—including mathematicians—understand these primitive concepts and are able to handle brute facts involving consistency and modality. This is tantamount to attributing to them a *sui generis* logical knowledge. Mathematicians would have the ability to acquire knowledge about the logical properties of their theories. They will know which theories are consistent, which ones are not, and will devote their efforts to study the logical properties of the former—i.e., which sentences follow from them. Thus, the fictionalist can employ a primitive notion of consistency to state her philosophical proposal.

This solution seems promising. It would ensure that fictionalism has an appropriate concept of consistency to explain mathematical knowledge. However, here comes the problem. Despite its appeal, the solution is not available for all types of fictionalism. In particular, it is not available for algebraic fictionalism:

Algebraic fictionalism cannot use a primitive notion of consistency or logical possibility.

The point is that the primitive concepts of consistency and possibility are eminently *semantic*. Logical knowledge, including knowledge about these two phenomena, is knowledge concerning a relation of entailment. And, this is a linguistic relation. A set of sentences entails other sentences—alternatively, a collection of propositions entails other propositions. Now, *language* is a semantic category. Grammatical categories, such as the categories of names, predicates, or sentences, are already semantic categories.

Take the case of proper names. An object is a proper name because it performs an specific function: among other things, it attempts to refer to another object. Even if it does not succeed, its function is essential to its identity and its status as a proper name. For example, consider the name “Aristotle”. The spot on the screen is just that, a spot. The sound emissions of this name are just that, sound emissions. What makes “Aristotle” a proper name is its semantic and communicative functions: it attempts to denote a very specific object, Aristotle. The same applies to predicates. Consider the predicate “to be a Greek philosopher”. As in the previous case, this predicate has a number of complex semantic and communicative functions. Among them, it applies to those objects which are Greek philosophers. If it did not perform this function, then it would not be a predicate—it would be another object, such as a stain or a sound. Although not all grammatical distinctions are logically relevant, they introduce semantic or communicative distinctions. Thus, the point is: it makes no sense to vary the semantic content of a name, a predicate or a sentence without varying its identity.

If the linguistic expressions e_1 and e_2 perform different semantic functions, then $e_1 \neq e_2$.

Thus, let us imagine that there were two expressions e_1 and e_2 with the same spelling, “Aristotle”. However, while e_1 would refer to an object—Aristotle—the other would perform a very different function: the same function that the predicate “to be a Greek philosopher” has in English. In this situation, while the first would be a name, the second would be a predicate.

This is where the problem lies. The semantic nature of language puts the algebraic fictionalist in a bind. Algebraic fictionalism does not attribute any content to mathematical language. This is precisely where it differs from full fictionalism. Therefore, if the former denies the semantic nature of mathematical language, then it is denying that mathematicians speak any language at all. In principle, this need not be a problem—although it would complicate things for fictionalism. It could be the case that mathematical practice is linguistic only in appearance. Mathematical practice may be linguistic in appearance only. At first glance, mathematicians appear to be talking, asking questions and making assertions. In reality, however, they would be doing something quite different. Although this “alien” view of mathematics is coherent, it is hardly plausible: no one will accept it. In addition, the real problem for algebraic fictionalism is other. As I said, primitive notions of consistency and possibility are eminently semantic. Logical knowledge is knowledge about the entailment relation, thus about a linguistic relation between sentences—i.e., between semantic entities belonging to a particular language. This leads to the following conclusion: if mathematicians do not speak any mathematical language, then they cannot have logical knowledge about the consistency—in a primitive sense—of theories stated in such a language, nor about what is mathematically possible. Consequently, algebraic fictionalism cannot explain mathematical knowledge by using a primitive notion of consistency or possibility.

Can algebraic fictionalism object to this challenge in any way? Let me briefly explore two options. One option is to attribute some kind of semantic content to mathematical language. As a result, it will be endorsing a semantic concept of language and will be able to attribute logical knowledge to mathematicians. But, what kind of content can it attribute to mathematical language? No kind of fictionalism can resort to paraphrase functions to explain mathematical language. Otherwise, it would lose its identity. For instance, if fictionalists accept Hellman’s (55) modal-structuralist translation of mathematics, they

still can deny the existence of mathematical objects. However, they will do something more: additionally they read mathematical language in a highly non-standard way. Whoever does so will not be a fictionalist, but a modal-structuralist *à la* Hellman. The unique option for algebraic fictionalism is to endorse mathematics with standard, platonic semantic content. Consequently, algebraic fictionalism collapses into full fictionalism. But, in the previous section I argued that full fictionalism undermines its favour.

A second option is to introduce a formal concept of language. There are perfectly good syntactic notions of language, formula and theory used in formal logic. None of these is —or presupposes— semantic categories. Therefore, could algebraic fictionalism use such formal languages to explain mathematical language and mathematical knowledge? The answer is no. Formal languages are mathematical objects. They are studied in model theory, proof theory and other branches of mathematics. At best, they are useful models —in the scientific sense of this word— which allow some properties and relations of the natural language to be studied mathematically. For example, entailment and consistency. Consequently, formal languages cannot be used by algebraic fictionalism to explain mathematical knowledge. First, when the logician claims that a collection of formulas is consistent this is employing the *model-theoretic* or the *proof-theoretic* notion of consistency. But, fictionalism needs primitive concepts of entailment and consistency. Second, the expressions of formal languages are not *expressions* in the sense employed by the linguist. They are expressions just *by stipulation*. For instance, the terms of formal languages “refer” to an object because there is a mathematical function that assigns the latter to the former. But, they do not perform a semantic function as names of natural language do.

This leaves algebraic fictionalism with no options. In order not to collapse into full fictionalism, it cannot use a primitive concept of consistency or possibility to explain mathematical knowledge either. Now, all relevant notions of consistency reduce to the model-theoretic concept of consistency, the proof-theoretic concept of consistency, or a primitive concept of logical consistency. None of this is available to algebraic fictionalism. This is enough to conclude (**):

(**) Algebraic fictionalism lacks an adequate concept of consistency to explain mathematical knowledge.

Consequently, algebraic fictionalism fails as an account of mathematics.

3.6 CONCLUSION

In this chapter, I have analyzed fictionalism, the thesis that mathematical theories are fiction and that mathematicians create and study what happens in these fictitious stories. I argued that fictionalism fails as an account of mathematics for three reasons. First, it cannot explain mathematical reliability. Therefore, it fails to solve Benacerraf’s reinforced version. Second, I distinguished two types of fictionalism *full fictionalism* and *algebraic fictionalism*. Then, I argued that if full fictionalism can explain how mathematical vocabulary acquires its content and how mathematicians acquire their linguistic competence, then so can platonism. This is fatal for the former. Finally, I argued that algebraic fictionalism lacks a suitable notion of consistency for reducing mathematical knowledge to logical knowledge. For this, it needs to explain how mathematical vocabulary acquires its content and how mathematicians acquire their

linguistic competence. However, this reduces it to full fictionalism. In conclusion, fictionalism must be complemented by an alternative explanation of these two phenomena.

4 Realist Pluralism

Multiversism

4.1 INTRODUCTION

In this chapter, I will analyze multiversism. Multiversism contends that there is no single universe of mathematical, abstract objects, but many of them. In this way, multiversism is a new, pluralist form of mathematical platonism. It accepts that all those objects populating the multiverse are abstract objects which exist independently of us. Thus, the original mathematical realm is replaced by the multiverse. Moreover, the plural nature of multiversism allows it to reduce mathematical knowledge to logical knowledge. This is so because it establishes a close connection between logic and ontology: every consistent theory is true in some universe. In this way, multiversism rescues platonism from Benacerraf's challenge: mathematical knowledge is knowledge about the consistency and consequences of mathematical theories.

First, I will explain that there are many different types of multiversism. Among them, Hamkinsian multiversism and full-blooded platonism are the most extreme: they impose no limits on the extent of the multiverse. This is why they can reduce mathematical knowledge to logical knowledge. Second, I will explain the two main arguments in favour of multiversism —apart from its ability to solve Benacerraf's challenge. Both arguments are based on the limitation theorems affecting contemporary mathematics. Third, I will argue that the more radical versions of multiversism are not good accounts of mathematics. Thus, I will argue that Hamkinsian multiversism cannot explain how mathematical vocabulary acquires its meaning, contrary to what it says. Fourth, I will argue that full-blooded platonism lacks an appropriate notion of consistency for explaining mathematical knowledge. The reduction of mathematical knowledge to logical knowledge presupposes the linguistic competence of mathematicians. However, full-blooded platonism fails to explain how mathematical vocabulary acquires its meaning and how mathematicians acquire their linguistic competence. Finally, I will argue that on the face of a realist ontology, the reduction of logical knowledge to mathematical knowledge imposes unreasonable restrictions on the latter.

4.2 KNOWLEDGE IN THE MULTIVERSE

Anti-realist alternatives to mathematical platonism claim to solve Benacerraf's challenge by weakening Responsiveness until the factors on which mathematical knowledge is responsive become epistemically tractable. For example, fictionalism achieves this by reducing mathematical knowledge to logical knowledge. As a result, fictionalism gives up a strong notion of mathematical objectivity. No mathematical

theory is *the* correct theory from a semantic or metaphysical point of view. Any consistent theory is acceptable as a fiction. In this chapter, I will focus on a family of realist accounts of mathematics that also weaken Responsiveness and objectivity: multiversist accounts of mathematics. Being a realist position about mathematical objects, multiversism is a form of platonism.¹ It accepts that mathematical theories are about abstract objects which exist independently of us —i.e., it accepts Existence, Abstractness, and Independence. At the same time, multiversism is a plural approach to mathematics —as fictionalism. It denies that there is a single, well-determined, and maximal realm of mathematical objects: *the* mathematical universe. This clashes with more traditional forms of platonism, which do accept a strong degree of objectivity. According to the latter, mathematical terms and quantifiers are interpreted in the mathematical universe. Then, mathematicians aim to obtain a suitable description of this universe. That is, the universe guarantees the existence of a privileged mathematical theory, the true mathematical theory, and the task of the mathematician is to identify it. In turn, multiversism rejects the idea of such a privileged theory: as for fictionalism, no theory is the correct theory from a semantic or metaphysical point of view. Recall that most mathematical theories can be interpreted in set theory. Thus, the opposition between universism and multiversism can be expressed as follows —indeed, the main discussion of multiversism takes in set theory:

Set theoretic Universism There is a unique, determined and maximal universe of sets where the set-theoretic terms and the set-theoretic quantifiers are interpreted. *The* universe is what set-theoretic concepts are about as well as what set-theorists seek to depict.

Set theoretic Multiversism There is a multiplicity of metaphysically equivalent universes of sets, each one of them providing an appropriate interpretation for set-theoretic terms and quantifiers. Accordingly, there are several concepts of set corresponding to the variety of universes and set theorists, instead of depicting one, privileged universe, seek to understand what is true in different universes and how they relate to each other.

Similarly to fictionalism, multiversism attempts to solve Benacerraf's challenge by weakening Responsiveness

Responsiveness $C\varphi + \varphi + X \Vdash K\varphi$,

—where $C\varphi$ expresses some counterfactual relation between agents and the facts depicted by φ , and “ X ” is an unknown preventing any simplistic definition of knowledge— in the same way. According to them, mathematical facts do not play any significant role in the explanation of mathematical knowledge. Instead, what they require is to replace Responsiveness with the following “equation”

$X \Vdash K_M\varphi$

and solve X with a treatable epistemology of mathematics. Then, multiversism solves X , once more, with logical facts concerning consistency and entailment. Consequently, knowledge about a mathematical

¹Indeed, there are anti-realist forms of multiversism. However, I believe they can be understood as versions of fictionalism. For more on the different types of multiversism, see the next section.

sentence φ reduces to two things: (i) knowing that φ follows from a given mathematical theory Γ , and (ii) knowing that Γ is consistent:

$$K(\Gamma \vdash \varphi + \Gamma \text{ is consistent}) =_{df} K\varphi^M,$$

where —informally— φ^M means that φ is true in some universe M . Thus, to know that “ $7 + 5 = 12$ ” is true in some universe it is enough to show that it follows from some theory that is relatively consistent to arithmetic —for instance, arithmetic is that theory. And, to know that CH is true in some universe it is enough to know that it follows from some theory relatively consistent to ZF. As can be seen, the strategy is the same as that used by fictionalists: reject any counterfactual connection between mathematical facts and subjects to explain mathematical knowledge. Thus, by grounding mathematical knowledge in logical facts, mathematical facts become accessible to human beings.²

Additionally, mathematical practice and methods become extremely important for some forms of multiversism —for instance, Hamkinsian multiversism. Specifically, the particular procedures used to obtain independence results in set theory. Nevertheless, these procedures consist of model-theoretic constructions involved in proofs of relative consistency —inner models constructions, set forcing, class forcing, ultraproduct constructions, etc. What is relevant, therefore, are the mathematical practices used to acquire knowledge about consistency and entailment. These concepts have become an integral part of contemporary mathematics. Fictionalism and multiversism coincide in this, since both positions are based on contemporary mathematical practice.

The main difference between multiversism and fictionalism is that multiversism can explain knowledge about true mathematical sentences taken at face value. Even though mathematical knowledge is grounded on —supposedly— weaker, logical facts, this is genuine mathematical knowledge. As a result, the multiversist has no problem explaining mathematical assertion. Multiversism can accept that mathematicians assert simple arithmetical statements —such as that $7 + 7 = 12$ — at face value. Moreover, when a mathematician knows that $7 + 5 = 12$, she earns the right to assert truthfully “ $7 + 5 = 12$ ” — per the K-Rule. Consequently, multiversism is in a position to solve the challenge posed by the explanation of the reliability of mathematicians. Unlike fictionalism —Chapter 3, 3.3.5—, it can solve X in the following argument-schema:

- (2.i) Multiversism can explain why mathematicians are highly reliable when gathering knowledge about X .
- (2.ii) If (2.i) is the case, then multiversism can explain why mathematicians accept a mathematical sentence φ only if φ accommodates to X .

Then,

- (2.iii) Multiversism can explain why mathematicians accept mathematical sentences φ only if φ conforms to X .

²Despite mathematical knowledge is explained by logical knowledge, mathematical facts do not become logical facts. Consistency does not enter in any truthmaking relation with mathematical theories, i.e., mathematical theories are not true *because* they are consistent.

Moreover,

- (2.iv) According to multiversism, all sentences φ conforming to X are true—at least φ is true relatively to some universe of objects.

Therefore, from (2.iii) and (2.iv),

- (2.v) Multiversism can explain why mathematicians accept mathematical sentences φ only if φ is true—at least relatively to some universe of objects.

Unlike fictionalism, multiversism simply solves X with properties such as being consistent, being derivable from a consistent theory, or related properties—for instance, being entailed by some model-theoretic construction. Because of its realism and its pluralism, (2.iv) turns to be true—any sentence satisfying such properties will be true relatively to some universe of sets. On this point, multiversism scores better than fictionalism.

4.3 MULTIVERSIMS

There are several versions of multiversism—see (19) for a cartography of the landscape they form. As I said before, multiversism is a proposal that has arisen within set theory. Moreover, classical mathematical theories are interpretable in set theory. Therefore, from now on I will focus on set-theoretic multiversism. A first case of multiversism is (154), who has been read as a *height* or *vertical* multiversist—he has also been read as a *potentialist* (19). According to height multiversism, while the concept—or better concepts—of set does capture facts about the width of the hierarchy of sets up to isomorphism, it does not uniquely capture its height.

Height Multiversism: There is a multiplicity of metaphysically equivalent universes of sets *varying in height, but not in width*, each one of them providing an appropriate interpretation for set-theoretic terms and quantifiers.

Accordingly, there are several concepts of set corresponding to this variety of universes. Set theorists, instead of depicting one, privileged universe, seek to understand what is true in different universes *varying in height*, and how they relate to other universes. For instance, CH poses a question about the width of the hierarchy—i.e., a question about how many sets there are in its stages. According to height multiversism, all universes tall enough to introduce the reals agree on this question. Consequently, height multiversism determines a definite answer to CH—even though some universes have more layers than others, they share the same structural properties concerning each layer.³

³Zermelo proved that second order Zermelo-Fraenkel set theory (ZF^2) is *quasi-categorical*. This means that for all models $V_1 = \langle V_\alpha, \in \restriction V_\alpha \rangle$ and $V_2 = \langle V_\kappa, \in \restriction V_\kappa \rangle$ —where α and κ are inaccessible cardinals—, then there is an embedding of V_1 in V_2 or there is an embedding of V_2 in V_1 . Thus, while V_1 and V_2 diverge in the number of stages of the hierarchy—one has α stages while the other has κ stages—and thus on its height, both share the same stages up to α or κ . As a result, ZF^2 fixes all questions about the width of $V_{\omega+3}$ —the stage where relations among reals appear—and, thus, semantically determines CH positively or negatively. However, ZF^2 is incomplete and does not entail either CH or its negation. In particular, for all sets a inhabiting both V_1 and V_2 , $\wp(a)$ is the same in both structures.

The dual position of height multiversism is *width* or *horizontal* multiversism. According to this, the concept—or better concepts—of set fixes the height of its models, but not their width.

Width Multiversism There is a multiplicity of metaphysically equivalent universes of sets *varying in width, but not in height*, each one of them providing an appropriate interpretation for set-theoretic terms and quantifiers.

Accordingly, there are several concepts of set corresponding to the variety of universes and set theorists, instead of depicting one, privileged universe, seek to understand what is true in different universes *varying in width* and how they relate to other universes. Unlike height multiversism, for width multiversism there are universes which verify CH and universes that verify its negation.

Additional versions of multiversism arise from the combination of the two previous types —i.e., height and width multiversism. For instance, the multiverse view advocated by Hamkins (54) combines both height and width multiversism. His rejection of a fixed width is based on Quine's (114, 66) criticism of second-order logic. Second-order logic would be part of set theory, and is affected by the relativism inherent to the concept of set.

The point is that a second-order categoricity argument, even just for the natural numbers, requires one to operate in a context with a background concept of set. And so although it may seem that saying “1, 2, 3,... and so on”, has to do only with a highly absolute concept of finite number, the fact that the structure of the finite numbers is uniquely determined depends on our much murkier understanding of which subsets of the natural numbers exist. (54, p. 428)

Consequently, the quasi-categoricity result obtained by Zermelo (see note 3) does not explain how the concept of set fixes the width of the universe. On the contrary, it only tells us what holds on a particular subclass of universes inhabiting the bigger multiverse.

A more fine-grained classification of multiversist positions arises when the ontology underlying them is also taken into account. As I said at the beginning, I will limit the exposition to *realist* versions of multiversism. According to these, the objects inhabiting the universes and the universes themselves, if they can be understood as objects, are abstract objects that exist independently of agents—in the sense of Independence Chapter 2, 2.2.

Realist Multiversism The objects inhabiting the universes and the universes—if “universe” expresses a concept characterizing a particular kind of object—are abstract objects that exist independently of agents.

Realist versions of multiversism constitute a special form of platonism. They accept the three theses composing this philosophical account of mathematics: Existence, Abstractness, and Independence. But, at the same time, realist multiversism rejects the traditional belief in a single realm of mathematical objects. As a result, multiversism introduces a new division inside platonism: while traditional forms of platonism are *monist*, the former is a form of *pluralist platonism*. Indeed, the seminal works which gave birth to multiversism (2; 54) overtly declare it as a new form of mathematical platonism.

[Multiversism] holds that there are diverse distinct universes [...]. Each such universe exists independently in the same Platonic sense that proponents of the universe view regard their universe to exist [...] The multiverse view is one of higher-order realism-Platonism about universes —and I defend it as a realist position asserting the actual existence of the alternative set-theoretic universes into which our mathematical tools have allowed us to glimpse. (54, 416-417)

The plural spirit of multiversism allows it to combine the lightweight approach to mathematical knowledge advocated by fictionalism, and a realist ontology. To understand the force of this, consider any incomplete mathematical theory —for instance ZF. We know that if ZF is consistent, ZF+CH and ZF+¬CH are consistent too —CH is independent of ZF. Now, let us return to the template schema needed to explain mathematical reliability —Chapter 2, 2.3.2.1. As fictionalism, multiversism solves X with the property of being consistent —or similar properties. Thus, ZF+CH and ZF+¬CH accommodate to X . But now, unlike in the case of fictionalism, (2.iv) turns out to be true, ensuring that both theories are true relative to some universe of sets. Multiversism exchanges fiction for universes. As in the case of the former, there is no single metaphysically or semantically correct universe. However, they are not fictional entities. This exchange allows multiversism to preserve truth, assertion, and to solve many of the problems faced by fictionalism.

In turn, *anti-realist* versions of multiversism deny —or at least refrain from asserting— that sets and universes exist in any metaphysically loaded notion of existence.

Anti-realist Multiversism The objects inhabiting the universes and the universes —if “universe” expresses a concept characterizing a particular kind of object—, do not exist.

The supporter of this view regards the multiverse discourse as a useful tool to describe the practice of contemporary set theorists. They spend a lot of time investigating which theories are consistent and which are not, and what follows from them and what does not. And all the resulting theories are perfectly acceptable from the point of view of set theorists. Indeed, they also make use of the multiverse —entering the region of meta-practice— to guide their research about consistency and to produce new proofs of independence. According to this view, a universe is no more than the correlate of a theory proved to be consistent by some well-established model-theoretic method. Some of these universes —i.e., consistent theories— fit better with our informal understanding of sets as well as with additional criteria for theory choice —aesthetics, simplicity, or the broadness of the resulting concept of set (126). Otherwise, “all the universes” are metaphysically at the same level: they do not exist.

Some antecedents of this view are Hilbert’s *coherentism* (57) or Carnap’s pluralism (20), who advocated a deflationary reading of existence.⁴ Recently, Shelah (126) has explicitly advocated an anti-realist

⁴In “On the Infinite” Hilbert differentiates between *real* or *finitely verifiable* statements and *ideal elements*. The former are the statements corresponding to speakers’ informal understanding of arithmetic. Hilbert took them to be provable employing only finitist resources, in some unspecified sense of “finitism” —see (138, Ch.1-3) for an attempt to precise the concept of *finitism*. Similarly to observational sentences in empirical science, Hilbert took these as “verifiable” in the previous sense. In turn, ideal elements are sentences that are not finitely verifiable, usually quantified sentences — Π_n^0 and Σ_k^0 sentences for $n \leq 1$ and $k \leq 2$ — purporting to make claims about in-

conception of the multiverse. According to him, multiversism is a false theory, and consistent extensions of ZF are not true in some domains of set-like objects. Since multiversism and fictionalism elaborate a very similar explanation of mathematical knowledge, I believe that anti-realist multiversism can be assimilated into the latter. While realist multiversism rejects the sceptical ontology of fictionalism, anti-realist versions of multiversism accept it. Thus, universes can be given the same treatment as fiction. The only difference between Shelah, on one side, and Hilbert and Carnap, on the other, is that the latter are *non-factual*. They deflate existence inside of mathematical discourse, remaining neutral regarding the existence—in an ontologically loaded sense—of mathematical objects. On the contrary, Shelah, like fictionalists, positively asserts that sets do not exist.

I will focus on realist versions of multiversism—specifically, on Balaguer’s (2) and Hamkins’ (54) proposals. Thus, I will focus also on those versions which combine height and width multiversism. These proposals use minimal, epistemologically treatable resources to explain mathematical knowledge. Remember, mathematical knowledge is reduced to logical knowledge—knowledge about consistency and entailment:

$$K(\Gamma \vdash \varphi + \Gamma \text{ is consistent}) =_{df} K\varphi^M,$$

where—informally— φ^M means that φ is true in some universe M . This minimisation simultaneously leads to a maximisation of mathematically acceptable theories: all consistent theories—or the theories proven to be consistent by model-theoretic methods—will be acceptable. Further limitations on acceptable theories, or on the extent of the multiverse, will be the result of additional constraints imposed on the explanation of mathematical knowledge. That is, logical knowledge will not be sufficient to explain it. Consequently, radical pluralism will be limited, reintroducing a concept of mathematical objectivity—although a minimal one—stronger than logical objectivity. This is precisely what I will argue for in Chapters 5-7. The exact relation between the proposal advocated there and height and width-multiversism remains open. In this chapter, I will argue in favour of this conclusion: that is, I will show why the extreme pluralism of Hamkins’ and Balaguer’s positions is problematic.

4.4 MOTIVATING THE MULTIVERSE

Before going into the critique, I will explain which are the main reasons in favour of multiversism. There are two broad families of arguments. First, the promised solution to Benacerraf’s epistemological problem—including its strengthened version. The rejection of responsiveness cancels the challenge. Then, multiversism positively explains mathematical knowledge through logic. The second family of arguments consists of arguments against universism. Since these are opposed positions, showing that universism has a problem is equivalent to showing that multiversism has an advantage. Otherwise, realist multiversism

finitely many objects. He understood them analogously to the instrumentalist’s understanding of non-observational sentences: as ontologically neutral devices employed to move from real statements to real statements. Thus, ideal elements are just a convenient tool for satisfying mathematicians’ pragmatic objectives—for instance, to simplify reasoning in mathematics. Aside from its usefulness, all that is required is that ideal elements are consistent with the theories composed of finitely verifiable sentences.

faces the metaphysical debate against anti-realist accounts of mathematics, such as fictionalism. The advocate of multiversism must provide additional reasons to support her favourite against the latter. Some authors —(2, Ch. 8), (94, Ch. 4)— have argued that this is an unsolvable dispute: there is and could be no facts to determine which is the metaphysically correct conception of mathematics. However, the arguments exposed in Chapter 3 suggest that this is too hasty a conclusion. For example, multiversism accounts for mathematical reliability, while fictionalism does not. Thus, the former scores better on this point. Nevertheless, I will not enter into the metaphysical debate here. Instead, I will expose the main arguments opposed to universism.

According to multiversists, universism is false because it cannot explain how mathematical terms and quantifiers become interpreted in the metaphysically privileged universe of mathematics. Equivalently, this can be put in terms of intensional entities —if wording about these entities is allowed— saying that universism can not explain how mathematical predicates and functors express unique, well-identified concepts, relations and operations which are about the objects in *the* universe. More specifically, the multiversist endorses

- (4.i) The logical and linguistic features of mathematical theories do not fix —up to isomorphism— a unique universe of objects where terms and quantifiers are interpreted,

—in terms of concepts: the logical and linguistic features of mathematical theories do not fix unique concepts, relations or operations holding among the objects in *the* universe.

In addition, the multiversist advocates

- (4.ii) Mathematical reference can only be explained using descriptive resources, specifically using the logical and linguistic properties of axioms coached in mathematical language.

The reasons for (4.ii) are both metaphysical and methodological. First, Abstractness and Independence claim that mathematical objects are abstract objects which exist independently of agents. All counterfactual connections with facts involving them are forbidden. Therefore, no connection of this kind plays a relevant role in explaining reference and meaning acquisition. The only relevant connections left to be used in metasemantic explanations are purely descriptive —i.e., non-causal connections based on the semantic properties of the mathematical language. These properties, together with the target objects, should impose sufficient constraints to achieve reference to them. Moreover, contemporary mathematical practice is eminently *algebraic* —(56),(76, 120),(35, Ch. 6) According to this, mathematicians identify a set of axioms and study what follows from them. The axioms are intended as a description of the objects under consideration: any consistent system of axioms contextually defines its subject matter. Thus, the role of axioms can be compared to the role of a system of equations with several unknowns: together they determine the collection of acceptable solutions to the system —the interpretation of mathematical vocabulary. In consequence, the descriptive resources relevant to mathematicians are those used in the axioms of their theories.

From (4.i) and (4.ii) immediately follows that:

(4.iii) If (4.i) and (4.ii), then it is not explainable how mathematical terms refer to a privileged universe —or how concepts, relations and operations holding among the objects in this universe become uniquely expressed by the mathematical vocabulary.

If reference and semantic content are not uniquely determined by logical properties of axiom systems —according to (4.i)— and no other resources are available to explain reference —as (4.ii) demands—, then it is simply not explainable how mathematics is only about a privileged universe of mathematical objects. Now, we have that:

(4.iv) If it is not explainable how mathematical terms refer to a privileged universe —or how concepts, relations and operations holding among the objects in this universe become uniquely expressed by the mathematical vocabulary—, then universism is false.

But, from (4.i), (4.ii) and (4.iii) it follows that:

(4.v) It is not explainable how mathematical terms refer to a privileged universe —or how concepts, relations and operations holding between the objects in this universe become uniquely expressed by the mathematical vocabulary.

Consequently, from (4.iv) and (4.v), the multiversist concludes that:

(4.vi) Universism is false.

The unique premise of the argument that remains to be justified is (4.i). There are two main arguments supporting it —both are based on the methodological procedures used in meta-mathematics: *the argument from logical independence* and *the argument from categoricity*.

4.4.1 The Argument from Logical Independence

The argument from logical independence builds on independence results —see 4.7 for a brief survey of the incompleteness affecting set theory. Zermelo-Fraenkel set theory (ZF) is highly incomplete. Indeed, it is incomplete at very low levels of the hierarchy. There are several Π_1^0 -sentences —as its Gödel sentence G_{ZF} or $\text{Con}(ZF)$ — that the theory does not decide. All these are examples of a *benign* kind of incompleteness: once we are justified in accepting Peano-Dedekind arithmetic (PA) the independence of Π_1^0 -sentences is enough for their truth. Moreover, the combination of *intrinsic* and *extrinsic* sources of evidence makes a strong case for completing ZF with new axioms. In his seminal (49), Gödel proposed to extend the theory by adding principles inductively supported by their fruitfulness in consequences —as non-observable statements in the case of empirical science. More precisely, if a candidate for extending ZF (i) affords us proofs of new theorems and (ii) it also simplifies to a large extent the proofs of old ones —Gödel takes these as *verifiable* consequences—, then it can be regarded as sufficiently justified. Thus, fruitfulness becomes a criterion for the objective truth of set-theoretic statements *in the universe*.⁵ The

⁵See (94, Ch. 3) for a recent defence of fruitfulness as a criterion of set-theoretic objectivity.

hope was to press inductive justifications of this kind far enough to trigger a considerable reduction of the plural space of ZF's extensions, obtaining a complete theory—at least complete for natural statements—the correct theory of *the universe* of sets.

Gödel proposed statements asserting the existence of large cardinals as the most natural candidates for this task. As a result, he gave birth to the *program for large cardinal axioms*, one of the most ambitious attempts to complete ZF. While Π_1^0 -sentences are incomplete relatively to particular, recursively axiomatized systems—such as Robinson arithmetic (Q), PA or ZF—the notion of a *large cardinal axiom* is informal in nature, and it is not subject to a precise, recursive definition. Therefore, these present themselves as a perfect background to test candidates for *absolute undecidability*—sentences undecidable from any natural, justified extension of ZF. Gödel envisioned an informal notion of *absolute provability*—i.e., provability from true large cardinal axioms—and conjectured a generalized completeness theorem: “every proposition expressible in set theory is decidable from the present axioms plus some true assertion about the largeness of the universe of all sets” (1946, 151). Large cardinal axioms are a resounding success below CH. They yield a completeness result for the theory of $L(\mathbb{R})$ in the sense that this theory is preserved by set forcing constructions.⁶

However, large cardinal hypotheses present several limitations that seem hard to overcome. First, this type of model-theoretic construction does not alter the truth of Σ_1^2 sentences as CH. Hence, large cardinal axioms cannot settle the status of such complex open problems. Some philosophers and mathematicians (69; 70) have argued that this is not the last word on the debate. According to them, the weakness of large cardinal hypotheses does not entail the existence of an absolute undecidable sentence and, thus, a negative answer to Gödel's conjecture—if it is appropriately strengthened. The door is open to supplement the program with additional, stronger and natural principles. While this is true, and hope can be sustained, the contrary is also true: there is no reason to believe that all natural open problems in set theory will be solved one day or another. After several years of research, no final decision has been taken over any of the available candidates. Moreover, inductive justifications are not truth-conducive, and the criterion of *fruitfulness* as well as the notion of *naturalness* employed to characterize stronger set-theoretic principles are pragmatic. Therefore, it is far from clear that these attempts to complete the hierarchy succeed in recovering a high degree of objectivity. The multiversist denies that this is the case, and claims that set theory does not capture a single, well-identified concept of set—thus, (4.i).

4.4.2 The Argument from Categoricity

The second argument builds on the fact that first-order theories with infinite domains are not categorical. The concept of *categoricity* is model-theoretic—thus, a mathematical concept. It involves the—also mathematical—concepts of language, theory, and model. A theory T is categorical if and only if all

⁶This is the theory of definable sets of reals. A definable set of reals is a set of reals characterized by a Σ_n^2 -formula—this can be seen as a formula of second-order arithmetic. $L(\mathbb{R})$ is the hierarchy that results from starting with the reals and iterating the definable powerset operation—the operation of taking definable subsets—into the transfinite. Those sets of reals in $L(\mathbb{R})$ are the definable sets of reals.

models M_1 and M_2 of T are isomorphic.⁷ Results about categoricity are proved *in* mathematics. Then, it is reasonable to expect some explanation of what these results and the very notion of *categoricity* have to do with semantics, a philosophical discipline. The answer is that formal semantics, that is, the mathematical study of semantic concepts such as *reference* or *truth*, provides an analytically tractable environment for obtaining knowledge about these concepts — similar to computational models applied in meteorology, which supply us with affordable information about such a complex system as it is the weather. Also, the notion of independence from a recursive collection of axioms is a mathematical concept.

Nevertheless, it is known that sufficiently strong first-order theories are far from being categorical. The *Downward Löwenheim-Skolem theorem* entails that any first-order theory T that has an uncountable model M_1 also has a countable model M_2 —thus, M_1 and M_2 cannot be isomorphic.⁸ Furthermore, completeness does not fix categoricity. There are first-order complete theories with non-isomorphic models. For instance, arithmetic has non-standard models. As Skolem showed, the theory of the structure $\langle \mathbb{N}, < \rangle$ —which, by definition, is a complete arithmetical theory containing PA— has non-standard models $\langle \mathbb{N} \cup \{a\}, < \rangle$ where a is a non-standard number such that for all $n \in \mathbb{N}$, $n < a$. Subsequently, Robinson developed *nonstandard analysis* as the result of making precise the notion of *infinitesimals* —reals smaller than all standard reals (96).

The downward side of the Löwenheim-Skolem theorems can be applied to ZF —(128),(111)— supplying us with a countable, transitive model of sets. For, assume that ZF has a model M . By the axiom of Pairs, M must be infinite —countable or uncountable. Therefore, by the Downward Löwenheim-Skolem theorem, ZF has a countable model N . At the same time, by Cantor's theorem, ZF entails that $\wp(\mathbb{N})$ is uncountable. Moreover, it will be uncountable in N —this satisfies the Powerset axiom. How is this possible? This result, which has the air of a contradiction, is famously known as *Skolem's Paradox*. Yet, this curious result does not show any hidden incoherence in our understanding of sets. What happens is that the open formula “ x is uncountable” is satisfied by countable sets inside N —informally, we could say that N has no resources to capture the intended notion of *being uncountable*, that is, it cannot see which sets are uncountable and which are not. In N there is no one-to-one map from $\wp(\mathbb{N})^N$ —the set denoted by “ $\wp(\mathbb{N})$ ” in N — onto $\mathbb{N}$, even though such a map exists in V . Indeed, this is how Skolem originally understood the result, as a failure of formal resources to uniquely fix the concept of set.

Thus, axiomatizing set theory leads to a relativity of set-theoretic notions, and this relativity is inseparably bound up with every thoroughgoing axiomatization. *The relativity is due to the fact that to be an object in B means something different and far more restricted than merely to be in some way definable. That this relativity must be inseparably bound up with every thoroughgoing axiomatization is clear; for it rests upon the general theorems of mathematical logic mentioned above* [Italics in the original] (128, p. 296)

⁷A model M of classical first-order language is a n -tuple $\langle D, \langle R_i \rangle_{i < \alpha}, \langle f_j \rangle_{j < \beta} \rangle$, where D is a non-empty set of objects, $\langle R_i \rangle_{i < \alpha}$ is a sequence of relations of length α , and $\langle f_j \rangle_{j < \beta}$ is a sequence of operations of length β . Given two models M_1 and M_2 , an isomorphism from M_1 to M_2 , is a bijection from D_1 to D_2 preserving relations and operations.

⁸Besides, the companion *Upward Löwenheim-Skolem theorem* entails that if T has a countable model M_3 it also has an uncountable model M_4 .

A purely algebraic approach to reference—in accordance with (4.ii)—entails an extreme relativism of set-theoretic concepts. This relativism of set-theoretic concepts has been welcomed by (54), who has strengthened the paradox including among the axioms of his theory of the multiverse the *Countability Principle*—this says that every universe V is countable from the perspective of another universe V^* —as well as the *Well-foundedness Mirage*—it says that every universe V is well-founded from the perspective of another universe V^* .⁹

The universalist might object that categoricity is retrieved in second-order logic with standard semantics. Thus, second-order PA (PA^2) is categoric. Also, (154) proved that second-order ZF (ZF^2) is quasi-categoric, in the sense that for any two models M_1 and M_2 of ZF^2 , there is an embedding from M_1 to M_2 or vice-versa. Nevertheless, the standard semantics of second-order logic is highly controversial. According to the standard semantics, n -adic predicates are interpreted in $\mathcal{P}(D^n)$ —the power set of the n -ary Cartesian product of the domain. Thus, standard models are stated in terms of the set-theoretic notion of *powerset*. This is the reason why Quine qualified second-order logic as “set theory in sheep’s clothing.” (114, p. 66) This opinion is shared by many philosophers and mathematicians, who believe that it hides strong mathematical notions in the metatheory (36). Alternatively, if weaker semantic systems are used to interpret the second-order language, such as Henkin’s frames or plural logic, categoricity vanishes again as these systems are similar in expressive power to first-order logic. To sum up, if logical resources are kept at the safe level of first-order logic, most mathematical theories—from the elementary PA to the powerful ZF—are not categorical. Again, this has been used to argue that these resources fail to fix unique, well-defined mathematical concept. Multiversists such as (54) only accept first-order logic.

4.4.3 Morals from the Limits of Logic

According to the argument from logical independence, ZF is highly incomplete and, as a result, its axioms are not enough to single a unique, metaphysically privileged universe where set-theoretic terms and quantifiers are interpreted—similarly, a unique, privileged set-theoretic concepts about the objects in this universe. Furthermore, the argument from categoricity tells us that even if ZF is improved up to a complete theory ZF^+ , this will not be categoric. The logical resources afforded by the language where ZF^+ is formulated—if these resources are not controversial—are again too weak to fix reference and quantification into a unique, privileged universe. This is precisely what (4.i) states. Consequently, the multiversist claims that her argument against universalism is solid.

However, if multiversists want to be in a better position than universalism, they must show that multiversalism does not face an analogous objection to that posed by (4.v), i.e., they must show that the following is false:

⁹The relativity of the concepts *finite* and *infinite* had already been pointed out by Skolem: “Even the notions “finite”, “infinite”, “simply infinite sequence”, and so forth turn out to be merely relative within axiomatic set theory. A set M is finite, according to Dedekind’s definition, if it is not equivalent to any of its proper subsets. But the fact that the axioms hold does not rule out the possibility that we can define, first, parts of M that are not subsets or, second, correspondences that are not mappings, that is, “sets” of pairs. It is therefore quite possible that, within a domain B in which Zermelo’s axioms hold, there exist sets that are “finite” in the sense of Dedekind and for which there are one-to-one mappings onto some of their proper parts; but these “mappings” are not sets of the domain.” (128, p. 295)

(4.v*) It is not explainable how mathematical terms refer into different universes inhabiting the multiverse —or how concepts, relations and operations holding between the objects in the multiverse become expressed by the mathematical vocabulary.

Once more, we meet the Language Acquisition Challenge (LAC). The multiversist must explain how the mathematical language acquires its plural meaning —or how mathematicians become equipped with the appropriate resources to navigate the multiverse. As I said in 4.2, multiversists claim that once Responsiveness is rejected they can explain mathematical reference, mathematical knowledge and the reliability of mathematicians all at once —unlike fictionalism. Instead of causal or similar strong connections between speakers and mathematical facts, all multiversism requires are weaker connections between speakers and the logical properties of mathematical theories. Specifically, it only requires theories to be consistent — mathematical knowledge reduces to logical knowledge. According to the multiversist story, mathematical terms and quantifiers featuring in consistent theories are interpreted into some universe, and these mathematical theories are true relative to this universe.

In the following sections, I will argue that this is not the case. Neither Hamkins' nor Balaguer's form of multiversism can cope with the explanatory pressure posed by LAC. *Against appearances*, I will argue that such extreme forms of realist pluralism cannot solve *X* in the template argument of 4.2. As in the case of fictionalism, any solution provided by them fails to make (2.iv) true. They fail to explain how mathematical theories are true in different universes *because*, and they fail to explain how mathematical terms and quantifiers become interpreted into these universes. Furthermore, Balaguer's full-blooded platonism uses the notion of *consistency* to explain how the mathematical language acquires its meaning, and how mathematicians acquire their linguistic competence. Thus, I will argue that this fails as an account of mathematics for the very same reasons as fictionalism does. Despite this negative critique, note that the door remains open to less extreme forms of pluralism —for example, versions of multiversism that impose some limits to the extent of the multiverse.

4.5 THE SET-THEORETIC MULTIVERSE

Contemporary set theorists have devoted huge efforts to understand the phenomenon of independence. As a result of these efforts a variety of model-theoretic techniques —such as inner models, ultrapowers, set forcing, or class forcing— have been developed to show which extensions of ZF are consistent —relative to ZF's consistency— and, thus, which sentences are independent of it. Astonishingly, set theory has been populated with an unexpected variety of set-theoretic possibilities, and research about models of ZF has become a central part of its agenda.

The fundamental objects of study in set theory have become the models of set theory, and set theorists move with agility from one model to another. While group theorists study groups, ring theorists study rings and topologists study topological spaces, set theorists study the models of set theory. (54, p. 418)

Hamkins (54) takes this phenomenon as strong evidence that our concept of set is inherently relativistic and, thus, as evidence against universalism —he takes the *argument from independence* and the *argument*

from categoricity as supporting (4.i). Furthermore, Hamkins does not accept second-order logic as “true logic.” This has two immediate consequences. First, the concept of set framed in ZF cannot be completed by employing logical resources alone. Second, set theorists study models of first-order ZF and its extensions. As a result, he endorses an extreme form of pluralism.

Hamkins advocates multiversism as a satisfactory account of set-theoretic practices. The multiverse is navigated through the variety of model-theoretic procedures used to build ZF models. This navigation always takes place against a background concept of set —i.e., it requires a departure point. The reason is that most model-theoretic techniques are *relative construction methods*: they are performed against a background, previously given notion of set. Once the set theoretician is given this background, different model-theoretic techniques tell her how to build different models —the journey starts.

Hamkinsian Multiversism There are several universes of sets manifested through the application of model-theoretic techniques. Each universe elucidates a particular concept of set.

From the perspective of the set theorists, each of these universes is a model of first-order ZF —i.e. a set. Moreover, each *model* of ZF exerts a partition of the language in true and false sentences, and thus elucidates a particular, well-defined *concept* of set —indeed, Hamkins identifies concepts and models.

But, which kind of entity are models? From the perspective of the mathematician, models are just sets —i.e., mathematical objects. Note that this is a mathematical statement made against the background of an —at least partially specified— concept of set S . According to universism, this is the unique concept of set we have. As a result, given an object, S determines whether this is or is not a set in an absolute sense. Conversely, the very notion of set turns out to be relative on the face of multiversism. As I said, multiversism denies that there is a single concept of set. As a result, an object is a set —or not— depending on the background concept of set that we are taking into account. For example, take two models V_α and V_κ , such that $\alpha < \kappa$. Then, if V_α is a substructure of V_κ , what in the former is a proper class, in the second is a set. In the multiverse, everything is a set —or not— depending on where you look from. This reveals a surprising —but entirely coherent with its extreme relativism— feature of Hamkins’ multiversism: it cannot be characterised from within the multiverse, but only from outside it.

The assertion that there are diverse concepts of set is a metamathematical as opposed to a mathematical claim, and one does not expect the properties of the multiverse to be available when undertaking an internal construction within a universe. That is, we do not expect to see the whole multiverse from within any particular universe. (54, p. 417)

This kind of pluralism implies an extreme relativism about mathematical concepts. This situation entails certain limitations on what can be expressed with them. What the universes are or what the multiverse looks like cannot be stated from the point of view of the mathematician, who is always working inside the multiverse.

Hamkins defends multiversism not only as an account of set theorists’ practice, but also as a suitable account of the metaphysics of mathematics, and as a suitable account of mathematical reference and knowledge. The idea is that model-theoretic techniques have semantic consequences for the language: they explain how mathematical reference is fixed inside the multiverse. The explanation involves two

steps. First, set theorists start a model-theoretic construction from a particular universe —the universe specified by the background concept of set. Second, the construction *shifts reference* to another universe. Consequently, mathematical reference is explained as the result of shifting the reference of terms — originally fixed inside a background universe— to another universe by performing a model-theoretic construction. Terms and quantifiers are interpreted in the resulting universe.

If this explanation is in order, the multiversist succeeds in accounting for LAC. The practice of model theory explains how mathematical language acquires its meaning, and how mathematicians acquire their linguistic competence. Additionally, multiversism provides a suitable explanation of mathematical knowledge along the lines of 4.2 —thus, solving Benacerraf’s challenge and the challenge posed by mathematical reliability. It is enough to solve X with the property of *having a model* in the template-argument of 4.2. By construction, the set theorist knows how the resulting model looks like and knows if a given sentence is true in the model. Furthermore, (2.iv) turns out to be true:

(2.iv) According to multiversism, all sentences ϕ which have a model are true —at least ϕ is true relatively to some universe of sets.

If a set-theoretic sentence has a model, then it is true at least relatively to some model —i.e., relatively to some universe of sets. This guarantees a strong correlation between model-theoretic techniques and the multiverse. As a result, the Hamkinsian multiverse can explain mathematicians’ reliability in the following way. Although there is no counterfactual connection between mathematicians and mathematical facts —according to the denial of Responsiveness—, there is a counterfactual connection between mathematicians and mathematical methods. Agents have to learn these methods, and that requires a lot of effort —becoming a set theorist is not something that is easily achieved. Additionally, the correlation between model-theoretic techniques and the multiverse guarantees that these techniques are in tune with the truth of sentences — relative to some universe. As a result, the Hamkinsian multiversist endorses (2.v)

(2.v) Multiversism can explain why mathematicians accept mathematical sentences ϕ only if ϕ is true —at least relatively to some universe of sets.

with an explanatory connection between set theorists’ acceptance of sentences and their truth. Not only are all the lines of the argument true. They are also explanatory. For multiversism, being reliable is almost an analytical consequence of becoming a set theorist.

All this looks promising. However, I believe there are serious reasons to doubt that Hamkins’ explanation of reference works. Specifically, Neil Barton (6) has argued against this explanation on the grounds of the extreme relativism on which it is based. As I said, the procedures used in model theory consist of relative construction methods —they take place against a background concept of set. At the same time, Hamkins’ extreme relativism claims that there are no privileged concepts of set. Any model of ZF elucidates a perfectly valid concept. But then, how do set theorists fix the background concept of set from which they carry out their activity? Without further restrictions, all the constructions seem to hang in the air. In the next section, I will develop this argument against this explanation of mathematical reference. Like fictionalism, Hamkinsian multiversism cannot provide a suitable solution to LAC.

4.5.1 Starting Points and Infinite Descendant Chains

The argument exploits the aforementioned fact that model-theoretic techniques are relative construction methods. As Hamkins put it: “it is better to understand these descriptions as relative construction methods, since the resulting universe described depends on the initial universe in which the constructions are undertaken.” (54, p. 417) This extreme relativism entails that, if these methods are read as fixing the referential apparatus of mathematical language, they presuppose a background concept of set. Let V_i denote a given universe—or, equivalently, concept—of sets. Then, to interpret the terms and quantifiers in some universe V_i , a background concept of set V_j is needed. That is, we have

Relativism Given a concept of set V_i , then there is another concept of set V_j such that V_j partially explains how V_i is fixed.

Now, suppose the multiversist wants to explain how a given universe of set V_0 is fixed by the set theorist and how terms and quantifiers are interpreted in this universe. By Relativism we get

(5.i) There is another concept of set V_1 such that V_1 partially explains how V_0 is fixed,

V_1 is the background concept of set against which the set theorist works. Consequently, instead of one, we end having two concepts and thus the multiversist owes us an explanation of how V_1 is fixed and how terms and quantifiers are interpreted in V_1 . Thus, again by Relativism

(5.ii) There is another concept of set V_2 such that V_2 partially explains how V_1 is fixed.

Besides, those universes are different:

(5.iii) $V_1 \neq V_2$.

On the contrary, the explanation would simply be circular—it would claim that the set theorist picks V_1 partially because she picks V_1 . This is not allowed. Explanations are irreflexive, and the *explanandum* must be different from the *explanans*. It is not possible to explain reference is fixed in V_1 if this very fact is presupposed. Consequently,

(5.iv) There is another concept of set V_3 such that V_3 partially explains how V_2 is fixed.

As a result of this train of thought, the multiversist concludes from Relativism and (5.i) that

(5.v) A non-well-founded chain of concepts C is required to explain how reference and quantification is fixed in V_0 —each concept V_{n+1} explains how reference and quantification is fixed in V_n .

C constitutes a referential regress. The argument is sound and motivates a serious worry about Hamkinsian multiversism.¹⁰ To explain reference and quantification the Hamkinsian multiversist needs an infinite descending chain of concepts C . Each concept in the series explains the previous one. But, there

¹⁰Indeed, it can be proved by mathematical induction—assuming that there are at most countably many concepts or assuming AC. The following is a brief sketch of the proof. Let A be the concept we want to explain. Also, let

is no initial concept from which all these explanations take off. As a result, C gives rise to a referential regress.

The multiversist can retort that not all regress is *vicious* in nature. For instance, it is an open debate whether *grounding* is a well-founded operation (120). The idea of a non-atomic world —maybe ours— with increasing, non-ending degrees of fundamentality is shocking for some people. However, nobody discards it arguing that this is an inconsistent concept, or that it is impossible for a world to be atomless. Similarly, gunky objects —wholes such that all their parts have proper parts— are regarded as a mereological curiosity, but not as impossible entities. So, suppose that in such a non-atomic world reference to some object, O_1 is explained in virtue of reference to objects

$$O_2^1, O_2^2, O_2^3, \dots$$

of greater fundamentality. And, reference to each of these is explained by reference to other objects

$$O_3^1, O_3^2, O_3^3, \dots$$

of even greater fundamentality and so on (6, 201). This reasoning generates an infinite descendent chain C' of explanations in which reference to objects O_n^1, O_n^2, O_n^3 is explained through reference to objects $O_{n+1}^1, O_{n+1}^2, O_{n+1}^3$. Again, an infinite regress looms on the horizon. However, in this case, this is not obviously vicious. So —the multiversist goes on—, if both C and C' are analogous, and C' is unproblematic, C also must be unproblematic.

Nevertheless, here is where the answer breaks down: C and C' are significantly disanalogous. (5.v) entails that set theorists use infinitely many concepts to refer to sets in V_0 *on a specific occasion*. But, first, it is highly implausible that a given person actually possesses an infinite stock of concepts —as finite beings, human beings deal only with a relatively small number of conceptual resources. And this not only in a given moment of their lives, but in their whole lifetime. Second, the thesis that speakers are required to handle infinitely many concepts to explain how reference and quantification are fixed on a given occasion is even more implausible. Again, compare this to C'. C' involves infinite many objects exemplifying infinite many levels of fundamentality. But, in this case, reference does not need infinitely many concepts. Indeed, in addition to C' the explanation requires at most two concepts: one employed to pick O_1 and an additional one to pick C'. Moreover, if reference to C' is fixed by direct means, then only

$\varphi(x)$ abbreviate “there is a concept of set C such that C explains how the last concept of a sequence of concepts of length $x+1$ such that (i) each concept occupying the position i explains the concept at $i-1$ — $i \leq x+1$ — and (ii) has A as the initial element” is fixed. For the case of 0, C exists by (3.2.i). Now, assume $\varphi(n)$. So, we have a concept C corresponding to such a sequence of length $n+1$. Again, by (3.2.i), there is another concept C' explaining C. But, C is the last concept of a sequence of concepts of length $n+2$ such that: (i) each concept occupying the position i explains the concept at $i-1$ and (ii) A is the initial element. Therefore, $\varphi(n+1)$. Consequently, by induction on the naturals, for all n , $\varphi(n)$.

Now, let S_k be the set of sequences of length $k+1$ such that (i) each concept occupying the position i explains the concept at $i-1$ — $i \leq k+1$ — and (ii) has A as the initial element. Then, for each element $x \in S_k$ let C_k be the set of concepts of set such that partially explains the last concept in S_k . Assuming there are at most a countable number of concepts, C_k will be at most countable. So, we can pick an element of C_k . Call it C_{S_k} . If there are more than a countable number of concepts, AC does the job. Then, define by recursion: $f_0 = A$ and $f_{n+1} = C f_n$. This is a sequence of finite sequences of concepts with initial element A. Finally, $\bigcup f$ is an infinite descending chain of concepts explaining A.

one concept will be required. Consequently, C and C' are far from being similar. C is a *vicious regress* while C' is not.

4.5.2 Fixing the Multiverse

The argument shows that the explanation of reference afforded by Hamkinsian multiversism entails a vicious regress. Therefore, it fails to solve LAC: it cannot explain how mathematical language acquires its meaning. As a result, multiversism backs once more against the metasemantic and epistemological challenges. So, is there any other way out of the trouble? The unique option is for the Hamkinsian multiversist to modify the position in such a way that

1. The regress is no longer vicious
- or
2. Multiversism does not entail any regress.

Nevertheless, prospects for success are scarce.

4.5.2.1 Answer One

Consider the first option. (5.v) entails a *vicious* regress because C involves infinite many concepts. If compared to C' , the latter, despite being an infinite descendent chain too, only involves infinitely many objects and levels of fundamentality. Therefore, any regress is safe as long as it does not involve any speaker dealing with such an amount of cognitive resources all at once. Now, the elements of C are models V_i of first-order extensions of ZF. Hamkins identifies these with concepts. Then, one option is to break this identification: C would be composed of infinitely many models of ZF, but not of infinitely many concepts. All that is required to explain how reference and quantification are fixed on a given occasion is an infinite descendent chain of such models.

This answer is fatal for multiversism. It entails that, at some point in the chain, there is a model V_k such that the set theorist has no concept in mind corresponding to it when referring in V_0 . Nevertheless, reference into V_{k-1} is still explained as a result of performing some model-theoretic construction from V_k —at this point of the chain, this behaves as a background concept of set. That is, the multiversist still owes us an explanation of how reference is fixed in V_k . But, if conceptual resources do not make the job—as originally required—the unique alternative left is to advocate a theory of *direct reference*, as it was made in the case of reference to the objects across the layers of C' . However, direct reference involves a strong, counterfactual relation between speakers and facts involving the objects referred to: in this case, sets in V_k . This is impossible: according to the multiversist, sets are abstract objects and there is no counterfactual relation—for instance, a causal relation—between facts involving them and speakers. This answer breaks on the face of Benacerraf's considerations. Reference to abstract objects only can be secured by means of descriptive resources. Furthermore, I do not see any other way of turning C into a non-vicious regress. Therefore, the first answer is unsatisfactory.

4.5.2.2 Answer Two

What about the second option? Can the multiversist block the argument in such a way that no regress arises any more? The multiversist has three options.¹¹ First, she can reject Relativity. This claims that for any concept V_i there is another one V_{i+1} from which the set theorist secures reference to the sets into V_i —model-theoretic techniques are read as shifting reference from one universe to another. But, the multiversist can deny this. Hamkins proposes multiversism as a reaction to the inherent relativity of the set-theoretic language. There is no single, privileged concept of set. First-order theories do not express anything by themselves, but only in the context of set-theoretic practice.¹² For instance, the word “cardinal” expresses a particular property when displaying the features of the constructible universe of sets L , and other, quite different, when describing the generic universe $V[G]$ obtained from a non-constructive universe V and a poset G .

However, the Hamkinsian multiversist can argue in a slightly different way. She can propose a distinction between *absolute* and *relative* concepts of set. The former corresponds to Hamkins’ universes—absolute concepts which completely describe a particular realm of sets. On the contrary, relative concepts are not fully informative and apply to objects inhabiting several of these universes. Indeed, as Hamkins identifies absolute concepts with particular universes, relative concepts would be identified with all the universes inhabited by objects to which they apply. Thus, let M_Γ denote the collection of universes V which, for all sentences ϕ in Γ , V makes ϕ true. Once relative concepts are on the table, the multiversist can argue as follows against Relativity. Certainly, the set theorist requires a background concept of set S against which she performs any model-theoretic construction. But, S does not need to be an absolute concept. It can be a relative one of the form M_Γ . In addition, she needs to change the story about reference a little bit to match the new scenario. The multiversist could say something according to the following lines: model-theoretic techniques would not only shift reference from one universe V_i to another universe V_j . Additionally, they target particular universes—and thus, absolute concepts—from the partial information made available by the absolute ones. Thus, Relativity must be modified as follows

*Relativity** Given a concept of set V_i , there is another concept of set S ,—absolute— $S = V_j$ or—relative— $S = M_\Gamma$, such that S partially explains how V_i is fixed.

As a result, the conclusion (5.v) is blocked: no regress arises. For, now the multiversist does not need to explain how reference and quantification are fixed in a relative concept M_Γ .

All that is needed to explain is how the set theorist manages to express such a concept. But, the explanation of how M_Γ is fixed should be very different from the explanation of how the absolute V_j is. For example, relative concepts could be fixed through Γ and satisfaction as follows—where “ V is a model of ϕ ” is explained through satisfaction as usual:

(5.vi) Given a collection of universes U , M_Γ = the collection of universes V in U such that for all sentence $\phi \in \Gamma$, V is a model of ϕ .

¹¹Barton considers the last two options.

¹²Since Hamkins accepts non-transitive and non-extensional models as acceptable models of set theory, his extreme relativism entails that no set-theoretic concept is absolute.

Thus, (a) no model-theoretic techniques, nor (b) a direct theory of reference are required to explain how the set theorist picks M_Γ . So, the chain goes down until this runs into such a relative M_Γ . Then, Γ and a suitable relation of satisfaction do explain how this concept became understood ending with the chain. For all cases of reference into an absolute V_i , there must be a *finite* sequence of concepts ending in a relative one.

This answer looks promising. Nevertheless, (a) is false. For (5.vi) presupposes a given collection of universes U . That is, presupposes access to some extent of the multiverse. Now, how does the set theorist get access to this extent? Presumably, if pressed enough, the multiversist would say that it is gained by performing the model-theoretic constructions which “generate” the universes composing it. Therefore, M_Γ will only be fixed under a previous selection effected on the multiverse. Let “Extent $_S$ ” denote the extent of the multiverse accessible —by means of the corresponding model-theoretic techniques— from a concept S . The multiversist also accepts

(5.vii) For all M_Γ there is a concept S such that $M_\Gamma =$ the collection of universes V in Extent $_S$ such that for all sentence $\phi \in \Gamma$, V is a model of ϕ .

Once this is assumed, (5.vi) works perfectly. Any relation of satisfaction presupposes the models in its range. The problem is that now Relativity* and (5.vii) give rise to another infinite regress. So, the distinction between absolute and relative concepts —an additional layer of theorizing not contained in Hamkinsian multiversism— does not solve any problem.

A second option for the multiversist is to break the regress by other means than introducing relative concepts. As we have seen in the previous paragraph, the apparent advantage of relative concepts was that they transformed the infinite sequence of concepts C into a finite sequence —one ending with a relative concept. Similarly, the multiversist could argue that, because of other reasons, C —as well as any other chain of concepts involved in the explanation of reference— is finite. If she finds a stopping point in each of these sequences, then her explanation of how the set theorist does refer will be safe —at least it will avoid the challenge. The problem is: what reasons does she have to postulate such stopping points? Barton suggests —only to test it— the following answer: “instead of maintaining that the extent of the Multiverse is dependent of where we start, she could instead say that its extent is indeterminate *in virtue* of being so dependent. The Multiverse is then fixed from this particular stopping point, which may be different on any particular occasion.” [Italics in the original] (6, p. 203)

Nevertheless, this is no answer at all. On the contrary, this is just a circular and subtle form of skipping the responsibility of explaining reference and mathematical knowledge. For, if we look more carefully, all that this —allegedly— solution does is change the requirement of explaining reference for the requirement of explaining how a stopping point is fixed on a given occasion. But, no such explanation is developed and reference still hangs in the air. But, are there any reasons to suppose that chains of concepts involved in the explanation of reference do always meet a last stopping point? The answer is a resounding no. Furthermore, to postulate the existence of these points is to beg the question, or endorsing set theorists with a mysterious access to the multiverse. But, if this is so, there is no reason to deny similar powers to the universist: indeed, the stopping points could always be one and the same, *the* universe of sets. In

conclusion, the multiversist cannot provide evidence that reference in the multiverse is achieved by only employing finite chains of concepts.

Finally, the multiversist is left only with one alternative: to accept a privileged concept of set, imposing a limit on the extent of the multiverse. Reference to the objects inhabiting this universe would be explained by other means than the performance of some model-theoretic construction. Hence, this constitutes a *modest vindication* of universism: *the* universe of sets would be this familiar and initial universe from where all constructions and incursions into the multiverse start.¹³ The multiverse needs a starting point. Nevertheless, this answer by no means entails that all versions of multiversism are false—only the radical Hamkinsian version. Once a background concept of set *V* is fixed and reference to such sets is secured, then model-theoretic methods can be interpreted as shifting reference from *V* into other, different universes gaining new, previously unknown concepts of set. Set theorists become explorers of the logical space of sets and their journey starts at home. In particular, *V* is not metaphysically different from the other inhabitants populating such a vast space. All it does is to provide set theorists with a point of departure in their journey.

In any philosophy of mathematics that tries to make sense of a multiverse picture of set theory as concerned with reference, there is likely to be an element of relativism of certain concepts. However, there is a limit to how far this relativism can go. In particular we need to be precise about which universes (off the bat) are to count as privileged, and in doing so will have to specify a list of concepts that we are simply taking to have absolute significance to restrict our models. Moreover, this stock of concepts must be sufficiently rich to allow the relevant model-theoretic techniques to be absolutely specified, and multiverse given precise limit. The acceptance of ‘any particular’ first-order model as a legitimate universe results in a relativism so strong that it cripples our ability to refer at all. (6, p. 204)

If such a departure point is rejected, I believe that multiversism fails at explaining mathematical reference and knowledge. Then, multiversism fails to solve LAC and Benacerraf’s challenge. Consequently, it fails as an account of mathematics—in particular, of set theory.

Indeed, the view collapses into an algebraic, anti-realist position similar to algebraic fictionalism—see Chapter 3, 3.5. Without a minimal starting point, Hamkinsian multiversism turns out to be another algebraic approach to mathematics according to which set theorists do not deal with a variety of concepts of set—least of all with abstract objects, such as sets and universes. Instead, they would lay down some—consistent—axioms and study what follows and what does not from them. As a result, their activity should be understood not as asserting something about a domain of existing objects, but as studying what will happen if some axioms were true. Unfortunately, as I argued in Chapter 3, 3.5 this position is also full of problems.

¹³The question of whether the multiverse has a core has been widely debated. See (132) for a positive development of this position.

4.6 FULL-BLOODED PLATONISM

The other version of multiversism I am going to discuss is *full-blooded platonism* (FBP). This was elaborated in (2) explicitly as a solution to Benacerraf's Challenge. He proposes to solve the challenge by abandoning Responsiveness and denying that counterfactual relations between mathematical facts and mathematicians are needed to explain mathematical knowledge. As fictionalists and advocates of other versions of multiversism, Balaguer explains mathematical knowledge as a case of logical knowledge. Therefore, the rejection of Responsiveness leads, again, to a plural account of mathematics: FBP is a form of pluralistic platonism. There is no universe of metaphysically privileged mathematical facts. In the absence of such a universe, there are either many or none, and Balaguer chooses the former option.

FBP is a simple thesis: all possible mathematical universes exist.^{14,15}

Full-blooded Platonism All logically possible universes exist.

FBP is the result of supplementing multiversism with a maximal claim. Not only are there a variety of different —many of them strange— universes. Moreover, none of them is missing. Its mere logical possibility suffices for its existence. Hence, FBP envisions a full, plenitudinous multiverse of mathematical objects.

According to Balaguer, this is what allows for a weak, non-counterfactual explanation of mathematical knowledge.

If FBP is correct, then all consistent purely mathematical theories truly describe some collection of abstract mathematical objects. Thus, to acquire knowledge of mathematical objects, all we need to do is acquire knowledge that some purely mathematical theory is consistent. (It doesn't matter how we come up with the theory; some creative mathematician

¹⁴Unlike Hamkinsian multiversism, full-blooded platonism is a general proposal for all mathematical theories. Balaguer does not focus on set theory. However, for the sake of continuity, I will present the proposal in set-theoretic terms.

¹⁵See (2, pp. 5-7) for a characterization of FBP. Balaguer spends some time explaining how FBP can be stated formally. "I do think it might help clarify FBP to say a few words about how one might go about trying to state it in a formal language. Before I do this, however, I want to emphasize that my sole aim here is to help the reader get clear about what FBP says; nothing important depends on finding an adequate formalization of FBP." (2, p. 6) Unlike Hamkins, he employs second-order logic to make himself available resources to talk about universe-like entities —despite he does not talk about universes. Then, FBP is formulated as follows:

$$\exists x Mx \wedge \forall V (\Diamond \exists z (Mz \wedge Vz) \rightarrow \exists z (Mz \wedge Vz)),$$

where "*M*" is a predicate expressing the property of *being a mathematical object* and "*V*" is a second-order variable ranging over universe-like entities. Of course, one wonders whether this does not condemn Balaguer's proposal from the outset. If second-order resources are available to talk about universes, what prevents us from accepting the categoricity of mathematical theories? The answer is that, for the full-blooded platonist, second-order variables do not range over subclasses of a well-determined universe. On the contrary, they encompass a multiplicity of such universes.

The metaphysical status of the universes, as well as the linguistic resources used to talk about them, is especially pressing. The answer to these questions will determine the best formalisation of the proposal. Here I will keep the exposition on an informal level, assuming that the multiversist discourse is in good shape —in particular, that talking about the existence of universes is safe.

might simply “dream it up”.) But knowledge of the consistency of a mathematical theory—or any other kind of theory, for that matter—does not require any sort of contact with, or access to, the objects that the theory is about. Thus, the Benacerrafian objection has been answered: we can acquire knowledge of abstract mathematical objects without the aid of any sort of contact with such objects. (2, pp. 48-49)

Like fictionalists, Balaguer reduces mathematical knowledge to logical knowledge. Knowledge about a mathematical sentence φ reduces to two things: (i) knowing that φ follows from a given mathematical theory Γ , and (ii) knowing that Γ is consistent. Thus, he solves X in the argument-template of 4.2 with the property of being consistent. Now, as in the case of Hamkinsian multiversism, (2.iv) turns out to be true:

(2.iv) According to FBP, all consistent mathematical sentences φ are true—at least φ is true relatively to some universe of objects.

Indeed, this is tantamount to the very definition of FBP. Its plural realism guarantees an intimate correlation between consistency and truth. As a result, Balaguer explains mathematicians’ reliability exactly in the same way as the Hamkinsian multiversist does—replacing the property of having a model, with the property of being consistent.¹⁶ Again, the rejection of Responsiveness prevents any counterfactual connection between mathematicians and mathematical facts. However, there is a counterfactual connection between agents and the process that leads them to acquire their linguistic competence, as well as with the procedures that allow them to acquire knowledge about consistency. Agents have to learn to speak and to determine which theories are consistent and which are not. This is something that does not come without effort. Additionally, the correlation between consistency and the multiverse guarantees that knowledge of the former is in tune with the truth of sentences—relative to some universe. As a result, the FBP-ist endorses (2.v)

(2.v) FBP can explain why mathematicians accept mathematical sentences φ only if φ is true—at least relatively to some universe of objects,

with an explanatory connection between mathematicians’ acceptance of sentences and their truth. Unlike in the case of fictionalism—see Chapter 3, 3.5.3—, not only are all the lines of the argument true. The argument is also explanatory. Knowledge about consistency—i.e. logical knowledge—ensures mathematicians’ reliability.¹⁷

If this is correct, FBP solves the major epistemological challenges of platonism: Benacerraf’s epistemological challenge and the challenge posed by mathematicians’ reliability. As a result, it would be a refined and philosophically sound form of platonism. However, as in the case of fictionalism and Hamkinsian multiversism, I doubt that its explanation of mathematical knowledge works. Indeed, because FBP reduces mathematical knowledge to logical knowledge, I will argue it faces similar problems

¹⁶These properties are not equivalent in higher-order logic. However, because of the argument from categoricity, multiversists tend to accept first-order languages. But, completeness and correctness entail that both properties are equivalent at this level. In this case, if the notion of consistency used by the FBP-ist coincides with the notion of consistency at stake in set theory, the explanations of the Hamkinsian multiversist and the FBP-ist are the same.

¹⁷See [p. 51] (2) for some comments on mathematical reliability from the point of view of FBP.

to fictionalism. First, like algebraic fictionalism —see Chapter 3, 3.5.3, FBP lacks an appropriate notion of consistency to perform the desired explanatory function. Again, the two candidates are a defined notion of consistency or a primitive one. The former presupposes mathematical knowledge, leading unavoidably to circularity. In turn, the second presupposes an understanding of mathematical language. Let us compare its situation with fictionalism. In Chapter 3, 3.5.3, I argued that a primitive notion of consistency must be a linguistic notion, thus a semantic notion. This puts algebraic fictionalism in serious trouble, since the latter either attributes no content to mathematical language or collapses into full fictionalism —see Chapter 3, 3.5.3. Both options are dead ends. Now, the situation is similar for FBP. This only has to deal with the second horn of this dilemma —due to its underlying realism, it attributes semantic content to language. However, FBP fails to explain how mathematical language acquires its meaning, and how speakers understand it. That is, we meet once more the Language Acquisition Challenge (LAC). As it happens to full fictionalism and Hamkinsian multiversism, FBP can not solve this challenge. In this case, it attempts to explain reference through consistency. But, this is circular.

Second, the realist metaphysics supported by FBP puts it in additional trouble. Assuming a minimum of mathematical complexity, the criterion of mathematical knowledge —consistency— becomes stronger than the theories themselves. There is nothing suspicious in acquiring knowledge about the facts depicted by some theory from other, stronger theory. Indeed, this is quite usual. For example, the premises of a deductive argument are —often— stronger than the conclusion. But they allow us to know the latter. The problem is to establish a criterion of knowledge that is stronger than what is known as a general rule. Both criticisms are exposed below.

4.6.1 The Consistency Challenge Again

FBP turns out to be in a similar situation to algebraic fictionalism. While it attempts to reduce mathematical knowledge to logical knowledge —knowledge about consistency and entailment— it lacks an appropriate notion of consistency. Remember, there are two candidates for the job. First, the mathematical concept of consistency that is used in model theory and proof theory. Second, a primitive notion of consistency. However, the FPB-ist can use none of them.

Consistency receives a smooth mathematical treatment in model theory and proof theory. Take the former. Using the notion of model, the consistency of a collection of sentences Γ reduces to the existence of a model M for the theory.

Model-theoretic Consistency A set of formulas Γ is *model-theoretically consistent* if and only if there is a model M such that for all $\varphi \in \Gamma$, $M \models \varphi$,

where “ $M \models \varphi$ ” is defined as usual in terms of Tarski’s satisfaction clauses. Alternatively, consistency can be analyzed using a formal notion of proof —expressed by “ $\vdash$ ”. The consistency of a collection of sentences Γ reduces to the fact that it does not prove a contradiction.

Proof-theoretical Consistency A set of formulas Γ is *proof-theoretically consistent* if and only if there is no sentence φ such that $\Gamma \vdash \varphi$ and $\Gamma \vdash \neg\varphi$.

Both the model-theoretic and the proof-theoretic notions of consistency reduce consistency to a mathematical concept. As a result, both are unacceptable for FBP's purposes. FBP attempts to explain mathematical knowledge in terms of consistency. But, if consistency is already a mathematical concept, it cannot play this explanatory role. If this were so, the explanation of mathematical knowledge would already presuppose that agents possess mathematical knowledge —i.e., knowledge about consistency, a particular mathematical concept. However, this is a bad explanation. Explanations are not reflexive (120)—facts are not self-explanatory. For instance, if someone asks why the moon takes 27 days to orbit the Earth, it would be absurd to answer “*because* it takes 27 days to orbit the Earth.” Consequently, FBP needs a different, non-mathematical concept of consistency.

At this point, (2, 69-75) appeals to a primitive and intuitive notion of consistency —this is the same as the one I exposed in Chapter 3, 3.5.3. I believe this notion is in perfectly good condition. It is highly plausible that linguistic competence comes with an intuitive understanding of such matters as consistency or possibility. But, even if this was not so, assume for the sake of the argument that the FBP-ist has some understanding of an intuitive notion of consistency. Balaguer claims that such a notion of consistency does not involve any “contact” with those facts depicted by mathematical language.

In general, knowledge of the consistency of a set of sentences —whether the sentences are purely mathematical, purely physical, mixed, or whatever —does not require any sort of epistemic access to, or contact with, the objects that the sentences are about. (2, p. 73)

This is true when “contact” is understood as involving some strong, metaphysical relation. For instance, no counterfactual relation with El Capitan —a rock formation 914 metres high located in Yosemite National Park, California— is required to know that those sentences claiming that it is a monolith and that it is 500 metres high are consistent with each other. Similarly, no counterfactual relation is required to know that the sentences “El Capitan is 500 metres high” and “El Capitan is not 500 metres high” are inconsistent with each other —these are the kind of examples Balaguer uses in (2, p. 73). Indeed, one may not even know what El Capitan is. All that is required is some understanding of the expressions composing the sentences —this is the crux of the matter.

What is the problem? As I argued in Chapter 3, 3.5.3, primitive notions of consistency and possibility are eminently *semantic*. Logical knowledge is knowledge concerning a relation of entailment. But, this is a linguistic relation. That is, a relation between sentences, propositions, or other similar entities. And these categories, as well as the category of *language* are semantic. Grammatical categories, such as the category of names, of predicates, or sentences, are already semantic categories. For example, a name is a name because it performs a series of semantic and communicative functions. Among them, it attempts to refer to an object. Thus, the name “Aristotle” is a name because speakers —among other things— attempt to refer to an object using it, the Greek philosopher Aristotle. Similarly, the name “El Capitan” is a name because speakers attempt to refer to El Capitan using it. The semantic function of linguistic expressions is indissolubly tied to their identity.

If the linguistic expressions e_1 and e_2 perform different semantic functions, then $e_1 \neq e_2$.

The same happens with other expressions, as predicates. Consider the predicate “to be 500 metres high”. As in the case of names, this predicate has several complex semantic and communicative functions. For

example, to convey information about those objects which are 500 metres high. If it did not perform this function, then it would not be a predicate—at least not this particular predicate. It would be another object, such as a stain or a sound. Although not all grammatical distinctions are logically relevant, they introduce semantic or communicative distinctions. The point is: it makes no sense to vary the semantic content or communicative function of a name, a predicate, a sentence, an adverb, etc. without varying its identity.

Therefore, speakers *must* understand the language of a given theory or a collection of sentences to claim that they are consistent. The theory is a linguistic entity coached in a meaningful language. Therefore, to use an intuitive notion of consistency, the FBP-ist must solve LAC: she must explain how the mathematical language acquires its meaning, and how mathematicians are competent in it. Indeed, Balaguer accepts the challenge (2, pp. 49, 51-52): “After all, platonists need to explain not just how we could acquire knowledge of mathematical objects, but also how we could do things like have beliefs about mathematical objects and refer to mathematical objects.” (2, p. 49) Not only that. In addition, he claims that FBP solves the challenge: according to him, the FBP-ist can explain how mathematicians formulate meaningful mathematical theories—this is the first premise in his argument for FBP (2, p. 51).¹⁸ However, what is his explanation? The truth is that Balaguer does not elaborate any. He takes the explanatory pressure off himself by saying that mathematical language is *about* mathematical objects in a *thin* sense:

The argument for (i) has already been given: this premise is trivial because it is not making any strong claim to the effect that our purely mathematical theories have unique domains of mathematical objects that they are “about” in some metaphysically thick sense of the term.
(2, p. 52)

However, Balaguer does not explain what “to be thinly about” consists of. He simply points to some alleged examples of beliefs which are thinly about a domain of objects—for instance, people’s beliefs about mathematical objects.¹⁹

This is not enough to meet the explanatory pressure put by LAC. The challenge asks for a full explanation of mathematicians’ linguistic competence, as well as for a metasemantic explanation of how mathematical expressions acquire their meaning. To claim that mathematical language is about—in a *thin* sense—mathematical objects, or that people believe—in a *thin* sense—in these objects is to beg the question. It is far from clear how the notion of “to be thinly about” alone helps to understand any of these matters. That is, how it helps to develop a suitable account of mathematical language. At best, merely pointing to this notion is too weak to perform any explanation. Therefore, the notion of “to be thinly

¹⁸“FBP-ists can account for the fact that human beings can—without coming into contact with the mathematical realm—formulate purely mathematical theories.” (2, p. 51)

¹⁹“Now, no one doubts that human beings could have beliefs that are thinly about mathematical objects. Indeed, the existence of a single mathematical platonist establishes this, for in such a person, we have someone who wholeheartedly believes that there are abstract mathematical objects (e.g., the number 3) and that these objects have certain properties (e.g., primeness). Now, of course, it may be that there are no such things as mathematical objects and, hence, that there is nothing “out there in the world” corresponding to our platonist beliefs, but this does not change the fact human beings are capable of arriving at beliefs that are thinly about mathematical objects.” (2, p. 39)

about” must be properly developed and explained in order to elaborate a complete theory that meets the challenge posed by LAC.

Crucially, what FBP needs is an explanation that: (1) assigns to each sentence its truth condition and (2) according to the denial of Responsiveness, does so without involving counterfactual connections between speakers and mathematical facts. In this way, it can avoid at the same time the challenge imposed by LAC and Benacerraf’s challenge. Moreover, this would entail an explanation of what “to be thinly about” consists in. Now, what kind of relation can FBP use that fulfils this double function? The only relation available is, precisely, the one that follows from his definition. FBP is the thesis that every possible universe of mathematical objects exists. Then, given the relation between the primitive concepts of possibility and consistency, FBP entails that every consistent theory can be interpreted in a universe —i.e. every consistent theory is true in some universe. In consequence, consistency should be the vehicle that connects sentences with their truth condition —in general, consistency should connect each mathematical expression with its semantic value. Thus, the FBP-ist could elaborate the following three-step explanation —recall the similarity with fictionalism. First, classical mathematicians are interested in consistent theories. Thus, they understand which theories are consistent and which are not. Second, FBP equates consistency and truth in the multiverse. If mathematicians have logical knowledge about consistency, this connection guarantees that the language is meaningful and that mathematicians understand the language. As a result, a suitable metasemantic explanation can be developed to meet the challenge posed by LAC. Finally, this picture would be a perfect explanation of what “to be thinly about” means too: *to be thinly about* some objects in the multiverse is nothing more than to be a consistent sentence, belief or theory.

How successful is this explanation? Certainly, despite appearances, it is not a good explanation. It *reduces linguistic competence to logical knowledge*. Now, remember that the notion of consistency at use is a primitive one. And, a primitive notion of consistency must be semantic. That is, to know which theories are consistent and which ones are not, the mathematician must understand the mathematical vocabulary and the sentences composing those theories. The former cannot occur without the latter:

Logical knowledge takes place only if agents understand the language.

Nevertheless, here we are going the other way around. We are trying to explain language understanding through logical knowledge. Consequently, the explanation is circular. Furthermore, there is no other connection between sentences and their truth conditions in the multiverse to which the FBP-ist can appeal. As a result, she fails to meet the explanatory pressure put by LAC. The FBP-ist cannot explain either how mathematical vocabulary acquires its meaning, or how mathematicians speak about mathematical objects. The unique option left for FBP is to resort to formal languages. There are perfectly good formal notions of language, formula, and theory used in formal logic. All of these are purely syntactic categories —no one presupposes meaning or semantic content. Moreover, they allow to speak about consistency. However, as I argued in Chapter 3, 3.5.3, formal languages are mathematical devices employed in model theory, proof theory, and other branches of formal logic. As a result, when the logician claims that a collection of formulas is consistent she employs the *model-theoretic* or the *proof-theoretic* notion of consistency. But, as we have seen, these are not available to the FBP-ist.

In conclusion, FBP is not allowed to use primitive notions of consistency or possibility to explain mathematical knowledge. Both demand linguistic competence on the part of the mathematician: that is, they have logical knowledge concerning the language of mathematics only if they understand this language. However, FBP fails to provide a suitable explanation of the latter. Thus, as fictionalism and Hamkinsian multiversism, it cannot solve the challenge posed by LAC. They do not explain how vocabulary acquires its meaning or how speakers become linguistically competent. Moreover, FBP cannot use mathematical notions of consistency either. These concepts lead FBP into an irremediable circularity. But, the mathematical and primitive notions of consistency—or possibility—are the unique notions available for FBP. Consequently, FBP fails to explain mathematical knowledge. The latter does not reduce to logical knowledge. Something more is needed.

4.6.2 The False Weakness of Logic

Finally, I will expose an additional challenge to the reduction of mathematical knowledge to logical knowledge. Assuming a minimum of mathematical complexity, knowledge of the consistency of a theory exceeds the knowledge of the theory. Thus, knowledge of consistency is too strong an epistemological criterion.

I will oppose the challenge both to intuitive and mathematical concepts of consistency. For this, note that there is a close connection between intuitive notions of consistency and the model-theoretic or proof-theoretic definitions. Roughly speaking, from a platonist point of view the latter offers valuable information to determine the extension of the former. In the case of first-order logic, the three concepts become coextensive for the platonist.

(a) if a theory T is semantically consistent, then it is intuitively consistent; and (b) if T is syntactically inconsistent, then it is intuitively inconsistent. Moreover, if we combine these two points with the completeness theorem [...] we arrive at the result that (among first-order theories) the intuitive notion of consistency is coextensive with both formal notions of consistency. (2, p. 70)

Because completeness breaks down when we move up to higher-order logic, this result does not hold in stronger systems. However, I believe that it is plausible to accept a weaker relation between primitive and syntactic consistency even at these levels. Specifically:²⁰

(*) A theory Γ is consistent according to our intuitive standards if and only if Γ is syntactically consistent.

²⁰Fictionalism avoids the challenge exposed below because it forbids this equivalence. Technically speaking, the arithmetical sentence expressing consistency will be false: this sentence quantifies over numbers, but numbers do not exist. According to fictionalism, mathematical theories are fictions. Therefore, mathematical knowledge of arithmetical sentences taken at face value simply vanishes. Thus, it is trivial that logical knowledge is stronger than mathematical knowledge for all mathematical theories. Nevertheless, I believe the challenge could be reintroduced in a more subtle way.

Otherwise, we would take as intuitively consistent some theory Γ that syntactically implies a contradiction. This would hold against the different logical calculus, including the first-order calculus, a conclusion that is hard to bite: mathematical knowledge about logic would turn out to be false. Otherwise, note that the analogous equivalence concerning semantic consistency is far more dubious. There are no extra-mathematical reasons to think that in higher-order logic the consistency of a theory suffices for the theory to have a model—which is a set. In short, given any n -order system, it is reasonable for the platonist to accept that the intuitive consistency of a set of sentences Γ is equivalent to its syntactic consistency. This is enough to develop the challenge.

Let's abbreviate “ φ is consistent” as $\text{Con}(\varphi)$. As Gödel showed, $\text{Con}(\varphi)$ is a number-theoretic statement—modulo some appropriate coding of the vocabulary of the language. Now, suppose that a mathematician knows (*)—I believe this is highly plausible. Suppose also that knowledge is preserved by *Modus Ponens*—this is almost universally accepted. Then, if this mathematician knows that Γ is intuitively consistent, by (*) she also knows that $\text{Con}(\Gamma)$. Knowledge about a primitive concept of consistency is tantamount to arithmetical knowledge.

I am now in a position to pose the following difficulty. According to FBP, assume that mathematical knowledge reduces to knowledge of consistency and entailment. Thus, knowledge about a mathematical sentence φ reduces to knowing its consistency. That is, knowing that φ follows from knowing $\text{Con}(\varphi)$. Similarly, for a set of sentences Γ , knowing the theory Γ is achieved through knowledge of $\text{Con}(\Gamma)$:

$$K(\text{Con}(\Gamma)) \rightarrow K(\Gamma).$$

The problem is that if $\text{PA} \subseteq \Gamma$, then $\text{Con}(\Gamma)$ is a stronger statement than any of those sentences in Γ . Gödel second incompleteness theorem says that if PA is consistent, then $\text{PA} \not\vdash \text{Con}(\text{PA})$ and $\text{PA} \not\vdash \neg\text{Con}(\text{PA})$.

$$\text{Con}(\text{PA}) \rightarrow \text{PA} \not\vdash \text{Con}(\text{PA}) \wedge \text{PA} \not\vdash \neg\text{Con}(\text{PA}).$$

Assume, then, that a mathematician has arithmetical knowledge—i.e., she knows PA .²¹ According to FBP, she obtains this knowledge because she knows that PA is consistent. Hence, by factivity of knowledge, PA is indeed consistent—we have $\text{Con}(\text{PA})$. And, by Gödel's second incompleteness theorem, $\text{PA} \not\vdash \text{Con}(\text{PA})$. Consequently, knowledge of PA does not transmit through logical consequence to knowledge of $\text{Con}(\text{PA})$ —the mathematical sentence stating that PA is consistent. Moreover, given the connection between syntactic consistency and intuitive consistency, knowledge of PA does not transmit through logical consequence to knowledge of the intuitive consistency of PA . As a result, FBP entails that for any mathematical theory $T \supseteq \text{PA}$, the *unique* criterion of knowledge for T is stronger than T . For platonism, logical knowledge is stronger than mathematical knowledge in some cases.

This situation is challenging for FBP for two reasons. If logical knowledge is stronger than mathematical knowledge, and mathematical knowledge is regarded as problematic, is the problem not transmitted to logical knowledge? As far as arithmetical knowledge is concerned, logical knowledge—even knowledge about a primitive notion of consistency—makes more demands on the world than mathematical knowledge. Thus, if the latter is problematic, the former will be even more so. Second, it is controversial that knowledge about a sentence or a theory requires as a general rule knowledge of something

²¹Knowledge about Robinson Arithmetic (Q) is enough. This is PA without induction.

independent from them. Indeed this is often the case. The clearest example is —precisely— inferential knowledge. If we know that the glass is not on the table and that the glass is either on the table or under the table, then we can know that the glass is under the table. This is so even if the disjunction “the glass is either on the table or under the table” does not follow from the conclusion. Nevertheless, what is controversial is to establish that the unique criterion of knowledge of a sentence or of a theory must be always epistemologically stronger than them.

The FBP-ist can object that knowing that a sentence or a theory is consistent is not stronger than knowing them. Thus, she can endorse the following conditional:

$$K(\Gamma) \rightarrow K(\text{Con}(\Gamma)).$$

As a result, knowledge of theories becomes equivalent to knowledge of their —intuitive and mathematical— consistency. Thus, the previous criticism simply vanishes. Knowing a theory T we already know that it is consistent. But, to what extent is this plausible? At least, I believe this is highly controversial. It is not obvious that speakers’ knowledge of a theory already conveys knowledge of its consistency. On the contrary, it seems that knowledge of consistency requires a cognitive achievement beyond the efforts made to know the facts depicted by T . For instance, suppose that current physics affords us knowledge about how things are at an empirical level. Thanks to it, we know the laws governing elastic collision, how entropy characterizes the evolution of thermodynamic systems, or how elementary particles behave. However, this situation does not seem to involve knowledge about the state of the theory itself. That is, whether it is a good description of the world or not, whether it is consistent, etc. Something more is required to know this thing. Knowing that a theory is consistent —or what it depicts is possible— requires an extra cognitive achievement —even if the notions of consistency and possibility used are primitive.

Indeed, if knowledge of a theory already entails knowledge about its consistency, the FBP-ist must accept the following situation. Let’s define recursively the following sequence of theories as follows: $PA_0 = PA$ and $PA_{n+1} = PA_n \cup \text{Con}(PA_n)$. Then, knowledge of PA would be equivalent to knowledge of PA_n , for arbitrary n . Again, this is at least highly controversial.

To sum up, the reduction of mathematical knowledge to logical knowledge faces two challenges. First, FBP lacks an appropriate notion of consistency to perform this reduction. On one side, mathematical concepts of consistency are useless for this purpose. In turn, FBP can not make use of a primitive notion of consistency —or possibility— until the challenge posed by LAC is solved. However, it is not enough to meet this explanatory pressure —i.e., FBP cannot explain how the mathematical vocabulary acquires its meaning, and how mathematicians understand this vocabulary. Second, given a minimum of mathematical complexity —arithmetic—, knowledge of consistency turns out to be stronger than mathematical knowledge. As a result, it is highly controversial that it can be established as a general criterion of mathematical knowledge. Consequently, the reduction of mathematical knowledge to logical knowledge is highly controversial too. Nevertheless, as in the case of Hamkinsian multiversism, this does not necessarily entail a total abandonment of FBP. Specifically, what is at issue here is its explanation of mathematical knowledge. This requires something more than logic. Conversely, FBP is fundamentally a metaphysical thesis: the thesis that all possible mathematical universes exist. But, this metaphysical thesis and the epistemology envisioned by Balaguer are independent. Thus, it is perfectly consistent to

deny the reduction of mathematical knowledge to logical knowledge while accepting this metaphysical pluralism. As in the case of multiversism, we need to restrict the extent of the multiverse to which we have access. Logic places mathematicians outside this mathematical multiverse, giving them a whole perspective of its totality. But, this is too strong. Mathematicians are placed inside the mathematical multiverse and need a departure point from where their journey starts. Thus, what is required is an explanation of a minimal notion of mathematical objectivity that secures this starting point: we need an objective minimal mathematical theory. The resulting theory supported by such a minimal amount of objectivity should include enough mathematics to develop the mathematical techniques needed to understand alternative universes. Then, the metaphysical debate between fictionalists and multiversists would only affect the nature of these alternative universes. That is, while a minimal amount of objectivity justifies at least one theory as true, it will remain open whether the other theories formulated in its language are also true—in other universes—or just fiction.

4.7 CONCLUSION

In this chapter, I have analyzed multiversism, a philosophical position claiming that there is no single universe of mathematical, abstract objects, but many of them. In particular, I have focused on its two most extreme forms: Hamkinsian multiversism and full-blooded platonism. Neither of them imposes limits on the extent of the multiverse. This is why both can reduce mathematical knowledge to logical knowledge. I have argued that none of them provide us with a satisfactory explanation of mathematics. Hamkinsian multiversism cannot explain how mathematical vocabulary acquires its meaning. More specifically, its explanation of mathematical reference leads to circularity. In turn, full-blooded platonism lacks an appropriate notion of consistency for explaining mathematical knowledge. To use a primitive notion of consistency or possibility it needs to solve the language acquisition challenge. However, full-blooded Platonism cannot do it on its own.

APPENDIX 4.A

Contemporary mathematics is characterized by a fascinating phenomenon: its ability to establish its own limits. The case of set theory is particularly striking because independence results have become central to the discipline. Since Gödel's proof of the relative consistency of AC and CH with ZF in the 1930s, and the subsequent proof of the relative consistency of their negation by Cohen in the 1960's, a huge variety of model-theoretic techniques have been developed to show that for a sentence φ it is not the case that $ZF \vdash \varphi$ or $ZF \vdash \neg\varphi$ —usually assuming that ZF is consistent.

Theorem 1. (Gödel, 1938) *If ZF is consistent, then ZFC+CH is consistent.*

Theorem 2. (Cohen, 1963) *If ZF is consistent, then ZFC+ $\neg$ CH is consistent.*

It is known that ZF is highly incomplete. Indeed, as ZF is recursively axiomatizable and PA is interpreted by this system, Gödel's incompleteness theorems hold in ZF. Consequently, there are several Π_1^0 -sentences φ —among them G_{ZF} and $\text{Con}(ZF)$ —such that $ZF \not\vdash \varphi$ and $ZF \vdash \neg\varphi$. All these are *benign* cases of incompleteness (68, p. 36). More resistant examples are sentences of third-order arithmetic,

such as CH—which is a Σ_1^2 -sentence—or the Axiom of Constructibility ($V = L$)—which is Π_1^2 . In the terminology of (68) these are *serious* cases of incompleteness, because unlike Π_1^0 -sentences their independence is not materially sufficient for their truth. They are so resistant to a solution that several set theorists regard them as cases of *absolute* undecidability, that is, undecidability from any natural and justified extension of ZF—for instance, originally Gödel considered $V = L$ as a plausible candidate for absolute undecidability (50; 69).

However, there were several attempts to overcome these limitations and find stronger principles to deal with hard cases of incompleteness. The most famous of these attempts is the *program for large cardinal axioms*—see (92), (94, Ch.2), (69; 70), and (132) for a brief survey of the program. Gödel proposed to complete ZF by broadening the relevant notion of mathematical justification. He distinguished between two different kinds of justification: *intrinsic* and *extrinsic*—see (91; 92) for a detailed analysis of the multiple sources of evidence used to justify set-theoretic axioms. Intrinsic justifications follow from the very concept of *set*. More precisely, a sentence is intrinsically justified when the sentence obtained relativizing its quantifiers to the concept of set holds analytically. The main example is the axiom of Extensionality. Allegedly, the following instance of the axiom

$$\forall z(z \in a \leftrightarrow z \in b) \leftrightarrow a = b$$

follows analytically from the sentence asserting that a and b are sets. Therefore, when people say things such as “the axioms are *self-evident*” or “the axioms are *obvious*” they are appealing to this particular source of evidence. Traditionally, almost all the axioms of ZF—with the exception of Powerset and Replacement—were accepted on the basis of arguments which tried to show the intrinsic evidence for their truth. Gödel considered that intrinsic forms of evidence were strong enough to support more statements than those settled by ZF. The main example are the axioms stating the existence of *small large cardinals*—inaccessible, hyperinaccessible, Mahlo, or weakly compact—and *reflection principles*, which entail the existence of large cardinals up to ineffable cardinals (68; 70).

However, these forms of justification are not enough for the universalist. They are far from narrowing down the space of ZF’s extensions up to a unique, complete theory. Here is where *extrinsic* justifications enter the picture. Gödel proposed that, similarly to non-observational sentences in the case of empirical science, new axioms could be justified inductively by pragmatic and heuristic reasons. Specifically, because their *fruitfulness* in consequences.

Even disregarding the intrinsic necessity of some new axiom, and even in case it had no intrinsic necessity at all, a decision about its truth is possible also in another way, namely, inductively by studying its “success,” that is, its fruitfulness in consequences and in particular in “verifiable” consequences, i.e., consequences demonstrable without the new axiom, whose proofs by means of the new axiom, however, are considerably simpler and easier to discover, and make it possible to condense into one proof many different proofs. (49, p. 521)

If a candidate for completing ZF afford us proofs of new theorems and considerably simplifies the proof of old ones it could be regarded as sufficiently justified—recently, fruitfulness has been considered by

(132) and (94) as a successful criterion for objectivity. Moreover, Gödel proposed large cardinal axioms as the main candidates for significantly reducing the incompleteness of ZF. He even proposed an informal notion of absolute provability, provability from true large cardinal axioms, and conjectured a generalized completeness theorem: “every proposition expressible in set theory is decidable from the present axioms plus some true assertion about the largeness of the universe of all sets.” (Gödel, 151)

Inductive methods have shown as a promising way to the highest reaches of the hierarchy. It is known that they support stronger axioms such as axioms asserting the existence of *large large cardinals*—measurable, strongly inaccessible, Woodin, or supercompact. This family of axioms satisfies Gödel’s two requirements. For instance, the existence of measurable cardinals was shown to be inconsistent with $V = L$ by Scott. This result supplies Gödel’s conjecture with some plausibility, since his original candidate for absolute undecidability is decided by large cardinal assumptions. Furthermore, large cardinal axioms decide a variety of properties of sets of reals codified in the axiom of Projective Determinacy (*PD*) and the Axiom of Determinacy (*AD*) relativized to definable sets of reals—i.e., sets of reals in $L(\mathbb{R})$ —($AD^{L(\mathbb{R})}$) (69, p. 170).

Theorem 3. (Martin and Steel) *Assume there are infinitely many Woodin cardinals. Then PD.*

Theorem 4. (Woodin) *Assume there are infinitely many Woodin cardinals and a measurable cardinal above them. Then $AD^{L(\mathbb{R})}$.*

Indeed, Shoenfield has shown that large cardinal axioms *freeze* Σ_2^1 -sentences—a subcollection of analysis—in the sense that set forcing constructions preserve their truth value. That is, if any of these axioms is plainly true—or false—in V and an extension $V[G]$ is obtained adding a generic $G \subseteq \mathbb{P}$ —where $\mathbb{P}$ is a poset—, then φ will be also true—or false—in $V[G]$. This means that Σ_2^1 -sentences are *generic absolute*. Their consistency cannot be shown by means of set forcing, one of the strongest model-theoretic techniques.

Theorem 5. (Shoenfield) *Let φ be a Σ_2^1 -sentence, $\mathbb{P}$ a poset, and let $G \subseteq \mathbb{P}$ be V -generic. Then $V \models \varphi$ if and only if $V[G] \models \varphi$.*

The situation generalizes to Σ_n^1 , for greater n , but always under the level Σ_1^2 -sentences. Woodin improved the result, proving that the theory of definable sets of reals—all Σ_n^1 sentences, for all n —and, indeed, the theory of $L(\mathbb{R})$, are generic absolute under stronger large cardinal assumptions. In this sense, large cardinal axioms are “effectively complete” with respect to sentences of analysis concerning $L(\mathbb{R})$, in the sense that their truth value remains unaltered by forcing. That is, if forcing yields a model $V[G]$ making any of these sentences φ true—or false—in $V[G]$, then φ is plainly true—or false—in V .

Theorem 6. (Woodin) *Assume there is a proper class of Woodin cardinals. Let $\mathbb{P}$ be a poset, and let $G \subseteq \mathbb{P}$ be V -generic. Then $L(\mathbb{R}) \models \varphi$ if and only if $L(\mathbb{R})^{V[G]} \models \varphi$.*

The result can be generalized beyond descriptive set theory to further, more complex sets of reals, freezing subtheories in Δ_1^2 sentences—see (70), (94, Ch. 2) for a brief exposition of the state of the art of the discipline. Thus, the program for large cardinal axioms is a complete success below CH. They yield a completeness result—in the sense of theorems 2.5 and 2.6—for all questions of structural complexity concerning the reals. However, here is exactly where their strength and ability to supply

us with knowledge about the universe stops. Levy and Solovay showed that large cardinal axioms are preserved by forcing constructions too.

Theorem 7. (Levy and Solovay) *Suppose that κ is a measurable cardinal and $\mathbb{P}$ is a poset such that $|\mathbb{P}| < \kappa$. Let $G \subseteq \mathbb{P}$ be V -generic. Then $V[G] \models “\kappa \text{ is measurable}”$.*

The result was generalized to all familiar large cardinal axioms. On the contrary, CH and other Σ_1^2 -sentences can be altered by this kind of “small” —because of the limited size of $\mathbb{P}$ — forcing constructions. “None of the familiar large cardinal axioms can settle CH. Thus there can be no generalized completeness theorem of the sort Gödel entertained in 1946 and the program for large cardinals must be considered a failure at the level of CH.” (69, p. 167)

5 Abstraction and Mathematical Knowledge

5.1 INTRODUCTION

Fictionalism and the more extreme forms of multiversism cannot explain mathematical knowledge. The reduction of mathematical knowledge to logical knowledge presupposes that mathematicians already have a good understanding of the mathematical language. However, neither of these two proposals can offer a suitable solution to the Language Acquisition Challenge (LAC). That is, fictionalism and multiversism cannot explain how mathematical vocabulary acquires its meaning. Consequently, an alternative explanation is needed.

In this chapter, I will explain how the abstractionist approach to mathematical objects advocated by (82) solves the challenge. First, I will explain what abstraction principles are. These principles introduce a reductionist perspective in metasemantics. As a result, they can be used to explain how a community of speakers enrich their linguistic resources without attributing to them any relation to the objects the language is about—in this case, mathematical objects. That is, they can be used to solve LAC without ‘presupposing any counterfactual or logical relation between mathematicians and mathematical facts. However, abstraction principles are not free of problems. Although they are promising, they face two famous challenges: the Caesar problem and the Bad Company problem. I will explain what each of these problems is and how they can be solved. Finally, once we are sure that abstraction principles are in good condition, I will show how they can be used to explain mathematical knowledge too. The abstraction process by which speakers extend their language also provides them with knowledge of a collection of truths—and not only logical truths. Thus, as Linnebo (79; 82) and Rayo (116) claim, mathematical knowledge comes with the understanding of language.

5.2 ABSTRACTIONISM

In his *Grundlagen der Arithmetik*,¹ (46, § 62, p. 73) Frege asks the following question:

How, then, are numbers to be given to us, if we cannot have any ideas or intuitions of them?

The echoes of this resounding question reach out to us, while it waits to be answered. And, the answer comes in the form of what has been called *abstractionism*. This is a *top-down*—the terminology is from (135, p. 1)—and *stipulative* approach to metasemantics. Even though mathematical objects are abstract

¹Frege (46). References refer to this edition.

and resist all attempts of ostensive and descriptive nomination, a community can stipulate mathematical reference by laying down the appropriate definitions. As a result, abstractionism explains how terms such as “the number of X s” —in symbols “ $\#(X)$ ”—, where “ X ” is a second-order variable ranging over second-order entities —such as concepts, properties or pluralities—, successfully refer by decomposing the answer of the above question into two simpler steps.² First, the content of a whole collection of sentences in which the term features is explained in terms of stipulation. Frege, as well as most abstractionists, focuses on identities “ $\#(X) = \#(Y)$ ”. The semantic content of these identities is stipulated according to a definition of the form³

HP

$$\#(X) = \#(Y) \leftrightarrow X \approx Y,$$

where “ $X \approx Y$ ” abbreviates “ X and Y are equinumerous.”⁴ This definition has become known as *Hume’s Principle* (HP). It stipulates that the sentence “the number of X s is equal to the number of Y s” has the same content as its right side, “ X and Y are equinumerous”.

Second, the semantic properties accorded on the identity impose enough constraints for the referents of “ $\#(X)$ ” and “ $\#(Y)$ ” to be selected. Specifically, if X and Y are indeed equinumerous, the identity “ $\#(X) = \#(Y)$ ” —by stipulation— comes out as true and, modulo the compositional nature of semantics, the semantic value of the sentential subexpressions is selected to meet the constraint that these compositionally determine the semantic value of the identity —which also turns to be true. According to Frege’s own words:

Our aim is to construct the content of a judgement which can be taken as an identity such that each side of it is a number. We are therefore proposing not to define identity specially

²Each step involves a major principle of Frege’s philosophical thinking: the *context principle* (46, p. 10), a metasemantic principle, and the semantic *principle of compositionality*. For a careful analysis of the status of both see (82, Ch. 7). Following Linnebo, I understand the context principle as an *explanatory principle* governing the explanations of how expressions acquire their meaning. The fact that an expression has meaning is not a metaphysical primitive —as, if there were such primitives, could be the self-identity of any object. Therefore, if an expression has meaning this must be accounted for. The context principle requires us to explain its meaning in virtue of its contribution to the meaning of the complex expressions in which it features.

Context Principle The meaning of an expression ε is explained in virtue of its contribution to the meaning of the complex expressions in which ε features.

On the contrary, the principle of compositionality is a *descriptive principle* of semantics. That is, it is a universal generalization which states that the meaning of any expression is a function of the simpler expressions composing it.

Principle of Compositionality The meaning of a complex expression e is a function of the meaning of simpler expressions composing e .

Sometimes, the two principles have been considered to be incompatible with each other. However, once the different roles of each one are understood, all tension vanishes.

³Throughout this chapter, sentences of this form must be read as the corresponding universal closure, unless otherwise stated.

⁴The notion of *equinumerosity* can be formalized in second-order logic as usual:

$$X \approx Y \equiv_{df} \exists Z(\forall x(Xx \rightarrow \exists!y(Zxy \wedge Yy)) \wedge \forall y(Yy \rightarrow \exists!x(Zxy \wedge Xx))).$$

for this case, but to use the concept of identity, taken as already known, as a means for arriving at that which is to be regarded as being identical. Admittedly, this seems to be a very odd kind of definition, to which logicians have not yet paid enough attention; but that it is not altogether unheard of, may be shown by a few examples. (46, §63, p. 73)

Another example are directions. The identity of directions is given by the following equivalence
Dir

$$d(l_1) = d(l_2) \leftrightarrow l_1 \parallel l_2,$$

where “ $d(l_i)$ ” abbreviates “the direction of line l_i ” and $\parallel$ is the relation of parallelism. Thus, two lines specify the same direction if and only if they are parallel. According to the first step of the abstractionist explanation, the semantic content of the identity “ $d(l_1) = d(l_2)$ ” is stipulated to be the same as the right-hand side of the equivalence. It will be true just in case the line l_1 is parallel to the line l_2 . Then, according to the second step, directions are selected as referents of “ $d(l_1)$ ” and “ $d(l_2)$ ” subject to the constraint that they compositionally determine the semantic value of the identity. This is why the explanation is endowed with a top-down flavour. The semantic value of a range of sentences is stipulated to accord with some previous standards. Once this is done, reference is explained in terms of these semantic values and compositionality.

Similarly, rational numbers can be given in terms of equivalences. Let “ $r(n, m)$ ” abbreviate “the rational number corresponding to the pair $\langle n, m \rangle$ of integers n and m ”. Then, rationals can be introduced using the equivalence relation $n_1 \cdot m_2 = n_2 \cdot m_1$ defined on pairs of integers.

RTL

$$r(n_1, m_1) = r(n_2, m_2) \leftrightarrow n_1 \cdot m_2 = n_2 \cdot m_1.$$

That is, the rational corresponding to the pair of integers $\langle n_1, m_1 \rangle$ is the same as the rational corresponding to the pair $\langle n_2, m_2 \rangle$ if and only if the product of n_1 and m_2 is the same as the product of m_1 and n_2 . Indeed, this accords with the usual definition of rationals. As Frege said, the history of mathematics is full of examples.⁵

An *abstraction principle* is an equivalence of the form

AP

$$f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta,$$

where α and β are variables ranging over objects⁶ of some kind, $\sim$ is an equivalence relation⁷ defined on these objects and “ f ” is a singular-term-forming-operator. Let’s call the complex terms of the form

⁵For a detailed exposition of the use of abstraction principles in the history of mathematics see (95). Also, Hale (51) has shown how Dedekind’s definition of reals (28) can be explained in terms of abstraction principles.

⁶Here, the notion of object must be understood in a broad sense. It includes first-order entities (individuals), as well as higher-order ones (concepts, properties, relations, etc.). Most of the time the context will suffice to determine the order of variables. I will use Greek symbols “ α ”, “ β ”, both as metavariables standing for terms of a particular n -order language and as n -order variables standing for the objects this language speaks about. Instead of quotes, the context will suffice to disambiguate both uses. This greatly simplifies the exposition.

⁷An equivalence relation is a relation which is reflexive, symmetric and transitive. Equinumerosity, parallelism or the relation holding between pairs $\langle n_1, m_1 \rangle$ and $\langle n_2, m_2 \rangle$ of integers when $n_1 \cdot m_2 = n_2 \cdot m_1$ are cases of equivalence relation.

“ $f(\alpha)$ ” *abstract terms* and their referents—if there were any—*f*-abstracts or simply *abstracts*. Also, let’s call the entities over which the variables α and β range *specifications*. The reason is that these objects specify the abstracts for which AP holds. Finally, let’s call the equivalence $\sim$ a *unity relation*. AP relates the identity of abstracts $f(\alpha)$ to the unity relation $\sim$ holding between the specifications.⁸ Abstractionism makes strong use of abstraction principles to explain how mathematical terms successfully refer and, *a fortiori*, to answer the aforementioned question of how numbers, as well as other inhabitants of the mathematical realm, are given to us.

5.2.0.1 The Structure of Abstraction Attempts

Consider an idealized community of speakers who initially have a language $\mathcal{L}_1$. Let’s call the language $\mathcal{L}_1$ the *base language* of abstraction. Let’s assume that $\mathcal{L}_1$ does not contain the singular-term-forming-operator f required to form abstract terms, as well as other similar expressive resources. Thus, the language $\mathcal{L}_1$ is free from ontological commitments concerning the abstracts satisfying AP. At most, it contains enough expressive resources to deal with the equivalence $\sim$ and other linguistic resources required to speak about the specifications on which this is defined. Then, speakers will use this resources to express a variety of facts concerning the equivalence. Specially, they will be able to state the right-hand side of AP. On the contrary, at this initial state the left-hand side remains unnoticed by them.

An attempt to extend $\mathcal{L}_1$ by means of AP takes three steps—for similar presentations of abstractionism in terms of communities of speakers see (82, pp. 148-151) and (135; 136).

- (i) First, the community extends the base language $\mathcal{L}_1$ by adding an operator f which applied to n -order variables α, β , etc. of the base language form expressions $f(\alpha), f(\beta)$, etc. —at this point of the explanation, it must not be assumed that these expressions are terms. Let’s call the resulting language the *extended language* $\mathcal{L}_2$. Additionally, it also contains a symbol “=” —at this point, it must not be assumed that this is the identity predicate. Then, the community uses the sentences of the extended language in accordance with some set of assertibility conditions stated in the base

⁸Abstraction principles provide a definition of abstracts. As Frege (46, §63) claims, “this seems to be a very odd kind of definition”. However, this sort of definition can be understood as a contextual definition, similar to the definition of truth afforded by the T-schema

$$\varphi \text{ is true iff } \varphi.$$

This definition offers a criterion—a truth condition—which allows us to identify when a sentence φ satisfies the concept of truth and when it does not—that is, when φ is true and when it is not. Thus, the concept of truth is explained in terms of its extension. Similarly, AP provides a criterion stated in terms of a unity relation that identifies those elements—the abstracts $f(\alpha)$ —satisfying the concept being defined. Indeed, if we had a predicate C expressing the concept being defined, AP would be equivalent to a principle identical in form to the T-schema

$$C(f(\alpha)) \leftrightarrow \alpha \sim \alpha.$$

The concept C satisfied by the abstracts is explained in terms of its extension. This reading of abstraction omits further comments on how the semantic content of sentences expressing facts about the equivalence relates to the content of identities between abstract terms. In particular, it omits Frege’s (46, §65) and Rayo’s (116) attributions of sameness of content, or Linnebo’s (81; 82) metaontological minimalism. This is equivalent to saying that such a reading does say nothing about how to interpret the equivalence $\leftrightarrow$. More on this on 5.3.

language. In particular, the use of *apparent* identities of the form $f(\alpha) = f(\beta)$ is governed by the assertibility condition $\alpha \sim \beta$ —see 5.6 for a precise statement of the assertibility conditions. Thus, the rule governing the use of such *apparent* identities is given by the corresponding instance of AP. By obeying this rule, members of the community come to speak *as if* there were abstracts $f(\alpha)$, $f(\beta)$, etc.

- (ii) Second, the content of new $\mathcal{L}_2$ -sentences φ is stipulated by φ 's assertibility condition. Specifically, the content of identities $f(\alpha) = f(\beta)$ is stipulated by the corresponding $\mathcal{L}_1$ -sentence $\alpha \sim \beta$ about the unity relation $\sim$ and specifications α and β . Thus, $f(\alpha) = f(\beta)$ is true if and only if $\alpha \sim \beta$ is also true. The assertibility conditions turn out to be truth conditions. Importantly, this specification does not rely on the ability of speakers to translate $\mathcal{L}_1$ expressions into $\mathcal{L}_2$.
- (iii) Third, semantic values for subexpressions of new $\mathcal{L}_2$ -sentences φ are selected to meet the constraint that these compositionally determine the semantic value of φ . Thus, if the abstraction attempt is successful, expressions $f(\alpha)$, $f(\beta)$, etc. are assigned abstracts as their semantic values. Thus, $f(\alpha)$ and similar expressions are terms referring to a particular object. Speakers, as a result of a successful abstraction attempt, are furnished with the ability to speak about these abstracts.

This three-step account summarizes the main features of abstractionist accounts of reference. The sketch can be specified in slightly different ways attending to how stipulation, sentential content, the relation between sentential content and reference, and so on are understood. Either way, I will assume this suffices to understand the kernel of the proposal.

Before going on, let me address a reasonable worry concerning the previous picture of abstractionism. The first step involves talking about how the lexicon of $\mathcal{L}_1$ is expanded. In particular, the operator f —and, presumably, other sorts of expressions—is added to $\mathcal{L}_1$. However, semantics only enters the explanation in the second step. Furthermore, the second step is only concerned with sentential content. The third step is the only one which mentions terms and reference. However, the arguments in Chapters 3 and 4 rely on the maxim that syntactic categories, such as the category of terms, are already semantic categories. So, does not the first step of the explanation lead back to the same problems concerning syntax faced by fictionalism and multiversism? For, when talking about the operator f is involved, this seems to qualify it as a purely syntactical element to which speakers accord some meaning. But, if this is the case, the concept of syntax and language at use must be a formal and mathematical one. Consequently, a vicious circle arises: the abstractionist explanation presupposes an understanding of mathematical facts and will be useless to explain mathematical reference.

The answer to this worry is to observe that the above three steps are *explanatory*. Any talk about syntactic and semantic categories featuring in them is performed by the *theoretician* who is trying to explain the speakers' linguistic practice. Conversely, speakers are not required to understand what they are doing and, thus, they are not required to understand this explanation. Therefore, the three steps composing the abstractionist explanation do not correspond to a temporal sequence of events underlying speakers' practice. In particular, speakers do not extend $\mathcal{L}_1$ by employing a formal language to which they attach some entities as semantic value—nor are they required to understand what a formal language

is. The unique stipulation they are required to do is a stipulation concerning the language they end up speaking. But, this is radically different from model-theoretic interpretations of formal expressions. So, abstractionism is free from the problems affecting pluralist philosophies of mathematics.

5.2.1 Language Expansions

Now, I will show that the following conditional is true:

If abstractionism can explain mathematical reference, it can also explain how speakers acquire a whole new language.

Again, take the case of directions. A variety of properties and relations between directions are introduced from properties and relations on lines. For instance, orthogonality. Two directions $d(l_1)$ and $d(l_2)$ are regarded as orthogonal if and only if the lines l_1 and l_2 are orthogonal. Let $\perp$ and $\perp^*$ be orthogonality between lines and directions respectively. Then, the previous claim can be symbolized as follows:

$$d(l_1) \perp^* d(l_2) \leftrightarrow l_1 \perp l_2.$$

Similarly, a variety of properties and relations on rationals are also “inherited” from properties and relations of pairs of integers. This is the case of the ordering between rationals. Let $<$ and $<^*$ be the usual orderings on integers and rationals respectively. Then

$$r(n_1, m_1) <^* r(n_2, m_2) \leftrightarrow n_1 \cdot m_2 < n_2 \cdot m_1,$$

—if m_i is negative, then other n_k and m_k such that $n_i \cdot m_k = n_k \cdot m_i$ must be found. These examples show that the relations of orthogonality between lines and the ordering relation between rationals are “inherited” from analogous relations on specifications. Indeed, this phenomenon takes place when the latter meets the appropriate condition. Namely, that the property or relation does not discriminate between specifications related by the unity relation $\sim$. Thus, assume that lines l_1 and l_2 are parallel, and that l_2 is orthogonal to l_3 . Then, l_1 and l_3 will be also orthogonal —parallelism is a *congruence* with respect to the relation of orthogonality. Generally, for any sentence ϕ which does not discriminate between parallel lines —any sentence such that $\parallel$ is a congruence with respect to ϕ —, a predicate ϕ^* can be introduced such that it holds of the directions of some lines if and only if ϕ holds of the very lines.⁹ A whole language about directions can be acquired through abstraction from a language only committed to lines.

This explanation can be generalized to other languages as follows. Let’s say that the unity relation $\sim$ is a congruence with respect to a $\mathcal{L}_1$ -formula $\phi(\alpha_n \dots \beta_n)$ if and only if it satisfies the following condition¹⁰

$$\alpha_1 \sim \beta_1 \wedge \dots \wedge \alpha_n \sim \beta_n \rightarrow (\phi(\alpha_1 \dots \alpha_n) \leftrightarrow \phi(\beta_1 \dots \beta_n)).$$

⁹“The meaning of any other type of assertion about directions would have first of all to be defined, and in defining it we can make it a rule always to see that it must remain possible to substitute for the direction of any line the direction of any line parallel to it.” (46, § 65)

¹⁰Again, the formula must be read as the corresponding universal closure.

The formula $\varphi(\alpha_1 \dots \alpha_n)$ does not discriminate between equivalent specifications. Moreover, for simplicity, consider formulas which only hold in the field of $\sim$:

$$\varphi(\alpha_1 \dots \alpha_n) \rightarrow \alpha_1 \sim \alpha_1 \wedge \dots \wedge \alpha_n \sim \alpha_n.$$

That is, if $\varphi(\alpha_1 \dots \alpha_n)$ holds of some specifications $\alpha_1, \dots, \alpha_n$, then these are related by $\sim$ to other specifications, and —because $\sim$ is reflexive— they are related by $\sim$ to themselves. Then, for any formula satisfying these two conditions, abstraction explains how an n -ary predicate about abstracts satisfying the following *inheritance principle* is introduced

Inheritance

$$\varphi^*(f(\alpha_1) \dots f(\alpha_n)) \leftrightarrow \varphi(\alpha_1 \dots \alpha_n).$$

In consequence, the abstractionist three-step explanation can be enhanced by adding the assertibility conditions corresponding to the each predicate φ^* —see 5.6. A whole new set of useful expressions is available to speak about the abstracts.

5.2.2 Reductionist Metasemantics

Abstractionism offers a workable metasemantics which looks promising to deal with the case of abstract objects. The key feature of this metasemantic account is its *reductionist nature*. To explain how direction-terms d successfully refer, the abstractionist only makes use of lines l and an equivalence relation $\sim$ with which speakers are familiar. Also, the explanation of how a relation of congruence $\perp^*$ holds on directions $d(l_1)$ and $d(l_2)$ only involves the lines l_1 and l_2 and, once more, $\sim$. Directions themselves are not used to explain these facts. While abstracts feature in the *explanandum* —what is explained is how terms of the form $f(\alpha)$ refer to these abstracts—, they do not feature in the *explanans*.

This is why abstraction and inheritance principles are so appropriate to deal with cases of reference to abstract objects. Together, they impose a *minimal requirement* for reference: to assign truth-conditions to identities and other sentences which respect the logic of identity¹¹. (81; 82; 84)

A minimal requirement is that truth-conditions have been assigned to all identity statements and other predication that involve the new singular terms to be introduced. Moreover, this assignment must be done in a way that respects the laws of logic. The most important part of logic in this connection is what we may call the logic of identity, which describes the identity relation and its interaction with predication. (82, p. 30)

Furthermore, abstraction principles are reasonably strong. For instance, HP is enough to interpret in second-order logic PA's axioms —this result, originally suggested by (99) and proved by (150), is known as *Frege's Theorem*. All this makes a strong case for abstraction principles to be considered as serious candidates for the design of an account of mathematics that avoids the challenges exposed in Chapters 3 and 4.

¹¹Frege's (46, §62, p. 73) words: “we shall be giving a general criterion for the identity of numbers. When we have thus acquired a means of arriving at a determinate number and of recognizing it again as the same, we can assign it a number word as its proper name.”

5.3 METAONTOLOGICAL MINIMALISM

Originally, Frege’s interest in abstraction principles was motivated by his desire to reduce arithmetic to logic. Simplifying things a bit, he regarded HP as *a definition* of the concept of cardinal number, and thus as a logical principle. Therefore, platonism, abstractionism and logicism —the thesis that arithmetic can be reduced to logic— are mixed together in his philosophical project. Similarly, neo-logicism (150) regards HP as an implicit definition —thus, as an analytic truth— of cardinal numbers. But, unlike Frege, they do not consider it a logical law. Neo-logicism relies on Frege’s Theorem and the analytic nature of definitions to claim that arithmetic is also analytic. Both logicism and neo-logicism read the biconditional $\leftrightarrow$ in HP as an analytic operator $\Leftrightarrow$

$$\#(X) = \#(Y) \Leftrightarrow X \approx Y,$$

expressing some kind of analytic relation between its two sides.¹² Modulo this reading of abstraction principles, both projects aim to reap the benefits of what seems a promising explanation of mathematical knowledge. If arithmetical truths are logical truths, or at least they are analytically true, it is possible to gain knowledge about them by laying down the right definitions —definitions of the form of AP.

However, the connection between abstraction and analyticity is not a matter of necessity. As Linnebo (81), (82, p. 3) —by appealing to Kant and Boolos— points out, it is controversial that analytic claims entail existence. Logic and conceptual analysis are not enough to settle the debate about the existence of mathematical objects. This is why he suggests an alternative reading of abstraction as a form of *metaphysical explanation*, which retains the epistemological benefits promised by analyticity. According to his reading, abstraction principles provide us with an explanation of what it takes, that is, of what is *metaphysically sufficient* for some objects —such as cardinal numbers, rationals, or directions— to exist. Similarly, inheritance principles explain what is *metaphysically sufficient* for some facts —those facts depicted by atomic sentences of the form $\varphi^*(f(\alpha_1)...f(\alpha_n))$ — to be the case. Consequently, instead of wondering whether there are analytic existence claims, this reading shifts attention to another question: are there *thin objects* “in the sense that their existence does not [...] amount to very much? Presumably, an analytic truth does not make a substantial demand on the world. But perhaps being analytic is not the only way to avoid imposing a substantial demand” (Linnebo, 2018, p. 3). A metaphysical reading of abstraction and inheritance principles suggests that the answer to this question is a resounding “yes” and the metaphysical picture arising from such a reading is known as *metaontological minimalism*. (81; 82)

Ontology is the study of what there is. Metaontology is the study of those concepts involved in ontology, such as the concepts of object, existence or fact. Then, metaontological minimalism is the thesis that such concepts are *minimal* or *thin*. And, a thin concept of object allows for thin objects: objects whose existence “does not impose substantive demands on the world.” Minimalists do not claim

¹²On several occasions $\Leftrightarrow$ was read as expressing some kind of “sameness” in semantic content or meaning. This is usually referred as Frege’s *recarving thesis*. The thesis claims that the identity $\#(X) = \#(Y)$ and $X \approx Y$ share the same content, which is *carved up* in different ways by each sentence. While the right side of HP is not committed to cardinal numbers, the identity recarves its content yielding commitment to such entities. However, this idea of “recarving” as well as its relevance in Frege’s philosophy is highly controversial. See (46, § 64), (150) and (52) for uses of the thesis. Recently, Rayo (116) has advocated the idea of sameness in content and recarving by means of his *just-is* statements. See (82, Ch 7) for a careful analysis of the role of recarving in the *Grundlagen* and *Grundgesetze*, and its relation to the context principle.

that all objects are thin in this sense. Trees, mountains, or —token— books are quite *thick* objects. Their existence requires very specific physical and contingent circumstances. Thus, for a —token— book to exist it is required a complex chain of events leading to some agents developing language, creating paper and finally creating the book by printing some ink marks on the paper and arranging this in the form of a book. However, minimalists claim that thickness comes from the kind of objects trees, mountains or —token— books are. Thickness should not be loaded on the concept of object itself. Historically, there have been a variety of versions of this idea. Each one of them provides an analysis of what “to be thin means” —that is, an analysis of what it means not to impose substantive demands on the world in order to exist. Coherentism is a well-known example. According to coherentism, the consistency of a theory *suffices* for the existence of the objects the theory describes to exist.¹³ For example, the existence of natural numbers only demands from the world that arithmetic is a consistent theory. More recent examples of minimalism are Eklund’s (32) *priority-based minimalism*, Thomasson’s (141) *easy ontology*, or Ferreirós’ (38) practice-based account of mathematical objects.

Linnebo’s proposal is that abstraction principles provide an alternative explanation of the metaphysical idiom of sufficiency. To the question raised by Koellner (70, p. 113, n. 48)

How cheap is existence in mathematics? Does consistency suffice? Consider ZFC+CH versus ZFC+¬CH. There is reason to believe that both are consistent. The trouble is that CH and ¬CH are existential claims and, on a straightforward reading, the objects that they assert to exist cannot coexist. I am inclined to think that existence in mathematics is “Consistency + X” but I do not know how to solve for X” [italics in the original],

abstractionist minimalism suggests solving *X* via abstraction and inheritance principles. The result is an abstractionist version of metaontological minimalism.

In consequence, abstraction is not only a metasemantic mechanism for fixing reference. Furthermore, it delivers an explanation of what it takes for some objects to exist. Again, consider material bodies —such as trees, mountains or books. Assume talk about parcels of matter is in order. Surely, a *necessary* condition for the existence of material bodies is that some parcels of matter —the paper and ink stains— are spatiotemporally coordinated in a very specific way. First, these parcels must be coordinated under motion in such a way that any two parcels change their spatio-temporal location if and only if their relative positions are preserved by the movement. More formally, two parcels *u* and *v* move from *a*₁ and *a*₂ to *b*₁ and *b*₂ respectively, if and only if $\rho(a_1) = a_2$ and $\rho(b_1) = b_2$ —where ρ is a permutation preserving the relative position of *u* and *v* to each other. Second, these parcels must be linked by a path of solid —in a non-technical sense of “solidity”— matter. For example, two spheres which satisfy only the former condition, and perform perfectly coordinated movements, do not constitute a physical body —in the usual sense. They must also be materially connected. Similarly, a flock of birds which have perfected their joint flight to the extreme does not constitute a material body either —even if the flock is given the

¹³Recall that criticism developed in Chapters 3 and 4 only affects consistency as a criterion for mathematical knowledge. On the contrary, coherentism is a metaphysical thesis claiming that consistency *suffices* for existence. For this reason, such criticism does not affect coherentism, as it does not affect FBP regarded just as a metaphysical position.

status of an “emergent object”, it is not a material body in the usual sense. Now, let $\equiv_b$ be the *partial equivalence*¹⁴ relation on parcels of matter defined by these two conditions. Then, for a material body to exist it is necessary for the parcels forming it to be related by the partial equivalence $\equiv_b$. However, it is highly plausible that this is also *sufficient* for its existence. This gives the following abstraction principle for physical bodies. Let $b(u)$ refer to the body to which the parcel of matter u belongs.

$$b(u_1) = b(u_2) \leftrightarrow u_1 \equiv_b u_2.$$

Moreover, similarly to the case of cardinal numbers, rationals and directions, a variety of predicates φ^* about material bodies can be introduced via inheritance principles. Thus, the equivalence $u_1 \equiv_b u_1$ is *metaphysically sufficient* for identities of the form $b(u_1) = b(u_1)$, and thus for the existence of $b(u_1)$ —assuming identity entails existence. Similarly, $\varphi(u_1 \dots u_n)$ is *sufficient* for $\varphi^*(b(u_1) \dots b(u_n))$ to hold.¹⁵

Generally speaking, abstraction reveals itself as an explanation of what it takes for some objects to exist—metaphorically speaking, of what the existence of these objects “demands on the world.” Consequently, the biconditional $\leftrightarrow$ featuring in AP is read as an stronger *sufficiency operator* $\Rightarrow$:

$$\alpha \sim \beta \Rightarrow f(\alpha) = f(\beta).$$

This reads: the fact that $\alpha \sim \beta$ —the specifications α and β are related by the unity relation $\sim$ —is *metaphysically sufficient* for $f(\alpha)$ and $f(\beta)$ to refer to the same abstract—thus, for this abstract to exist. Similarly,

$$\varphi(\alpha_1 \dots \alpha_n) \Rightarrow \varphi^*(f(\alpha_1) \dots f(\alpha_n))$$

says: the fact depicted by $\varphi(\alpha_1 \dots \alpha_n)$ is *metaphysically sufficient* for the fact depicted by $\varphi^*(f(\alpha_1) \dots f(\alpha_n))$ to be the case. This provides the following characterization of sufficiency via abstraction (82, p. 37).¹⁶

Consider [an abstraction principle] which purports to provide identity conditions for some kind F of object in terms that presuppose only some antecedently accepted ontology but not Fs. Assume that an agent has an appropriate grasp of the unity relation $\sim$ and uses the criterion and property inheritance principles of the form (Inher) to govern her discourse about Fs. Assume that the agent stands in an appropriate relation to specifications α and β , which are in the field of $\sim$, and uses these specifications as if to make claims about $f(\alpha)$ and $f(\beta)$. Then:

- (i) the agent is in fact referring to objects $f(\alpha)$ and $f(\beta)$,
- (ii) $\alpha \sim \beta \Rightarrow f(\alpha) = f(\beta)$,
- (iii) $\varphi(\alpha_1 \dots \alpha_n) \Rightarrow \varphi^*(f(\alpha_1) \dots f(\alpha_n))$.

¹⁴The relation is clearly symmetric and transitive. It must be partial because reflexivity can fail: not all parcels of matter give rise to a well-defined object in the previous sense. For instance, think about fluids or gases.

¹⁵The example of material bodies is from Linnebo (82, § 2.3). He himself acknowledges that it is a *toy case* to illustrate how abstraction explains the idea of metaphysical sufficiency.

¹⁶The original says “suffices for” instead of $\Rightarrow$. The same criterion for reference and sufficiency can be found in (81).

The strength of the operator $\Rightarrow$ is in the middle of the analytic $\Leftrightarrow$ and the strict conditional $\Box(\varphi \rightarrow \psi)$. It must be a less demanding notion than the analytic operator precisely because thin objects take attention away from analytic existence claims. Furthermore, it must be more demanding than the strict conditional. First, because otherwise, the well-known effect of the paradoxes affecting the strict conditional prevents a safe epistemological route to abstracts. Any conditional $\Box(\varphi \rightarrow \psi)$ such that ψ is necessarily true, is true. But, knowledge of φ does not ensure knowledge of ψ . For instance, suppose that set-theoretic truths are necessary truths. Moreover, suppose CH is false. Then, $\Box(\varphi \rightarrow \neg\text{CH})$ will be true even if $\varphi \equiv$ “Wittgenstein is the author of the *Tractatus*.” While this is known by every philosopher alive, nobody knows that $\neg\text{CH}$. This suggests that $\Rightarrow$ should be developed as a hyperintensional operator.¹⁷

If abstraction elucidates some form of metaphysical explanation, it can be used to characterize the idea of thin objects. Instead of asking whether there are objects whose existence is analytic—according to a more “traditional” account of abstraction—, we can ask which objects—to use the previous metaphor—“make little demands on the world.” Material bodies and similar physical entities are thick because their existence imposes contingent, physical requirements on the world. For the body $b(u)$ to exist, it suffices that $u \equiv u$:

$$u \equiv_b u \Rightarrow b(u) = b(u).$$

However, the partial equivalence relation $\equiv_b$ involves a causal chain of events leading u to be spatiotemporally located in the correct position for $u \equiv u$ to hold—recall the chain of events leading to the existence of books. Material bodies are thick because they have spatiotemporal location and are causally effective. This kind of objects are called *concrete*.

On the contrary, equinumerosity $\approx$ is not an spatio-temporal relation—provided the second order entities which X range over are not spatio-temporally located—, and does not involve any story about causal chains of events. Moreover, if HP is read along the lines of minimalism, then the fact that $X \approx X$ suffices for $\#(X) = \#(X)$. Thus, the fact that the concept X is equinumerous to itself suffices for the existence of the cardinal number to which $\#(X)$ refers:

$$X \approx X \Rightarrow \#(X) = \#(X).$$

Consequently, the existence of cardinals $\#(X)$ does not involve any substantial, causal relation. *Abstract objects*—such as cardinal or rational numbers—are objects whose existence does not impose any spatio-temporal demand on the world—they are specified by means of non-spatio-temporal entities and a non-spatio-temporal unity relation. Thin objects are so because little is required for their existence. In the middle of the concrete and the abstract realm, directions and types—such as sentence types—are an example of what is known as *quasi-concrete* objects. Somehow, these are thicker than numbers, but also thinner than mountains and birds. For instance, parallelism—the equivalence relation that suffices for the existence of directions $d(l_i)$ —is not a spatiotemporal relation:

$$l \parallel l \Rightarrow d(l) = d(l).$$

¹⁷Linnebo (82, p. 18) suggests a connection between his sufficiency operator and metaphysical grounding. Nevertheless, he also notes an important difference. As in the case of the strict conditional, even if φ grounds ψ it is doubtful that knowledge of φ transfers to knowledge of ψ . Also, abstraction principles seem related to the concept of *real definition*. Indeed, the mathematical examples provided by (120) are close to abstraction principles. The relation between all these metaphysical concepts remains open.

However, directions can be specified by concrete objects as well: some physical objects have direction.¹⁸

5.4 CHALLENGES TO ABSTRACTION

Abstraction principles must be in good condition if we want to use them to explain mathematical language and mathematical knowledge. However, it is unclear if this is the case. These principles have been opposed by two problems: the Caesar problem and the Bad Company Problem. In this section, I will explain what each of them is and how they can be solved. Thus, I will argue that the best solution to Caesar problem is the *piecemeal* solution advocated by (82, Ch. 9) and (135). In turn, I will argue that the best solution to the problem of bad company is to adopt a predicative reading of quantifiers and, thus, of abstraction principles.

5.4.1 The Caesar Problem

The Caesar Problem arises when one tries to answer the following awkward, but otherwise legitimate question:

- (1) Is Julius Caesar the number 3?

Let's call questions about mixed identities between abstracts —such as cardinal numbers, rationals or directions— and objects of another sort —for example, Julius Gaius Caesar— *Caesar questions*. The Caesar Problem is the silence of abstract principles when pressed to answer that kind of question. The problem was originally raised by Frege in the *Grundlagen*. As he recognizes, either HP or Dir

...will not, for instance, decide for us whether England is the same as the direction of the Earth's axis —if I may be forgiven an example which looks nonsensical. Naturally, no one is going to confuse England with the direction of the Earth's axis; but that is no thanks to our definition of direction. (46, § 66)¹⁹

Abstraction principles only provide abstracts with suitable criteria of identity —but no other kinds of objects. Given any two abstracts $f(\alpha)$ and $f(\beta)$, AP returns the condition $\alpha \sim \beta$ as a criterion for determining if the abstracts are identical or not — $f(\alpha) = f(\beta)$ if and only if $\alpha \sim \beta$.²⁰ However, AP is not powerful enough to tell what happens when the identity is a mixed identity involving an abstract $f(\alpha)$ and Julius Caesar, England, or a glass of water. Indeed, the problem motivated Frege's extension-based account of cardinals and the introduction of Basic Law V —leading his project to a contradiction.

¹⁸Are the concrete objects all and only the thick objects? Are the abstract objects all only the thin objects? I will leave this question open here. It would be interesting to know if the concepts of *thin* and *thickness* induce some ordering between metaphysical categories.

¹⁹Frege did not explicitly formulate the question. He originally mentions the undecidability of Caesar questions as a challenge to an alternative definition of number: “we can never —to take a crude example— decide by means of our definitions whether any concept has the number Julius Caesar belonging to it, or whether that same familiar conqueror of Gaul is a number or is not.” (46, § 56)

²⁰This does not entail that $\sim$ is decidable. See (97) for more on this question.

The Caesar Problem can be put as the conjunction of the following three, mutually inconsistent statements²¹—remember that $\mathcal{L}_1$ is the base language of abstraction, and $\mathcal{L}_2$ is the language resulting from the addition of abstract terms of the form $f(\alpha)$ and the predicates φ^* :

- C1 A successful abstraction attempt based on AP and Inheritance provides all abstract terms $f(\alpha)$ with a unique referent, while it does not affect the referent of terms in $\mathcal{L}_1$. Thus, after a successful attempt “Julius Caesar” is a term referring to Julius Caesar, conqueror of Gaul, and $f(\alpha)$ is a term referring to a unique abstract.²²
- C2 An abstraction attempt based on AP and Inheritance provides the identity predicate “=” with its usual syntax and semantics. In particular, if τ is a term in $\mathcal{L}_1$, then $\tau = f(\alpha)$ is a grammatical expression and it is true—if τ and $f(\alpha)$ have the same referent—or it is false—if τ and $f(\alpha)$ do not have the same referent.
- C3 An abstraction attempt based on AP and Inheritance does not decide the truth value of “Julius Caesar = $f(\alpha)$ ”: after the attempt, this remains neither true nor false.

C1 is supported by the usual reasoning underlying the abstractionist metasemantics. Advocates of this account argue that a successful abstraction attempt provides the abstract terms of $\mathcal{L}_2$ with a unique, well-determined referent: the corresponding abstract. In turn, C2 is a general feature of one-sorted languages. Besides, it seems reasonable that the resulting extended language $\mathcal{L}_2$ is one-sorted. If an abstraction attempt fixes the truth value of sentences committed to abstracts, the compositional nature of the base language $\mathcal{L}_1$ and the extended language $\mathcal{L}_2$ should guarantee that mixed sentences—sentences formed by logical operations which contain $\mathcal{L}_1$ -expressions committed to specifications and $\mathcal{L}_2$ -expressions committed to abstracts—, have also a well-determined truth value. Finally, C3 is the counterexample raised by Caesar Questions. Accordingly, solutions to the challenge can be classified depending on which one of the C1-C3 they reject.

5.4.1.1 Category Mistakes

An immediate and, to some extent expected, answer is to dismiss Caesar questions as meaningless—thus, to reject C2. According to this response, abstracts, as numbers or directions, are different enough from specifications to dismiss mixed identities of the form $f(\alpha) = \beta$. Caesar questions as (1) would involve a *category mistake*. Similarly, the following identities would be meaningless too:

(2) $d(l) = \text{England}$

(3) $r(n_1, m_2) = \text{the North Pole}$

²¹Here I follow (135)Studd’s exposition of the problem.

²²Remember that: (i) “abstract” does not mean here non-spatio-temporal entity, as required by Abstract; (ii) hitherto nothing forbids that abstracts are among the objects $\mathcal{L}_1$ speaks about even if these objects are concrete objects.

After all, what do directions or numbers have to do with roman emperors, countries or places? Generally, this response contends that all mixed sentences would be regarded as ill-formed and semantically flawed. In consequence, the extended language $\mathcal{L}_2$ would not be, against appearances, one-sorted. On the contrary, it must be divided into two non-overlapping sorts: one comprising $\mathcal{L}_1$, the other containing all the expressions in $\mathcal{L}_2$ committed to abstracts.

However, this answer is highly problematic. It does not do justice to speakers' practice.²³ Mixed identities and Caesar questions are more common and general than what examples (1)-(3) reveal. They are not limited to this kind of bizarre identities. On the contrary, they arise as soon as abstract terms are substituted in $\mathcal{L}_1$'s atomic expressions of the form $P(\alpha_1 \dots \alpha_n)$ by one or more α_i , $0 < i \leq n$. For example, advocates of this response must also give up questions such as:

(4) Is $\#(F)$ Frege's favourite number?

(5) Is $d(l_1)$ spatio-temporal?

This seems too high a price to pay. It is very common to ask for someone's favourite number. And, crucially, platonism must be able to claim that mathematical objects are not physical objects —i.e., platonism entails that Abstractness is true.

Furthermore, Caesar questions not only involve mixed sentences between abstracts and physical entities. They also include additional cases involving abstracts resulting from different abstraction principles. Consider a system containing several abstraction principles²⁴ —each of them containing a different unity relation $\sim_j$. Let $f_j(\alpha)$ denote the abstract specified by α relative to the unity relation $\sim_j$. Thus, f_j is an operator that to each α in the field of $\sim_j$ assigns a corresponding abstract, and the abstraction principles have the following form:

AP _{j}

$$\alpha \sim_j \alpha \wedge \beta \sim_j \beta \rightarrow (f_j(\alpha) = f_j(\beta) \leftrightarrow \alpha \sim_j \beta).$$

Then, the Caesar problem comes in the form of questions about mixed identities of the form $f_j(\alpha) = f_k(\beta)$ involving different kinds of abstracts. For instance, suppose that both HP and RTL are accepted. There seems to be no principled reason to reject the following as meaningless.

(6) $\#(x \neq x) = r(0, 1)$?

²³As Linnebo (82, pp. 162-163) notes, some authors dismiss this response adducing that the rejection of C2 is at odds with a realist ontology (40, p. 22). If there is an object a and another object b which exist independently of us, the identity statement " $a = b$ " will be meaningful and, therefore, true or false. Thus, C2 would be inconsistent with *Independence*. However, given the way in which the latter thesis has been formulated here, no inconsistency arises. If a and b are independent in the sense demanded by *Independence*, this only means that their existence is not metaphysically dependent on epistemic factors. Nothing is said about their identity. As Rayo (116, p. 80) points out, there is nothing in the idea of realism that forbids a many-sorted metaphysical or linguistic regimentation. See Linnebo (82, pp. 162-163) for similar comments. The latter takes Frege's metaphysics as a counterexample. Frege was a realist about objects and concepts. At the same time, he considered as a fundamental principle "never to lose sight of the distinction between concept and object" (46, p. x). These belonged to different categories, and identities between them lacked a definite truth value. In sum, the rejection of C2 does not pose a problem concerning the realist pretensions of the abstractionist project.

²⁴Here I follow Linnebo (82, p. 161)

That is, is the abstract 0 introduced by RTL the same as the cardinal 0 introduced by HP? Indeed, speakers' intuitions strongly support a positive answer: the cardinal number 0 is the rational number 0. Similarly, abstractionists want to claim that the natural number 0 is the cardinal $\#(x \neq x)$.

$$(7) \ 0 = \#(x \neq x)?$$

They want to reconstruct mathematics using abstraction principles. Thus, they would not succeed if the latter only introduced ersatz copies of numbers or other sort of abstract entities. For all these reasons, this first response is highly unpleasant.²⁵

5.4.1.2 Semantic Indeterminacy

A similar solution is to deny C1. Indeed, the objection that abstraction principles are not strong enough to secure reference has a long history. The main argument for this objection takes the form of a *permutation argument* (58; 1). Informally speaking, the argument claims that abstraction principles do not impose enough constraints to make abstract terms of the form $f(\alpha)$ refer to a unique element. For, the unique restriction they impose is that terms $f(\alpha)$ and $f(\beta)$ refer to the same object just in case $\alpha \sim \beta$. *No matter which object these terms end up referring to as soon as this condition is met.* Indeed, there are several objects which do the job. Therefore, abstraction principles are not enough to pick a unique referent for the abstraction terms.

Permutation arguments are model-theoretic arguments —where models are used to mathematically replicate different semantic notions. They put into question the ability of abstraction principles to single out a unique model. Suppose $\mathcal{M}$ is a model of the theory about $\sim$ formulated in the base language $\mathcal{L}_1$. According to permutation arguments if there is a model $\mathcal{M}_1$ of $\mathcal{L}_2$ extending $\mathcal{M}$ and satisfying AP, then there are infinitely many extensions $\mathcal{M}_k$ satisfying it.²⁶ Then, this mathematical result is taken as evidence against the metasemantic strength of abstraction principles. In the absence of additional constraints, they are not enough to secure reference.

Advocates of this second response reconstruct the semantics of abstract terms and, in general, the semantics of the extended language $\mathcal{L}_2$ as hosting some degree of semantic indeterminacy. Recently, this idea was elaborated by Boccuni and Woods (9), who developed a supervaluational semantics for abstraction. This approach models the semantic content of the new expressions in $\mathcal{L}_2$ against all the interpretations which make AP true. As I said, there are many such interpretations —i.e., models. Then, to secure

²⁵See (82, pp. 160-162) and (135, pp. 7-10) for similar considerations.

²⁶Let $\mathcal{L}$ be a first-order language. A model $\mathcal{M}$ for $\mathcal{L}$ is a pair $\langle M, v \rangle$, where M is a non-empty set and v is a function which assigns to each expression of $\mathcal{L}$ its extension in M . A permutation is a function $\rho : M \rightarrow M$ from M into M . Assume there is a model $\mathcal{M}$ such that

$$v(f(\alpha)) = a = v(f(\beta)) \text{ iff } \langle v(\alpha), v(\beta) \rangle \in v(\sim),$$

for all terms α, β . Then, permutation arguments show that there is a permutation defining a model $\mathcal{M}^*$, where $v'(f(\alpha)) = \rho(a) = v'(f(\beta))$, $v'(\sim) = v(\sim)$, $v'(\alpha) = v(\alpha)$, and $v'(\beta) = v(\beta)$ such that:

$$v'(f(\alpha)) = v'(f(\beta)) \text{ iff } \langle v(\alpha), v(\beta) \rangle \in v(\sim).$$

Indeed, permutation arguments show that there are infinitely many permutations making AP true.

reference, they stipulate that abstract terms refer to an object if and only if they refer to this object in all interpretations. However, as permutation arguments show, abstract terms refer to different objects in different interpretations. Consequently, they conclude that abstraction principles are not enough to secure reference —supervaluationist semantics does not assign to each abstract term a unique object as its referent. As a result, these are referentially indeterminate. Similarly, supervaluationist semantics settles that a sentence of $\mathcal{L}_2$ is true if and only if it is true in all interpretations, and false if it is false in all interpretations. Otherwise, its semantic content remains undetermined.

In spite of its simplicity and elegance, this solution is as unattractive as the former. It copes with the Caesar problem by endorsing the expressions of the extended language $\mathcal{L}_2$ with some degree of semantic indeterminacy. As a result, it also dismisses questions (1)-(7) as lacking a definite response. For example, in some interpretations the numeral “3” introduced by HP —assuming that HP introduces the natural numbers— will refer to Julius Caesar, and the answer to (1) will be affirmative. In others, this numeral will refer to an abstract object different from the Roman emperor, and the answer to (1) will be negative. As a result, (1) turns to be indeterminately true or false —i.e., it is neither true nor false. The same would apply to the other questions (2)-(7).

Nevertheless, as I argued, speakers’ practice reveals that many Caesar questions need a definite answer. Speakers have favourite numbers. Thus, if numbers are introduced by abstraction, it must make such claims involving propositional states true or false. Consequently, abstractionism must give definite answers to questions such as (4). Also, if abstractionism is used to explain reference to mathematical abstract objects —i.e. if abstraction principles are used for the benefit of platonism—, they must make Abstractness true. Platonism entails that mathematical objects are not spatio-temporal. Thus, abstractionism must give a definite answer to questions such as (5). Finally, abstractionism must determine the truth value of mixed identities involving systems of abstraction principles. These identities affect abstracts of different kinds. Therefore, if abstractionism wants to offer a serious theory of abstraction, it must explain how various abstraction principles behave together and how they relate to each other. In particular, if abstractionism wants to explain how mathematical objects —and not only ersatz copies of them— are given to speakers, it must answer questions of the form of (7) positively.

All this reveals that if the former response —the rejection of C2— to the Julius Caesar problem is in bad condition, then also is this second solution —the rejection of C1. To say that mixed identities involve categorical mistakes and are therefore meaningless, or to say —even if they are meaningful— that they lack a definite truth value, does not solve the problem. Instead, what is needed is a response that accepts Julius Caesar’s questions as meaningful —thus, that accepts C2—, and also preserves the ability to explain reference to particular abstracts —thus, that accepts C1. Only in this way will abstraction principles serve the purposes required of them.²⁷

5.4.1.3 The Wholesale Response

There is a unique option left: to deny C3. Abstractionism needs to explain how an abstraction step performed by using AP and Inheritance fixes the reference of abstract terms and delivers definite answers

²⁷See (135, p. 9) for similar considerations.

to Caesar questions. Thus, this response takes $\mathcal{L}_2$ to be a one-sorted language where mixed sentences involving abstracts and other kind of entities are perfectly meaningful. Responses denying C3 are classified into two major groups: *wholesale* and *piecemeal* responses.²⁸ Wholesale responses force abstraction principles to provide Caesar questions with determinate answers by appealing to a strong metaphysical regimentation of the world. This regimentation would impose enough constraints on abstraction for it to determine the content and the truth value of mixed statements. That is, according to the wholesale response, the semantic status of mixed statements would be a function of the process of abstraction and the world.

Wholesale The content and truth value of mixed statements is a function of AP, Inheritance, and metaphysics.

“According to what I will call *wholesale responses*, number-Roman identity conditions come bundled wholesale with number-number identity conditions.” (135, p. 2)

This is the solution championed by Hale and Wright in a series of well-known papers (150; 52).²⁹ The key notion of their proposal is that of a *sortal concept*. Informally speaking, a sortal concept is a concept whose predication requires the indefinite article “a”. For example, the concepts of *tree*, *mountain*, or *number*. These concepts allow for the introduction of the indefinite article, forming the expressions “a tree”, “a mountain”, or “a number” —compare with mass terms such as “water” or “knowledge”. While “this is a tree” is a grammatical expression, “this is a water” is not. The presence of the indefinite article in sentences involving sortals shows that these are concepts involving facts about identity —the identity of those objects satisfying them. More precisely, let’s say that the universal closure of a sentence of the form

$$F\alpha \wedge F\beta \rightarrow (\alpha = \beta \leftrightarrow \psi),$$

where ψ is a sentence expressing a suitable condition for the identity of the objects in the extension of F , is a *criterion of identity for F* if and only if it holds necessarily.³⁰ The criterion of identity must hold necessarily of F ’s because it rules the identity of elements in its extension, whatever this extension is. Then, sortals can be characterized as concepts having a criterion of identity:

Sortal Concept A concept F is a *sortal* if and only if there is at least one criterion of identity for F .

Hale and Wright’s wholesale response makes use of sortal concepts to explain how abstraction principles handle mixed identities.³¹ According to them, an abstraction principle of the form of AP fixes the

²⁸This is Studd’s (135) terminology.

²⁹Studd (135, p. 3, n. 3) mentions the recent work of Rosen and Yablo (122) as a competing wholesale approach.

³⁰The particular kind of modality is not relevant to the present discussion.

³¹The latest versions of their proposal appeal to an ontology of *categories*. A category is a maximally inclusive sortal.

A concept F is a category if and only if

- (i) F is a sortal for which there is a criterion of identity I , and
- (ii) for all β such that $\neg F\beta$ and any sortal G different from F , then I is not a criterion of identity for G .

reference of the abstract terms and, at the same time, captures a sortal concept: the sortal under which the abstracts fall. For instance, HP captures the sortal *cardinal number*. In turn, RTL captures the sortal *rational number*. But, sortals should be all which is required to deal with mixed and unmixed identities for the following two reasons. First, if an abstraction principle of the form of AP captures a sortal concept F , then the universal closure of the following sentence

$$F(\alpha) \wedge F(\beta) \rightarrow (f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta)$$

holds necessarily. Thus, AP is a criterion of identity for F . As a result, it rules the identity of the abstracts which fall under the corresponding sortal and settles the status of unmixed identity statements. In the case of HP, two cardinals $\#(P)$ and $\#(G)$ are the same if and only if the concepts P and G are equinumerous. In the case of DIR, two directions $d(l_1)$ and $d(l_2)$ are the same if and only if the line l_1 is parallel to the line l_2 . Generally speaking, two abstracts $f(\alpha)$ and $f(\beta)$ will be identical just in case α bears the unity relation $\sim$ to β .

Second, how can the notion of a sortal concept handle mixed identities—generally, mixed statements—between abstracts and other objects? According to the wholesale response advocated by Hale and Wright, such identities are false because of the mere fact that the abstracts and the other objects fall under different sortals.

Within a category, all distinctions between objects are accountable by reference to the criterion of identity distinctive of it, while across categories, objects are distinguished by just that—the fact that they belong to different categories. (53, p. 389)

Metaphorically speaking, sortals carve up³² the world into “regions.” Different sorts of objects belong to different “regions.” And, because of this, they must be different. The wholesale response answers the Caesar Problem by attributing to the world a fine-grained structure, and by placing objects at a fixed point in this structure. Thus, the number 3 belongs to the region carved up by the sortal *cardinal number*. On the contrary, Julius Caesar belongs to the region carved up by the sortal *human being*. These are non-overlapping regions, and the number 3 and Julius Caesar must be different objects. Similarly, all Caesar questions (1)-(7) are answered in a negative way.

5.4.1.4 Problems of Wholesaling

The wholesale response is not free of problems. First of all, it presupposes that sortals have a unique criterion of identity.³³ That is, the identity of objects falling under any given sortal concept is ruled by a unique criterion of identity for this concept.

However, nothing depends on the distinction between categories and plain sortals for my present purposes. So, to keep the text as simple as possible I will restrict my considerations to Hale and Wright’s sortal-based approach.

³²The origin of the “carving” metaphor is attributed to Frege (46).

³³In order to compare two criteria of identity, their identity conditions ψ_1 and ψ_2 must also be compared. Thus, a criterion of identity for relations is required at this point. I assume that there is such a criterion.

Uniqueness Thesis Let F be a sortal concept and I the criterion of identity for F . Then, any other criterion of identity I^* for F is identical with I .³⁴

This is the reason why mixed identities can be single-handled all at once in such a simple move. Given a criterion of identity I_1 , pick the collection S_1 of sortal concepts for which the former is a criterion of identity —note that the Uniqueness Thesis does not prevent two different sortals from sharing the same criterion. Now, take a different criterion of identity I_2 . The Uniqueness Thesis entails that the collection S_2 of sortal concepts to which this second criterion of identity applies does not intersect S_1 . Thus, the Uniqueness thesis guarantees that the sortals associated with a criterion of identity are different from the sortals associated with a different criterion. As a result, the metaphysical and semantic status of mixed identities can be explained as suggested above. The identity of objects falling under some sortal in S_1 is determined by the initial criterion of identity I_1 . In turn, the identity of objects falling under some sortal in S_2 is determined by the second criterion of identity I_2 . Finally, mixed identities between objects falling under sortals of the first type and sortals of the second resolve as false. Generally speaking, the identity of objects falling under sortals sharing the same identity criterion is ruled by this criterion. Conversely, the identity of objects falling under sortals with different criteria of identity is settled as false precisely because they belong to different sortals.

Despite the simplicity of this solution, the Uniqueness Thesis is highly controversial. It is a controversial assumption that sortals only have one criterion of identity. Consider what is known as the *Finite Hume's Principle*:³⁵

FHP

$$\#_F(X) = \#_F(Y) \leftrightarrow X \approx Y \vee (\neg \text{Finite}(X) \wedge \neg \text{Finite}(Y)).$$

Compare it with HP:

$$\#(X) = \#(Y) \leftrightarrow X \approx Y.$$

FHP assigns to second-order entities X and Y the same object if they are finite and equinumerous. But, if they are infinite, it will assign them the same object by default. Despite their similarity, HP and FHP are different abstraction principles —they involve different unity relations.

Now, consider the following Caesar question:

$$(8) \#(x \neq x) = \#_F(x \neq x)?$$

where $\#(x \neq x)$ is the cardinal associated by HP to the concept of being not self-identical, and $\#_F(x \neq x)$ is the abstract associated by FHP to the same concept. Let's denote by $n_\#$ the cardinals introduced by HP and by n_F the abstracts introduced by FHP. Then, (8) asks if the cardinal $0_\#$ is identical to the abstract 0_F . Intuition strongly supports an affirmative answer. Even though the unity relations of HP and FHP are different, they seem to agree on the objects assigned to finite second-order entities. On the contrary, the Uniqueness Thesis entails a negative answer: $0_\#$ and 0_F must be different from each other. HP captures one sortal concept —let's say the sortal *cardinal number*_#. In turn, FHP captures another sortal, the sortal

³⁴Hale and Wright (53), seem to believe that the uniqueness thesis follows from their characterization of sortal concepts. However, they also show some reservations about it at the end of the article.

³⁵This was proposed in (22) as a counterexample to the uniqueness assumption.

*cardinal number*_F. Hence, because HP and FHP involve different unity relations, the Uniqueness Thesis entails that *cardinal number*_# is a different sortal from *cardinal number*_F. Consequently, the wholesale response answers negatively to question (8): $\#(x \neq x) \neq \#_F(x \neq x)$.

Granted that both HP and FHP are acceptable at the same time —see the next section for more on this point— the wholesale response entails a highly counterintuitive result. Even worse, consider the Caesar question number (7): is the natural number 0 the same as the *cardinal*_# 0_# introduced by HP? Abstractionists need to answer affirmatively if they want to recover natural numbers using abstraction principles. However, because of the negative answer to (8), the *cardinal*_# 0_# is different from the *cardinal*_F 0_F. As a result, 0_N and 0_F are different objects too. The abstractionist ends up with two collections of numbers: the natural numbers properly, and a copy of them.

To fix this situation, one could accept that the concept of *natural number* is associated with several criteria of identity while accepting the Uniqueness Thesis too.³⁶ Among them are HP and FHP. Additional identity criteria agreeing with these on the natural numbers would be also included. The problem with this solution is that the Uniqueness Thesis immediately entails that the concept of *natural number* is no longer a sortal. However, sortals are the explanatory device used by the wholesale response to solve the Caesar problem. Consequently, if natural numbers are subsumed by no sortal —thus, if no abstraction principle captures a sortal that can be identified as the sortal of natural numbers—, then abstraction principles will not be strong enough to answer at least some Caesar questions —questions such as (7) and (8). Even worse, abstractionism could not explain how speakers grasp the concept of natural number —a basic mathematical concept.

This situation speaks against the Uniqueness Thesis. It is highly plausible that the wholesale response is defective when this thesis is assumed. Therefore, it must be abandoned. It is possible for a sortal concept to have several identity criteria. But, if this is so, the notion of a sortal concept turns out to be too weak to solve the Caesar problem. Now sortals may be subject to a variety of identity criteria agreeing on their identity conditions when the latter are restricted to the extension of the corresponding sortal. For example, this is the case of HP and FHP. The sortal *natural number* is subject both to HP and FHP —both abstraction principles provide natural numbers with identity criteria, even though they disagree on the identity of infinite cardinals.

But, take the case of RTL

$$r(n_1, m_1) = r(n_2, m_2) \leftrightarrow n_1 \cdot m_2 = n_2 \cdot m_1.$$

This principle says that the rational corresponding to the pair of integers $\langle n_1, m_1 \rangle$ is the same as the rational corresponding to the pair $\langle n_2, m_2 \rangle$ if and only if the product of n_1 and m_2 is the same as the product of m_1 and n_2 . Let's call the sortal grasped by this abstraction principle the sortal of *rational numbers*_r and denote these objects by n_r . Again, we can consider the question (6):

(6) Is $0_r = 0_{\#}$?

That is, is the rational_r 0_r the same object as 0_# and 0_F? RTL is very different from HP and FHP. It assigns an object to pairs of integers, not to second-order entities. Indeed, it presupposes that speakers already

³⁶This is hinted by Linnebo (82, pp. 166-167).

have some access to naturals —note that this does not invalidate it as a candidate criterion of identity for the sortal of natural numbers. Thus, (6) asks: does RTL agree with HP and FHP on natural numbers?

We cannot turn to the Uniqueness Thesis to give an answer. Even though the sortals rational number_r and natural number_r have different criteria of identity associated with them, in the absence of this thesis they can share several criteria of identity. Thus, what the wholesale response needs is a principled and general explanation of when two sortals—in this case the sortals natural number and rational number_r—share some criterion of identity as well as their extension. To date, no satisfactory account has been found by advocates of the wholesale answer. All attempts to introduce a general criterion for deciding when two abstraction principles introduce the same abstracts have met counterexamples. Given a system of abstraction principles, there is no uncontroversial solution for R in

$$f_j(\alpha) = f_k(\beta) \leftrightarrow R\alpha\beta.$$

Indeed, there are good reasons to think that any general account of the identity between abstracts which fall under different sortals will meet counterexamples —see Studd (135) for a detailed exposition of the situation.

Wholesale responses fail to solve the Caesar problem. They lack enough resources to explain how an abstraction step based on AP and Inheritance is enough to fix the truth value of mixed identities—and, in general, of mixed sentences. Therefore, they must be at least supplemented by additional explanatory resources. Moreover, wholesale responses face further difficulties. As it has been argued, abstractionism must make a choice concerning the reduction of natural numbers—as well as other mathematical objects—to those abstracts introduced by abstraction principles. For instance, this choice is what is at stake in question (7): is the natural number 0 the same as the cardinal_# 0_#? Furthermore, is it the same as the rational_r 0_r? The problem is that general explanations of the identity between different sorts of abstracts—explanations of the kind required by wholesale responses—make of this choice an insoluble dilemma. All the general solutions proposed entail that there are several, incompatible candidates to do this reduction. Consequently, wholesale responses indulge in a version of Benacerraf’s multiple reduction problem. Otherwise, these responses impose unreasonable constraints on mathematicians’ conceptual freedom—see (135) for a detailed analysis of all the problems faced by wholesale answers. An alternative response to the Caesar problem is required.

5.4.1.5 The Piecemeal Response

An alternative response to wholesale solutions which also denies C3 has been recently advocated by Linnebo (82, Ch. 9) and Studd (135).³⁷ This is called the *piecemeal response*. According to it, Caesar

³⁷Advances of the piecemeal response can be found in Frege’s *Grundlagen* §67 and *Grundgesetze*—where he stipulates that numbers are extensions. Also, Rayo’s account of *just-is* statements (116) is amenable to this kind of response. It is worth noting that Balaguer (2) already mentions that speakers’ concept of natural number, as well as speakers’ concept of set, involves more than mere structural properties codified in PA or ZF. For instance, speakers believe that the empty set $\emptyset$ is not a table. Similarly, speakers believe that numbers are not spatio-temporal entities. According to Balaguer, these attributions belong to the speakers’ *full conception* of natural numbers, or sets. The full conception of a given mathematical object nicely answers the kind of questions causing the Caesar problem. In turn, abstraction principles such as HP or Frege’s Law V were designed to recover, precisely, those mathematical

questions can be answered by laying down further stipulations —other than AP and Inheritance— when performing an abstraction attempt:

Wholesale responses grossly overestimate what is achieved just by laying down HP. On the face of it, however, there is a much easier way to solve the Caesar problem. According to my favoured response, the content of a mixed context such as ‘ $\#X = \text{Caesar}$ ’ is just as open to stipulation as the content of an unmixed context such as ‘ $\#X = \#Y$ ’. On this view, then, Caesar questions may be decided piecemeal, via further stipulations governing unmixed contexts laid down alongside HP. (135, p. 3)

The piecemeal response emphasizes the conceptual freedom of speakers and the importance of conceptual decisions when it comes to metasemantic explanations of reference.³⁸ According to this view, abstraction principles of the form of AP together with inheritance principles and metaphysics are not enough to settle the status of mixed sentences. They do not provide by themselves questions (1)-(8) with a definite answer. Something else is required to fix the reference of abstract terms and to determine when a given abstract is equal or different to other sorts of objects. This “something else” comes in the form of a conceptual decision.

Piecemeal The content and truth value of mixed statements is a function of AP, Inheritance, and conceptual decisions.

Speakers *tend to* accept that finite cardinals_# introduced by HP are the same as the corresponding natural numbers. Similarly, finite cardinals_F and finite rationals_r introduced by FHP and RTL respectively *are regarded* as cardinals_#, and *a fortiori* as natural numbers too. Conversely, speakers *tend to* deny that cardinals_# are Roman emperors. Likewise, directions are regarded as different from countries. In consequence, questions (1)-(3) and (5) are answered negatively. In turn, questions (6)-(8) receive an affirmative answer. The key feature of the piecemeal response is that opposite answers to each of these questions are consistent with AP, *Inheritance* and the metaphysical structure of the world. Conceptual decisions fix this limitation.

How can this be the case? When performing an abstraction attempt, a community of speakers has space to affect its result. Nothing prevents them from laying down additional stipulations constraining the content of mixed statements as well as the content of statements involving different sorts of abstract terms. Seen in this way, the piecemeal response is an extension of the top-down metasemantics that abstractionism already hosts. Abstractionism contends that speakers stipulate the content of unmixed sentences using both abstraction principles and inheritance principles. However, it turns out that these resources are not enough to settle Caesar questions and as other mixed statements —this is what the Caesar

theories —such as PA or ZF— encoding structural properties about these objects. This is the reason why they are not enough to settle the truth value of mixed sentences. Also, this is why Balaguer’s idea of a *full conception* is in tune with the piecemeal solution offered here.

³⁸The role of conceptual decisions in framing speakers’ linguistic devices was also emphasized throughout Rayo. (116) Rayo argues in favour of a platonistic philosophy of mathematics akin to the abstractionist view advocated here.

problem reveals. Then, why shouldn't they also stipulate the content of these sentences? The three-step metasemantic explanation of 5.2.0.1 is supplemented by adding *mixed postulates*³⁹ of the following form:

$$f_j(\alpha) = x \leftrightarrow \psi(\alpha x)$$

$$\varphi(x_1 \dots x_n) \leftrightarrow \gamma(\beta_1 \dots \beta_n),$$

where $x_i = f_k(\beta_i)$ or $x_i = \beta_i$, and $\psi(\alpha x)$ and $\gamma(\beta_1 \dots \beta_n)$ are suitable *criteria* expressible⁴⁰ in the base language of abstraction $\mathcal{L}_1$ which govern mixed sentences involving the abstracts $f_j(\alpha)$ and other objects x_i . The former postulate stipulates the content of mixed identities between the abstract $f_j(\alpha)$ and other objects x by means of the criterion $\psi(\alpha\beta)$. Similarly to the unity relation $\sim$ in AP, this criterion imposes additional constraints which compositionally constrain the reference of abstract terms. However, the criterion takes into account objects other than the abstracts. The result is a suitable solution to the Caesar problem: the status of identities between abstracts and other objects is settled in the same way as identities involving only abstracts. The second postulate does the same for mixed sentences $\varphi(x_1 \dots x_n)$ —where $x_i = f_k(\beta_i)$ or $x_i = \beta_i$ — using the criterion $\gamma(\beta_1 \dots \beta_n)$. In this way, the proposal generalizes to all forms of mixed sentences.

Unlike wholesale responses, the piecemeal solution offers different answers to the same Caesar questions depending on each particular abstraction attempt. The reason is that different abstraction attempts involve different stipulations, and the content of the resulting languages is also different. For example, suppose that an abstraction attempt based on HP is supplemented with a mixed postulate claiming that abstracts are different from spatiotemporal objects:

$$\#(X) \neq x \leftrightarrow x \text{ is a spatiotemporal object.}$$

Such an abstraction attempt ensures that Caesar questions of the form (1) will receive a negative answer. The number $3_\#$ is different from Julius Caesar. Thus, in this case, the extended language $\mathcal{L}_2$ obtained through abstraction introduces enough resources to speak about non-concrete objects such as cardinal numbers. In turn, suppose that another abstraction attempt based on the same abstraction principle HP is not supplemented by the previous unmixed postulate. HP alone is not enough to prevent the new abstract terms $\#(X)$ from referring to concrete objects. In particular, if no additional constraints are added, nothing prevents the abstraction attempt —if successful— from fixing Julius Caesar as the referent of the term “ $3_\#$.” In this situation, Caesar questions of the form (1) can receive positive answers —indeed, $3_\#$ will be a Roman emperor. Properly speaking, the extended languages obtained through abstraction as well as the corresponding Caesar questions would be the same only apparently. In the previous example, while the former speaks about non-spatiotemporal objects, there is no guarantee that the latter will be able to do so. As Frege puts it:

The definition of an object does not, as such, really assert anything about the object, but only lays down the meaning of a symbol. After this has been done, the definition transforms

³⁹Terminology is from (135, p. 17)

⁴⁰Recall that if metaontological minimalism is accepted, this requirement is lifted to the weaker requirement that $\psi(\alpha\beta)$ could be expressed in a suitable language $\mathcal{L}_1$. Speakers are not required to actually have this language.

itself into a judgement, which does assert about the objects; but now it no longer introduces the object, it is exactly on a level with other assertions made about it. (46, § 67)

Consequently, the piecemeal response can cope with the failure of the Uniqueness Thesis and other difficulties faced by general solutions offered to the Caesar problem along the lines of wholesale responses. Speakers have some leeway to affect the outcome of abstraction. When they are faced with a system of abstraction principles, as that formed by HP and FHP, there are no features on the world that decide if these introduce the same sortals or not —meaning and truth are not a function of the world. Instead, speakers have something to say about this too. To a large extent, their practices determine whether HP, FHP, and RTL capture the sortal of natural number, whether some do and others do not, or whether no one of them does. For example, the abstracts assigned by HP and FHP to finite concepts can be identified by adding the following mixed postulate to the abstraction system formed by them:

$$\#(X) = \#_F(Y) \leftrightarrow X \approx Y \wedge \textit{Finite}(X) \wedge \textit{Finite}(Y).$$

Both abstraction principles end up grasping the same sortal concerning these abstracts —presumably, the sortal of natural numbers. In turn, if no postulate of this kind is added, the result will be very different.

When we develop our linguistic practices, we have some degree of choice about whether or not to allow categories to overlap. To handle mixed identity statements, we often need *conceptual decisions*, not just factual discoveries. [...] My analysis emphasizes the importance of choice and the insufficiency of always looking to the world for an answer. When our ancestors first confronted Caesar-style questions, they had a choice which way to go; and this choice played a role in shaping the concepts that they thereby forged. Today we find ourselves in a different situation, since many choices are already implicit in the linguistic practices that we have inherited. [Italics in the original] (82, p. 160)

The piecemeal response avoids the difficulties faced by wholesale solutions by simply not committing to any general solution at all. Mixed statements are dealt with one by one and their status is settled on occasion of particular abstraction attempts. When speakers perform an abstraction attempt there is some degree of freedom which is not determined by abstraction principles or by the structure of the world. These are not enough to fix the content of those languages obtained through abstraction and to determine the truth value of mixed statements. Thus, metaphysics falls short in shaping the form of speakers' linguistic and cognitive resources. Their practices and conceptual choices fill the void left by the world.

5.4.2 Bad Company

Abstractionism faces the threat of paradox. This is a well-known story. As Frege's infamous Basic Law V shows, some abstraction principles are inconsistent. Similarly to HP, this "law" assigns to each concept X an object: its extension. The Basic Law V says that the extension of a concept X —in symbols $\hat{x}.Xx$ — is the same as the extension of another concept Y —in symbols $\hat{x}.Yx$ — if and only if X and Y are coextensive.⁴¹

⁴¹As before, throughout this section the formulas must be read as the corresponding universal generalization.

BLV

$$\hat{x}.Xx = \hat{x}.Yx \leftrightarrow \forall z(Xz \leftrightarrow Yz)$$

Russell proved that BLV entails a contradiction —this result is known as *Russell's paradox*. The contradiction follows as soon as the second-order variables X and Y are instantiated with the appropriate concept. Let's define a relation of membership between objects and extensions as follows:

y is a member of the extension u —in symbols “ $y \in u$ ”— $\equiv_{df}$ there is a second order entity X such that Xy and u is the extension of X .

Thus, the expression “ $y \in u$ ” abbreviates “ $\exists X(Xy \wedge u = \hat{x}.Xx)$ ”. Now, take the concept R whose characteristic property is⁴²

$$Ru \text{ if and only if } u \notin u.$$

R is the concept which subsumes all the extensions that are not members of themselves.⁴³ The contradiction follows when BLV is instantiated with R .

$$\hat{x}.Rx = \hat{x}.Rx \leftrightarrow \forall z(Rz \leftrightarrow Rz).$$

This equivalence guarantees that the extension of R , that is, the extension $r = \hat{x}.Rx$ to which all non-self-membered extensions belong, exist. From this and from the characteristic property of R it follows that r belongs to itself if and only if it does not:

$$r \in r \leftrightarrow r \notin r.$$

Consequently, a contradiction immediately obtains.

Russell's paradox shows that speakers run a serious risk when they perform an abstraction attempt. No abstraction attempt based on a contradiction will be apt to introduce a meaningful language.⁴⁴ As classical model theory shows, contradictions do not have models. In short, the possibility of extending their linguistic resources does not come free of dangers. Indeed, the risk is more delicate than it seems. For, there are abstraction principles which are very similar to other consistent abstraction principles, but which are inconsistent. Consider the “cousin” of HP (82, 54) that assigns a type $\dagger(X)$ to dyadic relations X :

H²P

$$\dagger(X) = \dagger(Y) \leftrightarrow X \cong Y,$$

where “ $\cong$ ” expresses a relation of isomorphism. Similarly to HP, H²P states that the type of a relation X — $\dagger(X)$ — is the same as the type of a relation — $\dagger(Y)$ — just in case X and Y are isomorphic — note that coextensionality is an isomorphism between monadic second-order entities. However, H²P is

⁴²Note that by definition of “ $\in$ ”: $u \notin u$ if and only if $\neg \exists R(Ru \wedge u = \hat{x}.Rx)$.

⁴³Second-order comprehension is required to prove the existence of R .

⁴⁴Let's leave aside paraconsistency here. The abstraction attempts I am interested in are those attempts which are apt to explain classical mathematics. Even if the idea of an abstraction attempt which makes a contradiction true is acceptable, surely this form of abstraction would not be able to explain classical mathematics.

inconsistent, as it entails the Burali-Forty paradox. This example shows that bad cases of abstraction can be very difficult to detect. Although similar to other principles which are in good order, they turn out to be inconsistent.

Besides, there are abstraction principles which are perfectly consistent, but which are inconsistent with other abstraction principles. This is the case of the parity principle (11) and the nuisance principle (152). Take the following equivalence relation: a concept X *differs finitely from another concept* Y if and only if there are finitely many objects that are X , but not Y , or the other way around. The nuisance principle assigns an object $n(X)$ to each concept X in virtue of this relation. Specifically, the principle says that the object $n(X)$ assigned to the concept X is the same as the object $n(Y)$ assigned to the concept Y if and only if X and Y differ finitely from each other:

NP

$$n(X) = n(Y) \leftrightarrow X \text{ differs finitely from } Y.$$

NP is satisfiable in and only in finite domains —hence, it is consistent. On the contrary, HP is satisfiable in and only in infinite domains —and is consistent too. Consequently, NP and HP cannot be satisfied in the same domain of objects. Despite each of these abstraction principles is consistent, they are mutually inconsistent. As a result, not all consistent abstraction principles can be accepted simultaneously.

Indeed, the piecemeal response could entail further inconsistencies. Some conceptual decisions regarding systems of abstraction principles can deliver a contradiction (135, pp. 22-23).⁴⁵ For example, suppose that speakers decide to identify the cardinals_# introduced by HP with the cardinals_F introduced by FHP by adding a mixed postulate which only involves the relation of equinumerosity:

F-H

$$\#(X) = \#_F(Y) \leftrightarrow X \approx Y.$$

This stipulation will lead them irrevocably to a contradiction. According to HP, $\#(\mathbb{N}) \neq \#(\mathbb{R})$. At the same time, $\#_F(\mathbb{N}) = \#_F(\mathbb{R})$. But, by F-H, $\#(\mathbb{N}) = \#_F(\mathbb{N})$. Therefore, $\#_F(\mathbb{N}) \neq \#(\mathbb{R})$: a contradiction. In this way, what has been designed as a solution, now leads us to another, equally serious challenge: the piecemeal solution hosts possible contradictions. Thus, in order to serve as an acceptable response to the Caesar problem, we must ensure that this drawback is brought under control.

The existence of this plethora of inconsistent or mutually inconsistent abstraction principles is known as the *Bad Company problem*. The problem is especially pressing for abstractionism. The “bad companions” —such as H²P or BLV— are structurally similar to the good ones. This is why it is so difficult to identify them. However, their identification is crucial to determine when a community of speakers is in a position to improve their linguistic resources and when it is not. Generally, what is required is a technical and philosophical understanding of what makes bad companions bad, and what differentiates them from the good cases. In turn, systems of mutually inconsistent abstraction principles need criteria to diagnose which of them are the good, and which of them are the bad companions —as well as to specify when speakers are allowed to safely identify different sorts of abstracts, securing the piecemeal response to the Caesar problem. Indeed, the literature on the Bad Company problem is extensive. Many formal

⁴⁵Similar worries were raised by Hale and Wright (52), as well as by (90).

criteria have been proposed to fix the situation:⁴⁶ *irenicity*, *stability*, *conservativeness*, or *unboundedness* are just a few of them. A careful review of these criteria would take me too far beyond the scope of the text. Instead, I will opt for a simpler diagnosis, adopting two quick complementary steps to cope with the problem.

5.4.2.1 Congruence in Systems of Abstraction Principles

First, the inconsistency underlying F-H is caused because the criterion of the mixed postulate $\neg X \approx Y$ is not respected by the unity relations of the abstraction principles involved.⁴⁷ That is, at least one of the unity relations of HP and FHP is not congruent with $\approx$. In this case, note that the latter is the unity relation of HP—trivially, it is a congruence with respect to itself. On the contrary, the unity relation of FHP is not a congruence with respect to $\approx$.

Let's denote the unity relation of FHP $\neg X \approx Y \vee (\neg \text{Finite}(X) \wedge \neg \text{Finite}(Y))$ —by means of “ $\approx_F$ ”. Then, the problem is that $\approx_F$ is not a congruence with respect to the relation $\approx$ governing mixed identities:

$$\text{it is not the case that for all } Z : X \approx_F Y \rightarrow (X \approx Z \leftrightarrow Y \approx Z).$$

For example, it is the case that $\mathbb{N} \approx_F \mathbb{R}$ —simply because both are infinite. However, while $\mathbb{N} \approx \mathbb{N}$, it does not happen that $\mathbb{R} \approx \mathbb{N}$ —ZF tells that there are more reals than naturals.

Given this diagnosis, the treatment suggests itself. Remember that when dealing with a single abstraction principle, the formulas φ that instantiate Inheritance

$$\varphi^*(f(\alpha_1) \dots f(\alpha_n)) \leftrightarrow \varphi(\alpha_1 \dots \alpha_n)$$

must satisfy Congruence—see section 5.2.1:

$$\alpha_1 \sim \beta_1 \wedge \dots \wedge \alpha_n \sim \beta_n \rightarrow (\varphi(\alpha_1 \dots \alpha_n) \leftrightarrow \varphi(\beta_1 \dots \beta_n)).$$

That is, the unity relation of abstraction principles must be congruent with φ . Now, a similar condition is required when dealing with systems of abstraction principles. As before, let's denote the unity relation of each abstraction principle by $\sim_j$. Also, let

$$f_i(\alpha) = f_k(\beta) \leftrightarrow \psi(\alpha\beta)$$

be a mixed postulate governing mixed identities between different kinds of abstracts. A *necessary condition* for identifying the abstracts introduced by two principles of the system—thus, a necessary condition for abstraction—is that their unity relations $\sim_j$ must be a congruence with respect to the criterion $\psi(\alpha\beta)$ featuring in the mixed postulate:

Congruence-2

$$\alpha_1 \sim_j \alpha_2 \wedge \beta_1 \sim_j \beta_2 \rightarrow (\psi(\alpha\beta) \leftrightarrow \psi(\alpha_2\beta_2)).$$

This condition prevents contradictions caused by the reasons underlying F-H. Furthermore, the solution can be extended to cover other formulas expressing properties and relations inherited by those abstracts that are rendered as identical by mixed postulates. This inheritance process will be governed by a new generalization of the inheritance principle which takes into account several abstraction principles at once.

⁴⁶See (143), (95, Ch. 4), (88), (80)

⁴⁷See (135) for a similar diagnosis.

5.4.2.2 Predicative Abstraction Principles

At this point, we have a partial solution to our problems. However, all this is not enough to avoid the Bad Company problem. The new principle Congruence-2 does not prevent the inconsistency of BLV or H²P. Moreover, when an abstraction attempt involves mutually inconsistent abstraction principles, this condition is silent on which of them are the good companions, and which ones are the bad. So, the response needs to be supplemented. Fortunately, there is a simple and elegant solution that solves all these difficulties in a single movement. It has recently been advocated by (82, Ch. 3 and 6).⁴⁸ Also, (116, pp. 81-82) proposal is similar in spirit. In any case, the solution is not new. It consists in adopting a *predicative* reading of quantification.⁴⁹

Russell's paradox shows that the collection of *all* non-self-membered sets does not form a set *r* —the Russellian set. As soon as the quantifiers in BLV —more specifically, the first order quantifier in the description of *R*'s characteristic property: “ $\forall z(Rz \leftrightarrow z \notin z)$ ”— are eliminated with the set *r*, a paradoxical reasoning follows deriving a contradiction. Moreover, it is usually thought that logic alone suffices for this reasoning —BLV is regarded as a logical falsity. However, this is not the case if a generous conception of logic is adopted. The contradiction arises only if speakers are fortunate enough to quantify over all the sets. Only then could they eliminate the first-order quantifier using *r*. But, here is where the problem lies. Logic alone is not sufficient to discard that speakers do not have such abilities.⁵⁰ Russell's paradox rests on the “pre-logical” assumption that quantifiers range over all the objects, including *r*.⁵¹ Once this assumption is resisted, there is plenty of space to block the paradox. Indeed, this was Poincaré and Russell's diagnosis. According to them the set-theoretic and the logical paradoxes were caused by some form of *vicious circularity*. The anomaly arises when some entity is defined or characterized in such a way that the definition or the expressions involved in its characterization presuppose the very entity. In some cases, this unacceptable circularity is explicit —for example, the liar paradox arises from a sentence that claims its own falsity.⁵² In others the circularity remains implicit. This is what happens with *r*. The characteristic property of this set only mentions those sets which do not belong to themselves. No reference is overtly made to *r*. But, the circularity results from the assumptions that the quantifiers range over all sets, and thus over the set *r*. The notion of *impredicativism* was intended to capture these forms of implicit circularity.

A definition is *impredicative* when it quantifies over a collection of objects among which the defined entity is found. Otherwise, the definition is *predicative*. Thus, the inconsistency surrounding BLV can

⁴⁸(135; 136) has also advocated Linnebo's solution.

⁴⁹Predicativism arises from Poincaré and Russell's early diagnosis of the set-theoretic paradoxes. It led the latter to propose type theory as an alternative solution to the one proposed by Zermelo. See (33) for a brief and self-contained overview of the history of predicativism.

⁵⁰Suppose an inaccessible cardinal κ exists, but speakers have only fixed reference to sets below it. Then Russell's paradox is blocked at the level V_κ . In this case, *r* exists and belongs to V_κ . All speakers can say meaningfully is that *r* is not among the sets they speak about.

⁵¹See (36) for similar comments.

⁵²Another example is Routley set. This is the set *s* whose members are precisely those sets which do not belong to *s*. In symbols:

$$\forall x(x \in s \leftrightarrow x \notin s).$$

Clearly, this set only makes sense in paraconsistent logic.

be blocked not just by abandoning the law, but by going predicative. Once the characteristic property of the Russellian set r does not quantify over it, no contradiction arises. This suggests a general solution to the Bad Company problem. Instead of imposing additional criteria to the unity relation, just accept *predicative abstraction principles*.

Impredicative Abstraction An abstraction principle AP is *impredicative* if and only if the abstract terms in its left side — $f(\alpha)$ and $f(\beta)$ — refer to objects that fall inside the range of some quantifier that occurs in its right side — $\alpha \sim \beta$.

Predicative Abstraction An abstraction principle AP is *predicative* if and only if it is not impredicative.

A *sufficient condition* for performing a successful abstraction step by using a system of abstraction principles is for these principles to be predicative. This is the predicative solution to the Bad Company problem.

The dynamic approach to abstraction gives a radically different answer to the bad company problem: *it accepts abstraction on just about any equivalence relation*. Paradox is avoided, not by being selective about which abstraction principles we accept but *by allowing the abstracta to lie outside of the domain on which we abstract*. (82, p. 52)

As a result of the predicative solution, abstraction becomes *dynamic*.⁵³ Against the *static* nature of more classical, impredicative abstraction, predicative abstraction principles allow speakers to expand their original domain of quantification. Before the community of speakers undertakes an abstraction attempt, the quantifiers range over a domain of objects. Then, an abstraction attempt performed by these principles assigns the abstract terms reference to “new” objects that lie outside this “old” domain. For example, in the case of set theory, BLV fixes reference to the Russellian set r —the set of all sets *in the old domain* that do not belong to themselves. However, r falls outside the range of quantifiers and does not belong to the domain where they are interpreted. In this sense, it is a new object for speakers. Abstraction provides them with an enlarged stock of objects. Then, quantifiers can be interpreted in the new domain formed by adjoining these new objects to the old ones. Then, the process can be *iterated* again, gaining access to more objects —I will back to this topic in Chapter 6.

5.5 THE METASEMANTIC EXPLANATION OF MATHEMATICAL KNOWLEDGE

In the last section, I explained how the two main challenges of abstractionism, namely, the Julius Caesar problem and the Bad Company problem, can be overcome. If the piecemeal and the predicative solution are in good condition, then abstraction principles can be used to solve the Language Acquisition Challenge (LAC) —see section 5.2.1. The reductionist perspective introduced on metasemantics by abstractionism allows us to explain how mathematical vocabulary acquires its meaning and how speakers acquire their linguistic competence without attributing to them knowledge of the extended language. The explanation

⁵³See (103), (82, Ch. 3), and (136).

does not presuppose logical knowledge about entailment and consistency in the latter. Consequently, we avoid the difficulties affecting fictionalism and multiversism.

Furthermore, abstractionist metasemantics offers a completely new perspective on mathematical knowledge. According to it, the linguistic competence acquired through abstraction principles already equips speakers with mathematical knowledge. This involves two steps. First, language extensions performed by successful abstraction attempts introduce a new language which comes with a collection of truths about the abstracts. Second, language understanding guarantees knowledge of some of these truths. When speakers learn how to use the expressions of the new language they obtain additional non-linguistic information: by acquiring knowledge of the truth conditions associated with sentences they already acquire some knowledge of what is true. In this way, mathematical knowledge is no longer reduced to logical knowledge. Instead, mathematical knowledge reduces to language understanding:

Mathematical knowledge reduces to language understanding.

This metasemantic explanation of mathematical knowledge has been advocated by Linnebo (79), (82, Ch. 11) and Rayo (116, Ch. 4). It heavily relies on Stalnaker's (131).⁵⁴ In this section I will develop the explanation and show how it solves both Benacerraf's challenge and the reinforced version posed by mathematical reliability.

Remember what the epistemological challenge asks for. First, it calls to accept mathematical platonism, the thesis resulting from the conjunction of Independence, Abstractness and Existence.

- (1) Assume mathematical platonism: there are abstract objects which are independent of speakers.

Second, on the face of this assumption, the challenger asks for an explanation of mathematical knowledge—this is Benacerraf's challenge. Indeed, a reinforced version of the challenge posed by Field (40, pp. 25-30, 230-239) and Linnebo (79) asks for an explanation of the reliability of mathematicians. More precisely, this reinforced challenge asks for an explanation of the *reliability claims*.

- (2) There is an explanatory pressure that comes from conditionals of the form:

Reliability Claim

$\forall \phi (\text{Mathematicians accept } \phi \rightarrow \phi \text{ is true}),$

⁵⁴In (131), Stalnaker tries to rescue possible-world semantics from the paradoxes of knowledge, the paradoxes of the strict conditional and other failures. He offers a metasemantic account of belief which is able to discriminate between expressions that necessarily have the same extension. For instance, it is necessary that Hesperus is Phosphorus—both are the same object, and identity is necessary. However, some person who does not know much about astronomy could believe that Phosphorus is the morning star, at the same time she believes that Hesperus is not. Therefore, *belief* contexts do not allow for substitution *salva veritate* of expressions which necessarily have the same extension. The phenomenon is problematic, because if “Hesperus is the morning star” and “Phosphorus is the morning star” express the same proposition, and it is believed that Phosphorus is the morning star, what else is required to believe that so is Hesperus? Currently, beliefs are regarded as hyperintensional phenomena. On the contrary, Stalnaker explained failures of substitution as metasemantic failures. When some person believes that Phosphorus is the morning star, but—incorrectly—believes that Hesperus is not, what happens is that she fails to understand what propositions are expressed by what sentences. In particular, she does not know which proposition is expressed by “Hesperus is the morning star”, and thus she does not realize that this is the same proposition as the one expressed by “Phosphorus is the morning star.” This metasemantic component of Stalnaker's explanation is crucial for Linnebo and Rayo's explanation of mathematical knowledge.

where φ is a mathematical sentence, “ $\rightarrow$ ” at least expresses a material conditional, and “to accept φ ” means to be disposed to assert φ .

As both Field and Linnebo note, the challenge is stronger than the original proposed by Benacerraf because it does not mention *knowledge* at all. The reliability claim is a consequence of the reliability of mathematicians’ beliefs. For, if their belief in what φ says is reliable, at least it must be factive —thus, if speakers reliably believe in what φ says, φ is true. Besides, note that if speakers are competent in the language to which φ belongs, they believe in what φ says if only if they accept φ . In consequence, the challenge targets the justification of those beliefs with platonistic content. If platonists can not explain the reliability claim, they will not explain the reliability of mathematical beliefs —and, *a fortiori* they will not be able to explain mathematical knowledge.

Third, the status of platonism depends on how it stands in relation to this explanatory pressure. If platonists fail to explain the reliability claims, they will not be allowed to assert (1).⁵⁵ That is, if platonists fail to explain how mathematicians’ beliefs with platonic content are reliable, then mathematicians will not be entitled to believe in such contents.

- (3) If platonism satisfies (2) —that is, if it provides an explanation of reliability claims—, then (1) is safe. If not, speakers are not justified in believing in (1).

Note that in the worst scenario, the argument does not conclude that mathematical platonism is false. It claims that speakers are not justified in believing in platonism —this is enough to abandon (1). However, the result is devastating for platonism as a philosophy of mathematics. For, if speakers are entitled to believe in the content expressed by true mathematical sentences, but not in platonism, it follows that the content expressed by mathematical sentences is not platonic. Consequently, if this were the case an alternative explanation of mathematics would be required. However, as it was shown in Chapters 3 and 4, the main alternatives do not fare any better.

5.5.1 Easy Options

Let’s briefly remember that there are two unsatisfactory ways out of the challenge.⁵⁶ Both proceed, positively or negatively, too quickly. The first consists in denying that platonism can satisfy (2) on the grounds of a causal theory of knowledge —or a causal theory of belief justification.⁵⁷ (7; 40) More generally, the possibility of providing reliability claims with a platonistic explanation can be resisted by arguing that mathematical justified beliefs with platonic content require some counterfactual relation between their content and speakers —as empirically justified beliefs do. This is the same as to claim that the reliability claims follow from the following instance of Responsiveness:⁵⁸

$C(\varphi, S) \wedge \varphi \Vdash$ speakers S are justified in believing that φ ,

⁵⁵This follows from the K-Rule: assertion of φ entails knowledge of φ . Moreover, knowledge of φ entails justified belief in φ . The conclusion follows by two applications of *Modus Tollens*.

⁵⁶See Chapter 2, 2.3.2, 2.6.

⁵⁷In this section I will speak in terms of mathematicians’ beliefs, instead of mathematicians’ acceptance of sentences. However, note that everything I will say can be put in those terms.

⁵⁸Here, responsiveness is modified to apply to justified belief, instead knowledge. Consequently, the unknown X can be removed. Reason as follows. We want to explain the reliability claim by adding that mathematicians’ belief

where φ is a mathematical sentence, $C(\varphi, S)$ expresses some counterfactual correlation between those facts expressed by φ and speakers' epistemic responses, and $\models$ means that $C(\varphi, S)$ and the fact that φ together explain what to be justified in believing that φ consists of. In this case, note that, if the reliability claim follows from Responsiveness, the conditional " $\rightarrow$ " must be read as a *counterfactual conditional*. This immediately rules out a successful account of (2) for two reasons. First, according to platonism mathematical objects are abstract—they are not spatio-temporal entities. Consequently, the counterfactual relation expressed by $C(\varphi, S)$ cannot be a causal relation, or any other spatio-temporal relation. However, aside of these, it is extremely difficult to think what such a relation would look like. It must be a counterfactual relation between non-spatio-temporal objects and spatio-temporal ones—the speakers. To date, nobody has found it. And, there is a consensus that mathematical knowledge cannot be accounted for by using this kind of counterfactual relations.

Second, if mathematical truths are necessary, it is not possible to make sense of a counterfactual relation between mathematical facts and speakers. Any counterfactual relation that does the job would entail that it makes sense to vary the mathematical facts. For, the relation would be of the form:

$C(\varphi, S) \equiv$ If mathematical facts φ were different, so would be the epistemic responses of speakers S .

Mathematical facts feature in the antecedent of the counterfactual because they are more fundamental than speakers' epistemic responses. Indeed, this is what Responsiveness aims to capture concerning speakers' beliefs or knowledge. Nevertheless, if mathematical facts are necessary, they cannot vary. As a result, we can not make sense of $C(\varphi, S)$. If platonic mathematical facts are necessary, any attempt to explain mathematical beliefs—or knowledge—as involving a counterfactual relation between these facts and speakers will fail. In short, if mathematical beliefs with platonistic content satisfy the corresponding instance of Responsiveness, platonism fails to satisfy (2). Nevertheless, there is no reason for endorsing platonism with the burden of explaining Responsiveness. Benacerraf's holistic preferences are not well justified. Mathematics is a sufficiently different enterprise from the empirical sciences to require a similar epistemology. Therefore, any denial of platonism which relies on the assumption that explanations of mathematicians' beliefs must conform to the standards of empirical statements or beliefs will be biased from the start. Once this is accepted, there is room for a satisfactory explanation. Indeed, this is accepted by fictionalism and by multiversism too: mathematical knowledge does not conform to Responsiveness.

The second attitude towards the challenge turns the second of the above weaknesses into a strength for platonism. Specifically, it holds that reliability claims do not pose any explanatory pressure on platonism provided that mathematical truths are necessary. This solution simply denies (2) on the grounds of mathematical necessity. As I said, if true mathematical sentences are necessary, then the conditional $\rightarrow$ in the reliability claim cannot be read as a counterfactual conditional. As a result, there is no pressure

in φ is a *justified* belief. So, suppose that mathematicians justifiably believe in φ . Because of Responsiveness, this is explained by the conjunction of $C(\varphi, S)$ and φ . Thus, it must be the case that φ . Indeed, using the counterfactual conditional involved in $C(\varphi, S)$

 $C\varphi \equiv$ If mathematical facts φ were different, so would be speakers' epistemic responses,

the conditional $\rightarrow$ in the reliability claim is strengthened to a counterfactual conditional.

to explain any correlation between platonistic mathematical facts and speakers other than the correlation expressed by a material conditional. Besides, if mathematical facts are indeed necessary, it takes nothing in the world for them to be the case. Therefore, it does not make sense to explain mathematical knowledge by requiring speakers to check the world to see if indeed there are numbers, sets, or functions—there is no chance for failure. Consequently, it seems that all that is required to explain mathematical knowledge is to attribute speakers a correct understanding of mathematical language. Suppose that ϕ is necessary. As soon as competent speakers understand ϕ 's truth conditions, they will know that it does not require anything from the world to be true. Thus, they will know that ϕ is indeed true. From this, it can be concluded⁵⁹ that reliability claims have no interesting explanation. Once mathematical language is acquired, justified belief in—or knowledge of—platonic facts does not impose demands on the world. It is trivial as soon as the content of mathematical language is grasped. Advocates of this second solution conclude from this that (2) is flawed. Platonism is not required to explain how mathematicians' beliefs are reliable, and the challenger must withdraw the challenge.

However, this response to the challenge goes too quickly. It reduces to say that mathematical platonism does not need to explain mathematical knowledge, or that this explanation is too cheap, just in case mathematical truths are necessary. This contrasts with the situation of mathematical students and ordinary speakers. Whether mathematics is necessary or not, as many people know, learning mathematics involves a great effort. Also, competent mathematicians spend a lot of time proving new, complex theorems. How can this be the case if to know—or to justifiably believe—that ϕ is true does require nothing once ϕ is correctly understood? The answer lies precisely in this qualification: *as soon as competent speakers understand ϕ 's truth conditions*, they will know that ϕ is true. The point is that, sometimes, it is extremely hard to understand what a sentence says. As Rayo puts it: “It is important to keep in mind that getting clear about the truth-conditions of a given mathematical sentence can be highly non-trivial. So determining whether the sentence is true is not, in general, a trivial affair.” (116, p. 98) It is one thing to deny that mathematical knowledge imposes substantive constraints on the world, and quite another to say that platonism is not obliged to explain mathematical knowledge—or at least, the reliability of mathematicians' beliefs. If this were the case, logical knowledge would not have to be explained either. Nevertheless, this is far from being true—the set of logical truths codified in first-order logic is not even decidable. In sum, mathematical knowledge is in need of explanation. And, the explanation has nothing to do with mathematical facts, but with the way speakers deal with language.

5.5.2 Mathematical Knowledge Through Abstraction

There is a perfectly good explanation of why mathematicians are so reliable—when it comes to mathematics. An explanation of what it takes for mathematician's beliefs—as well as mathematicians' acceptance of mathematical sentences—to be responsive to their truth. This is enough to account for (2), dispensing with the explanatory pressure posed by the challenger. Indeed, the explanation establishes a counterfactual connection between the mathematicians' beliefs and the truth of those very beliefs—they

⁵⁹For instance, Lewis (78, pp. 111-112) draws similar conclusions.

are not responsive to their truth just by chance.⁶⁰

5.5.2.1 The Two Dimensions of Empirical Truths

First of all, notice that when speakers are confronted with empirical questions, they need information about two different things to answer them. For example, suppose somebody is asked “Is El Capitan 914m high?”. To answer this question, first, she needs to know some facts about El Capitan—a rock formation in Yosemite National Park. Specifically, she needs non-linguistic information about its height.⁶¹ Second, she needs to understand the question. That is, she needs linguistic information about what object the name “El Capitan” refers to, as well as to know that “to be 914m high” is a predicate which is used in a very specific way. This piece of linguistic information is not part of the information about the height of El Capitan—the fact that El Capitan is 914m high does not entail that it is called “El Capitan.” Therefore, if speakers lack one of these two pieces of information, they will be unable to answer the question. For example, if they do not know how tall El Capitan is, but still manage to understand what are they being asked, they will answer something along the lines: “I have no idea.” In turn, if they do not understand the question, they will reply by asking for the linguistic information they lack: “Could you repeat, please? I did not understand you.” Only when speakers understand what words are uttered *and* are informed about the geology of Yosemite, they will be able to answer: “Yes. It is quite big, isn’t it?”

The success of speakers’ statements⁶² in picking a true sentence involves two explanatory levels [equivalently, the truth of speakers’ statements. At the *semantic level*, it depends on the content of the stated sentences. Speakers manage to state a *true* sentence ϕ partially because—to put it picturesquely—the world cooperates to make the case what ϕ says. In the previous example, when a speaker states “El Capitan is 914m high,” she makes a true statement in part because El Capitan is indeed 914m high. In turn, at the *metasemantic* level, their success depends on which is the stated sentence. Speakers manage to state a *true* sentence ϕ partially because ϕ is the sentence being stated. Thus, when a speaker states “El Capitan is 914m high,” she states a true sentence partially because this is about the height of El Capitan. Had she stated another sentence, her assertion could have been false. For instance, if her statement—resulting in the emission of the same chain of sounds—picked the sentence “The Everest is 914m high”, she would no longer be stating a true sentence. To sum up, speakers make true statements in virtue of two different kinds of facts: one kind contributes to its success in a *direct* way, while the other’s contribution is *indirect*.

⁶⁰Linnebo (79) notes that there is another unsuccessful explanation. He names it “the boring explanation”. According to this, a mathematician’s belief in the content of some theory T is responsive to their truth because: (i) she acquired these beliefs when she learned T and (ii) T is true. This explanation is as valid as it is obvious. The problem with the boring explanation is that it does not explain the relation between the fact that the mathematician learned T—and not other theory T’—and the fact that T is true. That is, when the theory is asked “why did the mathematician learn T?”, the theory does not answer “*because* T is true.” Consequently, at best it is a partial explanation of mathematicians’ reliability.

⁶¹Note that this piece of information is not necessarily coached in language.

⁶²Statements, as beliefs or suppositions are individual, localized actions. They are something that people do. The objects affected by statements are sentences—I will not specify if these are *type* or *token* sentences. Judgements do the same with propositions.

Direct Contribution The world —what sentences are about— directly contributes to speakers' success in making a true statement.

Indirect Contribution The metasemantic fact that speakers state a particular sentence indirectly contributes to speakers' success in making a true statement.⁶³

All this is in line with Benacerraf's defence of a causal theory of knowledge —or, at least a causal theory of justified belief. When asked about empirical matters, speakers need to gather two different pieces of information to provide a justified answer —and to possess justified beliefs about what they are asked. First, they need linguistic information about which sentence is involved in the question. That is, they need to understand what are they being asked.

Speakers need information about those facts that *indirectly* contribute to the truth of what is being asked.

Second, they need non-linguist information about the world. Notice that if speakers lack this second piece of information they could still have some beliefs about the issue in question. For some unexpected reason, they could believe that El Capitan is the tallest mountain in the world. Even better, they could believe without apparent reason that El Capitan is exactly 914m —being right by chance. However, the point is that if speakers lack information about those facts that directly contribute to the truth of what they are asked, their beliefs and responses will not be justified.

Speakers need information about those facts that *directly* contribute to the truth of what is being asked.

⁶³Linnebo's explanation involves a *semantic* truth predicate. (79, p. 567), (82, p. 203) The reason is that his explanation involves *sentences*, not *statements of sentences*. He argues that the truth of a sentence ϕ depends on two different kinds of fact. Directly, it depends on the world conforming to ϕ —what ϕ says must be the case. Indirectly, it depends on the metasemantic fact that ϕ “has the truth-condition it happens to have.” (79, p. 569) For example, if the sentence “The Everest is 914m high” had a different truth condition —one which attached a different meaning to the sentence's components—, the sentence could have been true. If the truth predicate were a deflationary one, this chain of reasoning would be simply false. For, according to the minimalist theory of truth, ϕ and “ ϕ is true” are synonymous. Consequently, the truth of a sentence would depend only on one sort of fact: the fact that ϕ is the case.

As he notes, advocates of minimalism could object to his explanation on this basis. However, this is not a problem for my purposes. I want to keep controversial assumptions to a minimum. But, if necessary I could make room for a semantic truth predicate. The real problem is that his explanation requires making sense of a sentence varying its truth-condition *without varying its identity*. However, the arguments in Chapters 3 and 4 strongly rely on opposite assumptions. As I said there, I take the identity of linguistic entities to be bounded to semantic factors. If this were denied, the arguments would no longer work. Moreover, we would be left with a purely grammatical conception of language. Then this conception would allow for a grammatical notion of consistency — which is fatal for the present project. Instead, I will expose the explanation in terms of statements. It is completely safe to think about speakers varying the statements they make. Furthermore, statements do not introduce a notion of consistency. Statements are actions, but the notion of consistency involved in Chapters 3 and 4 is not a pragmatic one. Also, statements presuppose a language: successful statements state meaningful sentences. Finally, this movement allows us to recover a minimal truth predicate. As a result, the explanation is improved against deflationists' complaints. Notice that the requirement of explaining the truth of a sentence ϕ by appealing to those facts depicted by ϕ does not contradict minimalism. As Horwich (63, pp. 23-24) notes, minimalism is compatible with additional explanations of truth. All it demands is to not build these explanations inside its definition —additional explanations of the concept will not be analytically true.

The best explanation of empirically justified beliefs will tell some story about how speakers become counterfactually entangled with the facts they are asked about. Generally, such a theory delivers the following instance of Responsiveness:

$$C(\varphi, S) \wedge \varphi \Vdash \text{speakers } S \text{ are justified in believing that } \varphi,$$

where φ is an empirical sentence. Consequently, if the first piece of information is given to them—if they understand what are they being asked—but they lack the second piece, they will be unable to honestly answer the question with a “yes” or a “no.” Finally, the explanation of empirically justified beliefs entails an explanation of the reliability claim. Reliable speakers accept a sentence φ just in case they are justified in believing that φ .

The moral is, when it comes to empirical sentences the explanation of the reliability claim requires much more than the speakers’ linguistic competence. It demands that two pieces of information be given to them:

Speakers S understand φ ,

and

$C(\varphi, S)$.

5.5.2.2 Reducing Mathematical Knowledge to Linguistic Competence

Mathematical discourse is different in this respect. Indeed, the lack of this two-step account of cognitive accomplishment is what makes it so distinctive. Consider an abstraction attempt performed by a system of abstraction principles of the form of AP and inheritance principles of the form of Inheritance. When the attempt is successful, it allows the community of speakers to extend the language $\mathcal{L}_1$, obtaining a richer language $\mathcal{L}_2$. Thus, the community obtains a new collection of meaningful sentences whose assertibility conditions are ruled by the right side of AP and Inheritance. Identities of the form $f(\alpha) = f(\beta)$ are assertible if and only if $\alpha \sim \beta$ is the case. Similarly, atomic sentences of the form $\varphi^*(f(\alpha_1), \dots, f(\alpha_n))$ are assertible if and only if the corresponding condition $\varphi(\alpha_1 \dots \alpha_n)$ is the case. Besides, these assertibility conditions can be extended to complex formulas. In general, the assertibility condition corresponding to any sentence φ_2 formed by composition from these atomic sentences is coached by a sentence φ_1 in the base language $\mathcal{L}_1$ —see 5.6.

Now, the question is: which kind of facts contribute to the truth of speakers’ statements involving $\mathcal{L}_2$ -sentences? As in the case of empirical sentences, there are some *metasemantic facts* that contribute *indirectly* to their success. Metasemantic facts explain how statements involving $\mathcal{L}_2$ -sentences pick these particular sentences—and not others. In this case, the relevant metasemantic facts are those facts concerning a particular abstraction attempt articulated by AP and Inheritance principles. How are speakers involved in the particular language $\mathcal{L}_2$ —and not in other? Because the community performs a successful abstraction attempt. For instance, take the case of DIR—the abstraction principle introducing directions. Speakers grasp the content of sentences committed to directions by relating their assertibility conditions

with facts about parallelism. For example, they grasp a notion of orthogonality between directions because this notion is precisely correlated through abstraction —specifically, through Inheritance— with facts involving the orthogonality of parallel lines. Thus, speakers understand a sentence such as “ $d(l_1)$ is orthogonal to $d(l_2)$ ” because its assertibility condition is the fact that the line l_1 is orthogonal to the line l_2 :

$$d(l_1) \perp^* d(l_2) \leftrightarrow l_1 \perp l_2.$$

Speakers use the new expressions in accordance with the principles ruling the abstraction attempt —AP and Inheritance. In general, speakers manage to state $\mathcal{L}_2$ -sentences just in case they follow the assertibility conditions governing the use of expressions of the extended language. These are the metasemantic facts contributing indirectly to the truth of speakers’ statements —they determine which is the language acquired by speakers.

In turn, which are the non-metasemantic facts that *directly* contribute to the truth of statements of $\mathcal{L}_2$ -sentences? Astonishingly, in the case of languages obtained through abstraction, these simply disappear. The metasemantic facts exhaust the whole story. An abstraction attempt not only supplies the new $\mathcal{L}_2$ -sentences with assertibility conditions, but also with truth conditions. Thus, the reductionist flavour of abstractionist metasemantics ensures that each $\mathcal{L}_2$ -sentence has a *non-homophonic* truth condition fully expressible in the base language of abstraction $\mathcal{L}_1$ —i.e., a truth condition grammatically different from the sentence in question. In particular, the truth condition of identities between abstracts corresponds to those equivalences generated by AP:

$$“f(\alpha) = f(\beta)” \text{ is true iff } \alpha \sim \beta,$$

And the truth condition of atomic sentences about abstracts corresponds to those equivalences generated by the Inheritance schema:

$$“\varphi^*(f(\alpha_1)...f(\alpha_n))” \text{ is true iff } \varphi(\alpha_1...\alpha_n),$$

where φ^* is a new predicate introduced from φ . These equivalences can be generalized to complex sentences —see 5.6. Thus, for each $\mathcal{L}_2$ -sentence φ_2 there is a $\mathcal{L}_1$ -sentence φ_1 such that

$$\varphi_2 \text{ is true iff } \varphi_1.$$

Notoriously, this situation contrasts with the case of empirical sentences. These sentences play an ineliminable role in stating their truth condition. Unlike the case of abstracts —among them, mathematical objects—, the semantics of empirical sentences is highly non-reductionist. Remember the above example. The facts depicted by a sentence such as “El Capitan is 914m high” are indispensable to describe its truth condition:

$$“\text{El Capitan is 914m high}” \text{ is true iff El Capitan is 914m high.}$$

In general, the truth condition of empirical sentences is stated in terms of those very sentences —this accords with the disquotational semantics introduced by the T-schema. Empirical facts involving mountains, trees, books, and other similar objects directly contribute to the truth of those sentences committed

to them. This is the reason why speakers must be counterfactually related to them —i.e., they must be C-related to them— to know if a sentence is true or not.

Conversely, an abstraction attempt confers a *reductionist semantics* to the resulting extended language. Thus, the facts depicted by $\mathcal{L}_2$ -sentences stop from playing a direct contribution to their truth —they do not feature in the corresponding truth conditions. Instead, their truth condition is given in terms of facts involving only their specifications. Moreover, these are the same facts that explain speakers' linguistic competence concerning the extended language —i.e., facts about the unity relation allowing to perform the required abstraction attempt. Consequently, in the case of abstracts, the facts that contribute in a *direct way* to the truth of speakers' statements reduce to those metasemantic facts that *indirectly* contribute to their truth.

Cognitive Reduction The facts that directly contribute to the truth of speakers' statements reduce to metasemantic facts that perform an indirect contribution.

This offers a new perspective on assertion in languages obtained through abstraction —thus about mathematical assertion if this is obtained in this way. The justification for asserting something about the abstracts reduces to a linguistic, metasemantic achievement: the acquisition of linguistic competence in mathematical language. Now, the facts that contribute to the truth of speakers' statements are only metasemantic facts that contribute in an indirect way to their truth. Therefore, speakers only need to know the latter —i.e., they only need to understand the extended language— to know if the corresponding sentences are true or false —thus, to warrant their assertion or denial. Language acquisition comes with a bunch of secure mathematical assertions.

Again, take the case of directions. The linguistic competence concerning expressions committed to directions is explained by facts about the unity relation of parallelism. For instance, speakers understand the content of the predicate “to be orthogonal*” ($\perp^*$) —which holds between directions— because they use it according to Inheritance schema —which governs its use:

$$d(l_1) \perp^* d(l_2) \leftrightarrow l_1 \perp l_2.$$

Thus, they understand what a sentence such as “ $d(l_1)$ is orthogonal* to $d(l_2)$ ” means because they are acquainted with facts about orthogonality between the lines l_1 and l_2 . These are the metasemantic facts which explain how speakers acquire their linguistic competence about directions —they indirectly contribute to the truth of their statements. Moreover, these are the only facts contributing to their truth. The Inheritance schema generates the corresponding truth condition for sentences about orthogonality*:

$$d(l_1) \perp^* d(l_2) \text{ is true iff } l_1 \perp l_2.$$

If the line l_1 indeed is orthogonal* to the line l_2 , “ $d(l_1)$ is orthogonal to $d(l_2)$ ” will be true. As a result, speakers make true statements about directions *just because* they understand the language. Generally, speakers make true assertions in languages introduced by abstraction *just because* they understand these languages.

Cognitive Reduction provides us with an explanatory connection between those facts that explain linguistic competence and the truth of speakers' assertions. As soon as they are acquainted with the

former, they learn to use the language in such a way that their assertions come out as true because of these very facts.⁶⁴ While in the case of empirical sentences, such as “El Capitan is 914m high”, the truth of statements is a function of two parameters —meaning and the world—, in the case of statements about abstracts, their truth is just a function of meaning.

Finally, the truth of beliefs can be explained in a similar way. As in the case of statements, let’s say that agents’ beliefs are true when their content is a true proposition. By “belief” I mean a mental state or an action made by agents, analogous at the level of psychology to statements. Thus, as in the case of statements, it makes sense to vary the content of beliefs, that is, the proposition that is believed —see n. 63. Now, the truth of beliefs can be explained in a similar way. Agents’ beliefs about the abstracts are true because of the same reason statements about the abstracts are true. Those facts indirectly explaining speakers’ linguistic competence, are the same that indirectly explain speakers’ belief formation —thus, linguistic competence and belief formation are different axes of the same process. These facts *indirectly* explain why agents’ beliefs are true. Moreover, no other facts contribute in a direct way to the truth of agents’ beliefs.

Two-dimensional Beliefs The truth value of empirical beliefs is a function of two parameters: the world and those facts fixing their content.

Uni-dimensional Beliefs The truth value of beliefs about abstracts is a function of those facts fixing their content.

This offers a new, and fresh perspective about knowledge —at least about justified belief— obtained through abstraction —thus, about mathematical knowledge if this is obtained in this way. The cognitive accomplishment achieved in mathematics reduces to a linguistic, metasemantic achievement: the acquisition of linguistic competence in mathematical language. The same facts that lead agents to understand the language, lead them to form some beliefs about the abstracts. Furthermore, the facts explaining linguistic competence and belief formation are the only ones that contribute to the truth of statements and beliefs. Again, *Cognitive Reduction* provides us with an explanatory connection between those facts and the fact that agents believe in true propositions. As a result, *mathematical knowledge —at least, mathematical justified belief— reduces to language understanding.*

5.5.3 Solving the Epistemological Challenges

There is a counterfactual connection between those facts explaining linguistic competence and cognitive achievement, and the fact that agents make true statements or have true beliefs. Facts about abstracts as well as counterfactual relations between them and speakers do not play any function in the explanation of the latter. In consequence, mathematical knowledge —at least mathematical justified belief—, if obtained through abstraction, does not satisfy instances of Responsiveness.

⁶⁴ Again, this does not mean that the theory obtained is decidable. If the unity relation $\sim$ or other relations between the specifications are undecidable, then so are identity or the corresponding relations between the abstracts. See (97).

Let φ_2 be a $\mathcal{L}_2$ -sentence and let φ_1 be the corresponding non-homophonic truth condition stated in the base language $\mathcal{L}_1$. Then, Responsiveness is replaced by the following schema:

(a)

Agents S understand $\varphi_2 \wedge \varphi_1 \Vdash$ Agents S are justified in believing φ_2

Agents are justified in believing any $\mathcal{L}_2$ -sentence *because* they understand it and its reductionist truth condition is the case. As can be seen, facts depicted by these sentences as well as counterfactual relations between them and agents disappear from the “equation.” *This is the main difference between empirical knowledge and mathematical knowledge.* While the former demands stronger relations between agents and the world, in accordance with Responsiveness, the latter does not. Furthermore, language understanding is explained by abstraction and inheritance principles. Agents acquire their linguistic competence in the extended language because they perform a successful abstraction based on principles of the form of AP and Inheritance:

Agents S behave according to AP and Inheritance $\wedge \varphi_1 \Vdash$ Agents S understand φ_2 .

Consequently, (a) is equivalent to (b):

(b)

Agents S behave according to AP and Inheritance $\wedge \varphi_1 \Vdash$ Agents S are justified in believing that φ_2 .

That is, agents are justified in believing that φ_2 because they use the extended language according to AP and Inheritance, and it is the case that φ_1 . This constitutes a reductionist explanation of mathematical knowledge. Nevertheless, it no longer reduces to logical knowledge, as fictionalism and multiversism propose. Instead, mathematical knowledge reduces to linguistic competence.

The acceptance of sentences in the extended language is explained in the same way.

(b*)

Agents S behave according to AP and Inheritance $\wedge \varphi_1 \Vdash$ Agents S accept φ_2 ,

where “to accept φ_2 ” means to be disposed to assert it. Mathematicians are disposed to assert $\mathcal{L}_2$ -sentences *because* they understand them and their reductionist truth condition is the case. Again, this constitutes a reductionist explanation of assertion and linguistic behaviour. Linguistic competence obtained through abstraction comes with a collection of assertible sentences and knowledge about the abstracts.

The reductionist nature of this explanation allows us to account for the reliability claim. Indeed, this follows from (b*). Assume mathematicians accept a $\mathcal{L}_2$ -sentence φ_2 . According to (b*), this is so *because* they behave according to AP and Inheritance principles, and φ_1 . But, together these entail φ_2 —see 5.6. Consequently, mathematicians accept a given $\mathcal{L}_2$ -sentence only if it is true:

Reliability Claim

$\forall \varphi (\text{Speakers S accept } \varphi_2 \rightarrow \varphi_2 \text{ is true}).$

As I said, neither (b), (b*), nor this conditional involves a counterfactual relation between facts depicted by $\mathcal{L}_2$ -sentences and speakers. Nevertheless, they involve a counterfactual relation of a slightly different nature: a counterfactual relation between those facts indirectly contributing to the truth of speakers’

statements —on which their acceptance of sentences rest— and speakers. This ensures two things. First, this ensures strong support for instances of the reliability claim. These are not true just by chance. For instance, because mathematicians are lucky enough that their statements tend to be true so far. On the contrary, their statements tend to be true because they are justified in asserting them. Similarly, their beliefs concerning mathematical matters tend to be true because they are justified in believing them. Second, even if mathematical facts are necessary, the explanation still works. The necessity and abstract nature of mathematical objects prevent any counterfactual relation between them and agents. However, these facts play no role in explaining mathematical assertion and belief. Thus, the reduction of mathematical knowledge to linguistic competence still allows for a strong explanatory relation between the facts involved in the latter and agents. Even if mathematics is necessary, we can explain mathematical knowledge.

5.6 CONCLUSION

In this chapter, I have explained how a community of speakers can acquire a whole new language by performing a successful abstraction attempt. The reductionist metasemantic perspective introduced by abstraction principles allows us to explain how speakers enrich their linguistic resources without attributing to them any relation to the objects the language is about. Consequently, abstraction principles can be used to solve the Language Acquisition Challenge without presupposing any counterfactual or logical relation between mathematicians and mathematical facts. In turn, they are not free of problems. They face the famous Caesar problem and the Bad Company problem. However, I explained how both problems can be overcome. The combination of the Piecemeal solution and a predicative reading of quantifiers guarantees their good condition.

Additionally, abstraction principles are rich enough to explain mathematical knowledge. The abstraction process by which speakers extend their language also provides them with knowledge of a collection of truths formulated in mathematical language —and not only logical truths. Thus, mathematical knowledge comes with the understanding of language. Instead of reducing the former to logical knowledge, it is explained as a form of linguistic knowledge. Mathematical knowledge involves a very particular form of cognitive achievement: the correct understanding of mathematical language. First, this proposal can be understood as a modest vindication of universalist platonism. Against the pluralist perspective of fictionalism and multiverse, the abstractionist approach allows us to reintroduce a minimum of mathematical objectivity: objectivity with objects. Besides, this is perfectly consistent with Kreisel's *dictum*: what matters is mathematical objectivity, not mathematical objects. Given the reductionist nature of abstraction principles, mathematical objects do not play an indispensable role in the explanation of mathematical knowledge. Second, this must not be understood as a refutation of fictionalism or multiversism. Once a minimum of mathematical objectivity has been introduced, these conceptions can be defended for other mathematical theories. The debate between them is beyond the scope of the present text, and —if possible— will have to be resolved by other philosophical arguments. The minimum of mathematical objectivity introduced here only constitutes the starting point in the exploration of the multiplicity of alternative mathematical fictions or universes.

APPENDIX 5.A

Let's state the assertibility conditions for languages introduced through abstraction. To simplify things, I will start exposing the assertibility conditions for the new sentences committed to abstracts in a two-sorted language. The expressions of sort 1 are the expressions of the base language $\mathcal{L}_1$ —among them, it should have enough resources to express the equivalence relation $\sim$. Let's use metavariables α and β for terms of sort 1. The sort 2 contains the operator f , which applied to terms α of sort 0 forms complex terms $f(\alpha)$ of sort 1, a countable collection of variables v_i , and a collection of new n -ary predicates P —among them an identity predicate $=_1$. As can be seen, the sort 1 represents the base language for abstraction $\mathcal{L}_1$. Let's assume this already has an interpretation I .

First, for sentences ψ of sort 1, the assertibility condition coincides with its truth condition:

$$\psi \text{ is true iff } \psi.$$

Second, to each n -ary predicate P of sort 2 other than the identity predicate $=_1$, let's associate a formula of sort 1 φ_P such that:

$$\alpha_1 \sim \beta_1 \wedge \dots \wedge \alpha_n \sim \beta_n \rightarrow (\varphi_P(\alpha_1 \dots \alpha_n) \leftrightarrow \varphi_P(\beta_1 \dots \beta_n))$$

and

$$\varphi_P(\alpha_1 \dots \alpha_n) \rightarrow \alpha_1 \sim \alpha_1 \wedge \dots \wedge \alpha_n \sim \alpha_n.$$

That is, the equivalence $\sim$ is a congruence with respect to φ_P , and this is satisfied inside the field of the former. To simplify things, I will state the assertibility conditions relative to an assignment σ of the terms of sort 1 into the domain of I :

If α is a constant of sort 1, $\sigma(\alpha)$ is its referent in I

If α is a variable, $\sigma(\alpha)$ is an element in the domain of I .

Then, the assertibility conditions for sentences of sort 2 are the following.

- (i) $f(\alpha) =_1 f(\beta)$ is assertible relatively to σ iff $\sigma(\alpha) \sim \sigma(\beta)$
- (ii) $P(f(\alpha_1) \dots f(\alpha_n))$ is assertible relatively to σ iff $\varphi_P(\sigma(\alpha_1) \dots \sigma(\alpha_n))$
- (iii) The clauses for connectives are as expected
- (iv) $\forall v \varphi(v)$ is assertible relatively to σ iff $\varphi(f(\alpha))$ is assertible relatively to every σ^* that differs from σ in the value of α and $\sigma^*(\alpha) \sim \sigma(\alpha)$.

The condition that $\sigma^*(\alpha) \sim \sigma(\alpha)$ in clause (iv) is required because the formulas of sort 2 only consider those elements in I related by the unity relation $\sim$.

Finally, to introduce assertibility conditions in a one-sorted framework we have to consider mixed postulates. Take a system of abstraction principles, each one generated by a unity relation $\sim_j$. Let f_j be the abstract-operator corresponding to each unity relation. The system is formed for abstraction and inheritance principles of the form:

AP_j

$$f_j(\tau) = f_j(\tau) \leftrightarrow \tau \sim_j \tau$$

Inheritance_j

$$P(f_j(\tau_1) \dots f_j(\tau_n)) \leftrightarrow \varphi_P(\tau_1 \dots \tau_n).$$

Then, mixed postulates of the form

$$f_j(\tau_1) = \tau_2 \leftrightarrow \psi(\tau_1 \tau_2)$$

$$P(x_1 \dots x_n) \leftrightarrow \gamma_P(\tau_1 \dots \tau_n),$$

—where $x_i = f_k(\tau_i)$ or $x_i = \tau_i$, and $\psi(\tau_1 \tau_2)$ and $\gamma_P(\tau_1 \dots \tau_n)$ are suitable *criteria* expressible in the base language of abstraction $\mathcal{L}_1$ —determine how abstracts behave in mixed contexts involving both abstracts of different kind and other kind of objects. Recall that, to avoid contradictions, when the abstracts introduced by two abstraction principles of the system are identified, the corresponding unity relations $\sim_i$ and $\sim_k$ must be a congruence relative to $\psi(\tau_1 \tau_2)$ and $\gamma_P(\tau_1 \dots \tau_n)$. That is:

Congruence-2

$$\tau_1^1 \sim_r \tau_1^2 \wedge \tau_2^1 \sim_r \tau_2^2 \rightarrow (\psi(\tau_1^1 \tau_1^2) \leftrightarrow \psi(\tau_2^1 \tau_2^2))$$

and

$$\tau_1^1 \sim_r \tau_1^2 \wedge \dots \wedge \tau_n^1 \sim_r \tau_n^2 \rightarrow \gamma_P(\tau_1^1 \dots \tau_n^1) \rightarrow \gamma_P(\tau_1^2 \dots \tau_n^2).$$

Let $\mathcal{L}_2$ be a one-sorted language resulting from extending the base language of abstraction $\mathcal{L}_1$ through this system of abstraction and inheritance principles. Let I be an interpretation for $\mathcal{L}_1$. Now, the assertibility conditions for the extended language can be stated in a similar way to the two-sorted case. Also, let σ be an assignment of terms τ into the domain of I :

if τ is a term of $\mathcal{L}_1$, $\sigma(\tau)$ is its referent in I

if τ is a variable, σ assigns it an object in I

if τ is of the form $f(\tau^*)$, then $\sigma(f(\tau^*)) = \sigma(\tau^*)$.

Then, the assertibility conditions for the vocabulary of the language is given by the mixed postulates:

- (i) $f(\tau_1) = f(\tau_2)$ is assertible relatively to σ iff $\psi(\sigma(\tau_1)\sigma(\tau_2))$
- (ii) $P(\tau_1 \dots \tau_n)$ is assertible relatively to σ iff $\gamma_P(\sigma(\tau_1) \dots \sigma(\tau_n))$,

where P is an n -ary predicate of $\mathcal{L}_2$. Finally, the condition for complex formulas is generated as expected.

6 The Potential Infinite

6.1 INTRODUCTION

Abstraction principles are extremely useful devices to explain mathematical language and mathematical knowledge. Once quantification is read predicatively—that is, at least some abstracts do not belong to the domain of the quantifiers with occurrences in the abstraction principle—they introduce a reductionist perspective in metasemantics. As a result, we can explain how speakers acquire a whole new language and knowledge about the new abstracts without presupposing the existence of counterfactual relations or a previous understanding of these objects by speakers. This makes abstraction principles appropriate to apply in mathematics. Mathematical knowledge is explained by linguistic competence. However, so far I have not shown how these principles are powerful enough to recover serious mathematical reasoning. In the previous chapter, I have only given a few examples of mathematical objects introduced by abstraction—for instance, natural numbers or directions. But, this is not enough to explain mathematical knowledge. We need to explain how those theories about these objects are obtained as the result of the process of abstraction.

In this chapter, I will address this question. I will explain how abstraction principles are sufficient to recover serious mathematical theories. It is known that Hume's principle and second-order logic are powerful enough to interpret Peano's arithmetic (150). However, here I will take a different path. I will explain how abstraction principles allow us to recover Zermelo-Fraenkel set theory (82, Ch. 3, 12). First, I will explain how predicative abstraction principles allow to iterate abstraction attempts. This leads to a series of domain expansions, generating enough objects to recover set theory. Then, I will explain how these expansions can be expressed at the level of the object language using modal operators. As a result, abstraction becomes dynamic. Second, I will contrast set-theoretic potentialism and actualism. Set-theoretic potentialism is the view that the universe of sets forms a plurality of objects. In turn, set-theoretic potentialism denies this. I will argue that the latter is in a better condition than the former, as it frees us from certain explanatory pressure that affects actualism. Third, I will introduce the iterative conception of sets. Then, I will show how the iterated application of abstraction principles provides us with an explanation of its two primitive operations: the Set of and the Powerset operations. Furthermore, this explanation entails set-theoretic potentialism. I will expose the modal principles that govern the iteration of abstraction attempts and the expansion of domains. These principles shape the form of the hierarchy. Also, I will expose those principles that characterize the nature of sets. All these principles are the axioms of *potentialist set theory*. Finally, I will show how the potentialist set theory interprets

Zermelo-Fraenkel set theory. However, the theory alone is not powerful enough to recover the axioms of Infinity and Replacement. Thus, it must be supplemented with additional axioms.

6.2 ITERATED ABSTRACTION

In Chapter 5 I showed how abstraction principles allow the introduction of a new language whose truths come equipped with a satisfactory epistemology. Directions, natural numbers, rational numbers and other sorts of entities can be introduced through successful abstraction attempts. The dynamics of this process are simple. Start with a n -order language $\mathcal{L}_1$ committed to unproblematic objects.¹ For example, lines—let’s assume that lines are unproblematic. Let’s call this the *base language* of abstraction. It has all the usual resources needed to talk about these objects, such as n -order terms and n -order quantifiers. In this case, the language has terms and quantifiers ranging over lines. Additionally, it has enough resources to express an equivalence relation between them: the relation of parallelism ($\parallel$).² Then, using facts about parallelism between lines a community of speakers can stipulate assertibility conditions governing a language $\mathcal{L}_2$. Let’s call this the *extended language*. This contains an abstraction operator f which applies to the n -order variables of the base language forming new expressions which are not part of $\mathcal{L}_1$. Additionally, the extended language contains a symbol “=” that features in *apparent identities* of the form “ $f(\alpha) = f(\beta)$ ”. Then, the assertibility conditions constrain the content of the expressions of the extended language. The assertibility condition corresponding to the apparent identities is given by the following abstraction principle:

DIR

$$d(l_1) = d(l_2) \leftrightarrow l_1 \parallel l_2.$$

This abstraction principle, together with additional stipulations—as the piecemeal solution to the Caesar Problem dictates, see 5.4.1.5—, fixes the content of the expressions composing the sentences. The content is selected as to satisfy the truth value and the role attributed by the assertibility condition. As a result, the complex expressions “ $d(l_1)$ ” and “ $d(l_2)$ ” turn to be terms referring to the directions of l_1 and l_2 respectively, and the apparent identity symbol is a predicate expressing identity. Additional predicates φ^* are “inherited” in a similar way from those formulas of the base language satisfying *Congruence* and *Field*. For example, the assertibility condition corresponding to a predicate expressing a relation of congruence between directions ($\perp^*$) is inherited from the congruence relation between lines ($\perp$):

$$d(l_1) \perp^* d(l_2) \leftrightarrow l_1 \perp l_2.$$

¹The notion of object at use is Fregean in spirit: objects are characterized semantically as possible referents of terms. The result is a broad notion of object. For instance, as soon as theoreticians refer to concepts, concepts are objects too. Nonetheless, the characterization allows for objects of different order and different kinds. This suffices to do justice to Frege’s maxim: “never to lose sight of the distinction between concept and object.” (46, p. x)

²Remember that the base language $\mathcal{L}_1$ is not necessary for an abstraction attempt to take place. All it requires is that agents have enough information about the objects and the equivalence relation. But, this information can be non-linguistic. Abstraction is not a phenomenon restricted to language. For the importance of non-linguistic elements in the practice of mathematics see (35). For simplicity, I will assume that $\mathcal{L}_1$ is given to speakers.

A community of speakers embarking on such an abstraction process ends up with a new language that allows them to talk about directions.

Generally, given an n -order language, abstraction principles—together with additional stipulations—of the following form

AP

$$f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta,$$

—where α and β are n -order terms and $\sim$ is the *unity relation*—allow a community to extend their language obtaining resources to speak and reason about the *abstracts* denoted by the complex terms $f(\alpha_i)$. Additional predicates expressing properties and relations of the abstracts will be introduced by inheritance principles of the form

Inheritance

$$\varphi * (f(\alpha_1) \dots f(\alpha_n)) \leftrightarrow \varphi(\varphi_1 \dots \varphi_n).$$

The extended language provides them with all the resources needed to refer to and make assertions about the abstracts.

However, although abstraction principles look like a promising device, so far I have not explained how an abstraction attempt introduces mathematical language in an *absolute* sense. I have only explained how abstraction principles introduce a language committed to abstracts *relatively* to a given kind of specifications. For instance, directions are specified by lines. Likewise, RT introduces rationals from naturals. And HP introduces naturals from concepts understood as intensional second-order entities. The problem is that all these specifications are as problematic as mathematical objects. Indeed, lines—according to a widely shared notion of line—and natural numbers are mathematical, abstract objects. In turn, intensions are abstract objects too. But, if the specifications are as problematic as the abstracts, what is the benefit afforded by abstraction principles? To solve the epistemological challenges abstractionism must rest on a secure basis. Therefore, something must be said about this basis. Furthermore, there is a second problem. Although I mentioned some abstraction principles used to introduce mathematical entities, I have not explained how abstraction motivates any serious mathematical theory. If abstractionism claims to be a successful explanation of mathematical language and mathematical knowledge, it must recover enough mathematics. For instance, a good test would be arithmetic. If abstractionism can explain how basic arithmetic is introduced through abstraction, then we would have a case for it. However, this does not come so quickly. It requires explaining how the whole domain of natural numbers, and the basic operations defined on them—the successor, the sum and the product—are introduced through this process.

The solution to the second problem is given by the predicative solution to the Bad Company problem—see 5.4.2. In the *Grundlagen* as well as in the *Grundgesetze*, Frege considered that abstraction principles did not enrich the power of quantifiers. For him, the objects introduced by abstraction principles fell in the domain of the quantifiers of the base language. For instance, despite DIR supplies us with additional linguistic resources, terms referring to directions, in a way, we already alluded to these directions when we spoke of things in general—i.e. when speaking of all the things. For Frege, quantifiers of the base language express a form of absolute generality. It is this enormous expressive power what leads Basic Law V into Russell's paradox. If the quantifiers in the expression “the set r of *all* non-self-

membered sets” —in symbols $\forall z(z \in r \leftrightarrow z \notin z)$ — range over absolutely all the objects, including the set characterized by the expression, a contradiction follows. Russell’s solution was to replace the impredicative Law V with its predicative version. Once the expressive power of the quantifiers is conveniently restricted, and the set r is outside the domain where the quantifiers in the base domain are interpreted, the contradiction disappears. Therefore, this restriction suggests itself as a general solution to the Bad Company problem. The problem is not solved by laying down criteria to identify which principles are safe, but by abandoning absolute generality: any predicative abstraction step is apt to expand the domain of quantifiers.

Impredicative Abstraction An abstraction attempt is *impredicative* if and only if all the abstracts fall in the range of the quantifiers of the base language.

Predicative Abstraction An abstraction attempt is *predicative* if and only if it is not impredicative.

A predicative abstraction step expands the domain. At least part of the new abstracts do not belong to the domain where the quantifiers were originally interpreted. Let’s call this the *base domain* of abstraction. The expansionist nature of this form of abstraction not only provides a solution to a long-standing problem —the Bad Company problem. It offers an additional advantage, allowing to iterate the process of abstraction. After an abstraction attempt, the quantifiers can be interpreted in the new, enlarged domain. Then, by the same reasoning as before, predicative abstraction takes place over this new domain, generating even more objects. The process can be repeated ad infinitum. This endows abstractionism with sufficient richness to recover minimally complex mathematical concepts. In this way, we get rid of the second of the problems discussed above. The former —the relative nature of abstraction— is overcome by choosing the appropriate base from which to start the iteration.³

So far, the idea of expanding the domain of quantification has only been sketched out. However, additional resources are required to express the domain expansions in the language. Once these are in place, I will show in 6.5 how the iterated application of abstraction principles allows us to recover basic mathematical concepts and a minimal mathematical theory. First of all, note that abstraction principles of the form of AP perform a double task. On one side, they introduce new entities. More specifically, they assert that for all specifications α in the field of the unity relation $\sim$, there is an abstract $f(\alpha)$. In symbols:⁴

Collapse

$$\forall \alpha(\alpha \sim \alpha \rightarrow \exists \beta(f(\alpha) = \beta)).$$

³A word of caution here. Theoreticians speak about the iteration of abstraction attempts. In turn, agents are not required to possess any sophisticated mathematical notion of iteration at this point. A successful explanation of mathematics does not need to presuppose scepticism. Theoreticians are allowed to use mathematics to explain mathematics. Such an explanation must be scrupulous only on one point: it must not be reduced to the assumption that mathematical methods are reliable.

⁴Adding a new predicate “ $Abs(\alpha, x)$ ” expressing that x is the abstract specified by α , this can be also expressed as follows:

$$\forall \alpha(\alpha \sim \alpha \rightarrow \exists \beta Abs(\alpha, \beta)).$$

Collapse asserts the existence of some abstract corresponding to each specification. Notice that the universal quantifier and the existential quantifier are interpreted in the same domain of objects. Consequently, the base language of abstraction already quantifies over the abstracts. On the other side, abstraction principles provide the abstracts with criteria of identity. According to AP, two abstracts are equal if and only if their specifications are related by the unity relation. In symbols:⁵

AP_I

$$\forall \alpha \beta (\alpha \sim \alpha \wedge \beta \sim \beta \rightarrow (f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta))$$

Abstraction principles of the form of AP are equivalent to the conjunction of *Collapse* and AP_I. Therefore, they can be “factorized” into these two components. This will facilitate the work and allow a better understanding of what is behind the expansions of domains.

The domain expansion performed by a predicative abstraction attempt entails that the universal and the existential quantifiers in *Collapse* are interpreted in different domains. The existential quantifier is interpreted in a domain which extends the domain of the universal quantifier —see figure 6.1. Therefore,

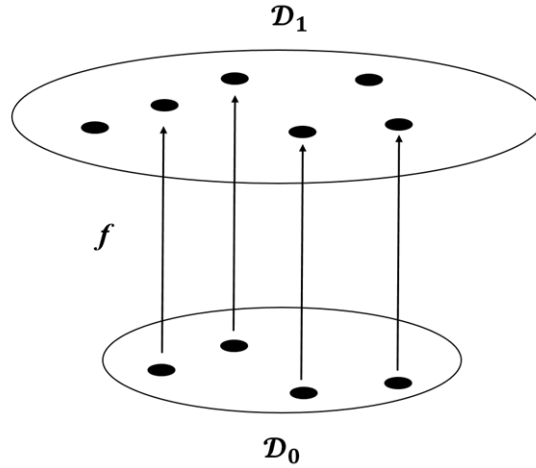


Figure 6.1: Domain expansion: the domain $\mathcal{D}_1$ expands the domain $\mathcal{D}_0$ by adding abstracts specified by the elements in this domain.

Collapse says that all the objects in the base domain of abstraction specify an abstract placed in another, larger domain. The existential quantifier requests a shift of domains. However, this shift is hidden by the classical formulation of abstraction principles. What is required is to make explicit such a shift. Here, there are two options. First, a predicate “*Old*(α)” could be added to the classical language to restrict abstraction to its extension. The collection of old objects —i.e. of objects satisfying “*Old*(α)”— would be the base domain of abstraction. However, this option becomes unpractical for two reasons. First, it complicates the formulation of abstraction principles in excess.⁶ Second, iterating abstraction will require more than one “*Old*(α)” predicate. Consequently, the process of abstraction will be strongly conditioned by the number and type of them.

⁵Again, if a predicate “*Abs*(α, x)” is added, this could be expressed as follows:

$$\forall \alpha_1 \alpha_2 \beta_1 \beta_2 (Abs(\alpha_1, \beta_1) \wedge Abs(\alpha_2, \beta_2) \rightarrow (\beta_1 = \beta_2 \leftrightarrow \alpha_1 \sim \alpha_2)).$$

⁶*Collapse* and AP_I should be written as follows:

A better solution is to adopt a richer language, able to deal with jumps from one domain to another.⁷ Indeed, shifts of this kind can be expressed by adding modal operators $\Box$ and $\Diamond$ to the classical language. The operators are to be understood as expressing a shift from smaller to larger domains performed by a series of abstraction steps. When applied to a formula φ , the operator $\Box$ says that abstraction does not affect the truth value of φ . On the contrary, the operator $\Diamond$ says that a series of abstraction steps can make φ true. Thus, the semantics of these modal operators should be understood informally as follows:

$$\begin{aligned}\Box\varphi &\equiv_{df} \text{no matter what abstraction steps speakers carry out, it will remain the case that } \varphi \\ \Diamond\varphi &\equiv_{df} \text{speakers can abstract until it is the case that } \varphi.\end{aligned}$$

This form of modality diverges from the most common ones, such as physical, metaphysical or conceptual. According to Fine's (42) terminology,⁸ this is *interpretational* modality, as opposed to *circumstantial* modality. While the latter involves a shift in the metaphysical situation, the former only affects the interpretation of the language. In particular, the interpretational possibilities are compatible with the actual metaphysical situation. By using the new modal operators, *Collapse* can be stated as follows:

Collapse ^{$\Diamond$}

$$\Box\forall\alpha(\alpha \sim \alpha \rightarrow \Diamond\exists\beta(f(\alpha) = \beta))$$

This modal principle explicitly says that the existential quantifier can be interpreted in other, larger domains. It says that no matter where speakers are in the abstraction process, for all the specifications in the field of the unity relation $\sim$, they can abstract as to refer to the corresponding abstract $f(\alpha)$. Similarly, the identity criterion can be reformulated as follows:

AP_I ^{$\Diamond$}

$$\Box\forall\alpha\beta(\alpha \sim \alpha \wedge \beta \sim \beta \rightarrow (f(\alpha) = f(\beta) \leftrightarrow \Box(\alpha \sim \beta))).$$

It says that no matter where speakers are in the process of abstraction, two abstracts are the same not only if their specifications are related by the unity relation $\sim$, but if necessarily so —i.e. if they are related after any abstraction step.

The use of interpretational modality comes with a second benefit. Predicative abstraction avoids paradox by allowing the domain to expand. However, this weakens the power of the language, since absolute quantification is lost. If the base domain of abstraction can be enlarged, this means that the classical quantifiers “ $\forall$ ” and “ $\exists$ ” do not cover all the objects. At least some abstracts in the expanded

Collapse

$$\forall\alpha(Old(\alpha) \wedge \alpha \sim \alpha \rightarrow \exists\beta(f(\alpha) = \beta))$$

AP_I

$$\forall\alpha\beta(Old(\alpha) \wedge Old(\beta) \wedge \alpha \sim \alpha \wedge \beta \sim \beta \rightarrow (f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta)).$$

⁷An early model-theoretic analysis of domain expansions in terms of supervaluations over an abstraction principle can be found in (61; 73). The supervaluation operation can be iterated, generating a sequence of models such that (i) the domain of one model expands the domain of the previous model in the series, (ii) all models in the series satisfy the abstraction principle, and (iii) the series reaches a fixed point, where all truths of the theory are determined. However, they do not incorporate modal resources in the object language to express the expansions of domains. Nor they do focus on predicative abstraction, offering an analysis of its more classical, impredicative forms.

⁸Linnebo (82, p. 61) also adopts this terminology.

domain do not belong to the base domain of quantification. Besides, if expansions are performed *ad infinitum*, there is no way to make quantifiers to quantify over absolutely all objects. Nevertheless, the modal operators retrieve absolute generality into the picture. When the quantifiers are prefixed with modal operators they generalize not only over the objects available at some step of the abstraction process. As the interpretation of the modal operators reveals, a much stronger form of generality arises: they quantify over all the objects at all steps of the abstraction process. A sentence of the form $\Box\forall\alpha\varphi$ says that for all objects at all steps of the abstraction process, φ is true. In turn, $\Diamond\exists\varphi$ says that for some object at some step of the abstraction process, φ is true. Let's call the expressions " $\Box\forall$ " and " $\Diamond\exists$ " *modalized quantifiers*. In 6.5 I will show how a minimally complex mathematical theory —set theory— can be explained by the iteration of abstraction attempts, and how the modal operators provide the necessary resources to express the concepts and to explain the axioms of the theory.

6.3 POTENTIALISM VS ACTUALISM

The kind of generality introduced by the modalized quantifiers can be understood as supporting a particular approach to the infinite and to mathematics in general: *potentialism*.⁹ This is opposed to the currently dominant conception in the mathematical community, *actualism*. Potentialism and actualism introduce a qualification in the nature of infinite collections of objects. Loosely speaking, potentialism concerning a given collection A is the view that objects in A do not form a "completed collection". In turn, actualism concerning a collection A is the view that objects in A do form an "actual" or "completed collection".¹⁰

These conceptions can be clarified using the concept of a *plurality of objects*. A plurality of objects is no more than these objects. Indeed, pluralities should not be thought of as an additional object over and above these objects —as, for instance, sets are. They are a convenient, nominal device to speak

⁹For expositions of the varieties of potentialism see (86; 87). Hints of this position are found in (137, pp. 73-74): "So we shall refer to Ω [the type of ordinals] as an *incomplete type*. Similarly, types built up from ω using the other logical operations will be called incomplete. [...] One may take towards the 'potential infinity' Ω the attitude that Brouwer took to the potential infinity ω ." [Italics in the original]. Here Tait anticipates the current distinction between set theoretic potentialism —loosely speaking, the idea that the collection of sets is an incomplete collection— and Aristotelian potentialism —again loosely speaking, the idea that the collection of natural numbers is an incomplete collection— (86; 87). Other references to potentialism can be found in (68), who already characterizes the notions of actualism and potentialism in terms of pluralities.

¹⁰The distinction has its origins in the historical debate on the nature of infinity. As it is famously known, Greek mathematicians and philosophers were extremely scrupulous about this idea. For them, there was only potential infinity. For example, Aristotle rejected that the kind *natural number* selected a complete infinite collection of those objects. Again loosely speaking, each natural number is generated from its predecessor, but the complete collection of naturals is never generated. This attitude towards the infinite lasted for centuries, and even modern mathematicians such as Gauss manifested sceptic qualms about infinite collections — including the collection of natural numbers:

I protest against the use of infinite magnitude as something completed, which is never permissible in mathematics. Infinity is merely a way of speaking. (Karl Friedrich Gauss, Briefwechsel mit Schumacher, *Werke*, Band 8, p. 216, 1831. Translation from (87, p. 1202)

The change of attitude came about in modern mathematics as a consequence of the acceptance of infinitary methods. This led Dedekind, Cantor and other contemporary mathematicians to the concepts of real number and actual infinity. See (133, Ch. 2-5) and (35, Ch. 8) for a survey of the development of the procedures that gave rise to these ideas and to contemporary set theory.

in the singular about the plural. In consequence, they are the quintessential example of an *extensional collection* of objects: their “identity” reduces to the identity of their members —see 6.5 for an analysis of what an extensional collection is. With this device on the table, the ideas behind actualist and potentialist approaches to infinity can be expressed simply and clearly. Actualism concerning a collection A can be understood as the thesis claiming that the members of a given collection A form a plurality.

Actualism There is a plurality of objects among which each object in A is found.

Therefore, the idea of an actual or completed infinite collection of objects becomes the idea that those objects form an infinite plurality. In turn, potentialism concerning a collection A claims that its members are not exhausted by any plurality of objects.

Potentialism For every plurality of members of a collection A, there is another member of A that is not among the objects forming the plurality.

Then, a potential infinity of objects is an infinite collection that is not encompassed by a plurality —the members of the collection do not form a plurality. As can be seen, actualism and potentialism are opposite positions: each one entails the negation of the other.

From this general characterization, a finer classification of potentialist and actualist positions emerges depending on which collections are taken to substitute A. For instance, one can be very restrictive and reject actualism even about the natural numbers —thus, rejecting actualism concerning any other infinite collection. The result is the traditional Aristotelian potentialism.¹¹ In turn, other, more moderate positions accept actualism about some infinite collections while endorsing potentialism about others. Thus, one can be an actualist concerning the naturals, the reals and other collections of numbers, while being a potentialist about the hierarchy of sets.

Set theoretic Potentialism For every plurality of sets, there is a set not in this plurality.

Set theoretic Actualism There is a plurality of objects among which each set is found.

As is well-known, set-theoretic paradoxes forbid the existence of a set encompassing every set —i.e. the set of all sets. Furthermore, because of the extensional nature of both sets and pluralities — which suggests a close connection between them—, this can be taken as evidence against set-theoretic actualism. One might argue that the reasons for the non-existence of the set of all sets have to do with the radical difference between the idea of a collection which encompasses every set and the sets themselves: the former is too big to be accommodated to extensional concepts.¹² But, if this is so, no plurality of sets will contain every set either. Being extensional in nature, pluralities are as unacceptable as sets for assimilating collections of this size. However, no matter how compelling this reasoning is, the fact remains that it does not follow from the paradoxes alone. It involves a certain degree of commitment to the idea that size is the ultimate cause of them. However, the most extended position maintains that the collection of all sets does indeed form a plurality.

¹¹ Terminology is taken from (86; 87).

¹² See (91, pp. 484-485) for a careful exposition of the *limitation of size* rule of thumb. The paper constitutes one of the most comprehensive studies of the different forms of evidence used in set theory.

In 6.5 I will show how the modal resources used to express the endless iteration of abstraction attempts supply us with enough resources to elaborate the core ideas of potentialism. In this way, abstractionism provides an explanation of set-theoretic potentialism. Hitherto, there is no crucial argument to choose between these two philosophical positions, that is, between potentialism and actualism. The reasons given in the debate are often mixed with personal preferences and other prejudices. The agreement is still a long way off. However, there is a point on which potentialism scores better. For, it dispenses with some explanatory pressure posed on set-theoretic actualism.¹³ Actualists owe us an explanation of why some pluralities form sets, while others do not. According to actualism, the collection comprising every set is a plurality: the plurality of all sets. Besides, set theory tells us that an enormous amount of its subpluralities give rise to a corresponding set: the set of all sets among the given plurality of objects. Then, the question arises: what does it take for certain pluralities to form a set and not others? A standard response offers a simple solution to this puzzle: just logic. Russell's paradox shows that the sentence stating the existence of the Russellian set is a logical contradiction. Thus, the plurality of non-self-membered sets can not form a set on the grounds of logic alone. Logic tells us that some pluralities form sets, while others do not. Nevertheless, this answer is unsatisfactory for at least two reasons. First, it is incomplete. At best it explains why *one* plurality does not form a set. It does not explain why other pluralities, such as the plurality of all sets, do not generate their corresponding set. Logic alone is not sufficient in these cases to derive a contradiction—the axiom of Separation is required to prove a contradiction from the assumption that the set of all sets exists. Second, it is false that logic alone suffices to prove that the plurality of all non-self-membered sets does not form a set. What Russell's paradox shows is that speakers cannot quantify over the Russellian set—i.e. the set co-extensional with the plurality of non-self-membered sets in the speakers' domain of quantification cannot be in this domain too. This is not the same as saying that such plurality is not coextensive with a set. To prove this an additional metasemantic assumption is required: the assumption that speakers quantify over absolutely all sets. However, this is far from being immediate. Indeed, the most natural answer to the metaphysical question “what does it take for a plurality to form a set?” is “nothing.” Nothing is required for a plurality to form a set. Thus, every plurality is coextensional with a set.¹⁴ But, this is something that potentialism can perfectly accept.

6.4 THE ITERATIVE CONCEPTION OF SETS

The *iterative conception* is a structural conception concerning the disposition of sets. According to it, sets are arranged in an ordered series of increasingly inclusive stages or levels. Thus, each stage contains all the sets contained in previous stages. The iterative conception is usually taken as an informal source of intrinsic evidence for more formal theories of sets which dispense with the language of stages.¹⁵ More precisely, the statements describing the conception would be analytically true of speakers' notion of set—at least they would be analytically true of some concept of set. Here, I will only provide an informal

¹³See (68, pp. 17-20) and (134, pp. 699-700) for similar arguments.

¹⁴Studd (134, p. 699) calls this the *Maximality Thesis*.

¹⁵Maddy (91) qualifies it as one of the main “rules of thumb” used for justifying ZF axioms less Replacement, and locates its origin in Zermelo's foundational paper. See (91; 92) for a survey of the main sources of evidence used in set theory.

gloss of the conception. However, in their seminal texts, Shoenfield (127) and Boolos (10) developed it as a formal *stage theory*.¹⁶

As usual, sets bear a relation of membership to sets —members of a set are also called its elements. Additionally, sets bear a relation to the stages in which they are contained. The iterative conception is coached then in a collection of principles describing the actions of two primitive operations: the *Set of* and the *Power* operations. Given a plurality of sets¹⁷ contained in a given stage, the *Set of* operation takes this plurality to a set whose members are exactly these objects —i.e. *the set of* these objects.

Set of Given a plurality of objects, *Set of* relates these objects to the set whose members are exactly those objects.

In turn, the *Power* operation takes a collection of sets A to the collection $\mathcal{P}$ of all sets of sets in A . A note of caution here: the *Power* operation is not assumed to be the same as the Powerset operation ((P)). While the former applies to collections —indeed, it is intended to apply to the stages, which are collections of sets—, the second applies to a particular kind of collection: sets. As I will show, the Powerset operation allows us to interpret the iterative hierarchy in set theory. By considering arbitrary sets, this means that the power collection not only contains those sets which are characterized by particular expressions. For instance, when dealing with countable languages the set of predicates will have at most as many members as elements have ω . But, by Cantor's theorem $\mathcal{P}(\omega)$ —which is coextensional with the power collection $\mathcal{P}(\omega)$ — has more members than ω . Thus, $\mathcal{P}(\omega)$ has more elements than predicates — some sets are not characterizable using these predicates. Note that the *Power* operation can be defined in terms of the *Set of* operation given some assumptions. First, if a quasi-combinatorial notion of plurality is introduced, allowing for non-describable pluralities, every set of elements in A matches a particular plurality of objects —the plurality of its members. This yields a notion of arbitrary plurality. Second, given a collection A assume that the action of *Set of* on the pluralities of objects in A also yields a collection of sets. Then, given a collection of sets A , by the first assumption *Set of* forms every set of sets in A . By the second, there is a collection in which every such set is found: the power collection $\mathcal{P}(A)$. However, these assumptions do not immediately follow from the description of the *Set of* operation alone.

Power Given a collection A , *Power* relates it to its power collection $\mathcal{P}(A)$ containing all arbitrary sets of objects in A .

The iterative hierarchy of sets is structurally articulated by the combination of the two primitive operations. All sets contained in some stage arise from the iterated application of them. At the initial stage S there are no sets, just an empty plurality of objects. Applying the *Power* operation to this the next stage $\mathcal{P}(S)$ is generated. This contains all sets of sets in S . Then, we can take the power again forming the next stage $\mathcal{P}(\mathcal{P}(S))$. As before, it contains all sets of sets in $\mathcal{P}(S)$. Going on with this process we get all the finite stages. However, the process does not stop here. At this point, the *Set of* operation takes the plurality of sets contained in finite stages and yields another stage. Thus, the first limit stage is

¹⁶See (100), (68, pp. 13-17), and (134, pp. 698-699) for expositions of the iterative conception.

¹⁷Remember that a plurality of objects is not an object over and above the objects. They are only a linguistic device used to talk in singular about the *many*.

reached, which comes immediately after all of them. This stage will contain all finitely hereditary sets. Then, a new round of applications of the Power operation yields increasingly inclusive stages until a new limit stage is formed by the Set of operation. The iteration continues endlessly. The iterative hierarchy can be represented in Zermelo-Fraenkel set theory (ZF). The Powerset operation ($\mathcal{P}$) retrieves the Power operation, and Replacement allows a sufficiently strong notion of ordinal codifying the endless iteration generating the hierarchy.¹⁸ The stages are then identified with a particular kind of sets, the cumulative

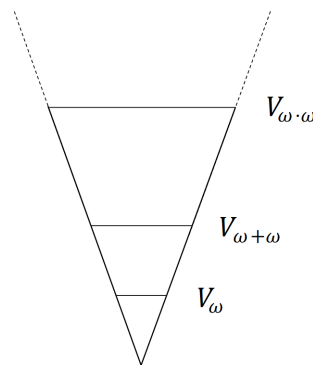


Figure 6.2: The iterative hierarchy of sets.

ranks, defined recursively over the collection of ordinals. The first stage is $\emptyset =_{df} V_0$. For each stage V_α , we let $V_{\alpha+1} =_{df} \mathcal{P}(V_\alpha)$ be its successor. And, limit stages $V_\lambda =_{df} \bigcup_{\alpha < \lambda} V_\alpha$ are defined as the general union of all previous stages.

Finally, the action of the Set of and the Power operations is coached by the following structural principles: the Priority and the Plenitude principles. Let's say that a set *is formed* at a stage if this is the first stage in the series containing it. Then, the Priority principle says that the elements of a set formed at a given stage are formed at previous stages. In turn, the Plenitude principle says that every set belongs to all the stages equal or greater than the stage in which it is formed.

Priority If a set is formed at some stage its elements are formed at previous stages.

Plenitude Every set belongs to all stages equal or greater than the stage at which is formed.

These two principles describe some particularities of the distribution of sets across the hierarchy. The Plenitude principle ensures that the hierarchy is cumulative —no set is lost when the two primitive operations are iterated. The Priority principle ensures that the order of stages respects membership. The description of the iterative conception is completed by adding the following four principles. Two principles of extent, Origin and Infinity:

Origin There is an initial stage.

¹⁸The importance of Powerset and Replacement was pointed out by (139, p. 42):

It is no exaggeration to say that these two principles (the axioms of the power set and of ordinals) are all that matter in set theory. Namely, every set can be obtained by starting with the empty set and by transfinite iteration of application of the principle of constructing a power set along with construction of ordinals. We can thus say that the study of the axioms of set theory will be complete if we can understand all about constructions of ordinals and power sets.

Infinity There is a stage different from the initial stage that is not the successor of any stage.

The Origin principle states the fact that the iterative process needs a starting point from which to begin. Of all the starting points the simplest and the one which generates all the pure sets is the empty stage—the stage which contains no objects. In turn, the principle of Infinity ensures the existence of a stage containing infinite many sets. Without this principle, the iteration could not enter the transfinite, and the hierarchy would be reduced to the hierarchy of hereditarily finite sets. Finally, two principles of nature are added: Extensionality and Foundation:

Extensionality Two sets are identical just in case they share their members.

Foundation Every non-empty set has a $\in$ -minimal element.

While the above principles characterize a structure shared by hierarchies of objects resulting from generative operations of the type of Set of and Power, the principles of nature refer specifically to sets. They describe some characteristic properties of sets. As expected, the principle of Extensionality offers a criterion of identity for sets—two sets are identical just in case they are co-extensional. The principle settles the extensional nature of sets, which distinguishes them from other intensional entities such as concepts or properties. The principle of Foundation ensures the good behaviour of the membership relation ($\in$). In particular, it entails that no set is a member of itself.

The iterative conception is considered one of the main sources of intrinsic evidence supporting the axioms of Zermelo-Fraenkel set theory (ZF)—see (10) for a formal exposition of the iterative conception in the form of a *stage theory*, as well as for a formal derivation of ZF's axioms except Replacement from this.

Definition 1 (ZF) Let $\mathcal{L}_\in$ be a first-order, classical language whose unique non-logical constant is the binary predicate “ $\in$.” The intended meaning of this predicate is the membership relation: “ $y \in x$ ” means that y is a member of x . Then, ZF is the $\mathcal{L}_\in$ -theory which is recursively axiomatized by the following axioms:

<i>Emptyset</i>	$\exists x \forall y (y \notin x)$
<i>Extensionality</i>	$\forall xy (x = y \leftrightarrow \forall z (z \in x \leftrightarrow z \in y))$
<i>Pairs</i>	$\forall uz \exists x \forall y (y \in x \leftrightarrow y = u \vee y = z)$
<i>Union</i>	$\forall u \exists x \forall y (y \in x \leftrightarrow \exists z (y \in z \wedge z \in u))$
<i>Separation</i>	$\forall u \exists x \forall y (y \in x \leftrightarrow y \in u \wedge \varphi(y))$
<i>Powerset</i>	$\forall u \exists x \forall y (y \in x \leftrightarrow \forall z (z \in y \rightarrow z \in u))$
<i>Foundation</i>	$\forall x (\exists y (y \in x) \rightarrow \exists y (y \in x \wedge \forall z (z \in x \rightarrow z \notin y)))$
<i>Infinity</i>	$\exists x (\exists y \forall z (z \notin y \wedge y \in x) \wedge \forall y (y \in x \rightarrow \{y\} \in x))$
<i>Replacement</i>	$\forall x \exists ! y \psi(xy) \rightarrow \forall u \exists z \forall x (z \in u \rightarrow \exists y (y \in z \wedge \psi(xy)))$

Separation is an axiom schema where $\varphi(y)$ is a $\mathcal{L}_\in$ -formula with a free variable “ y ”, and the variable “ x ” is not free in $\varphi(y)$. Zermelo set theory (Z) is ZF less Replacement. In turn, S is Z less Infinity.

A note of caution here. As usual, “ $\{x_1 \dots x_n\}$ ” denotes the set whose elements are $x_1, \dots, x_n$. Thus, the brackets “{” and “}” are a convenient terminology introduced by definition. ZF’s axioms can be stated without this resource.¹⁹

However, in the previous and the next sections, the brackets play the role of an abstraction operator—an operator that applies to terms forming complex terms. In the following section, I will explain how abstractionism provides an analysis of the Set of and Power operations in terms of abstraction principles. The iteration of abstraction attempts is rich enough to provide an explanation of the iterative conception. Using the interpretational modality sketched in [Section 2](#), this process can be described by a modal theory: this is *potentialist set theory*. Potentialist set theory is a powerful mathematical theory. It entails the appropriate modal formulation of S’s axioms—ZF’s axioms less Infinity and Replacement. Consequently, it recovers a minimally complex amount of mathematical thinking—at least it recovers Peano-Dedekind arithmetic.²⁰ However, can we do better? After exposing the theory, I will discuss whether abstractionism can recover both Infinity and Replacement, thus retrieving ZF in full.

6.5 POTENTIALIST SET THEORY

In this section, I will explain how the iteration of abstraction attempts allow us to explain the primitive operations of the iterative conception of sets: the Set of and the Powerset operations, as well as the principles characterizing the hierarchy. I will expose the modal principles that govern the iteration of abstraction attempts and the expansion of domains. These principles shape the form of the hierarchy. Also, I will expose those principles that characterize the nature of sets. The result is the *potentialist set theory*. Modal reconstructions of mathematics have been proliferating for some time. Among them, there are two main families: structuralist (106; 55; 8) and non-structuralist potentialist theories (101); (134); (103); (81, Ch. 3, 12). Here I will follow the exposition in (82, Ch. 3, 12) and (86; 87).

6.5.1 The Plural Hierarchy

First, we need to understand which is the main engine of “set production.” In line with abstractionism, this will be a plural version of Frege’s Law V—let’s call it the Plural Law V (PLV). Let “ uu ” and “ vv ” be plural variables ranging over pluralities of objects. Remember that a plurality of objects is not an object over and above these objects. Plural variables as well as plural names are no more than nominal devices to speak in singular about the plural. Also, let “ $\prec$ ” be a binary predicate representing the relation *to be among*: “ $y \prec uu$ ” means that y is among the uu . Then, Plural Law V is the following abstraction principle:

PLV

$$\forall uu \forall vv (\{uu\} = \{vv\} \leftrightarrow uu \equiv vv),$$

¹⁹The axiom of Infinity is rewritten as follows:

$$\exists x (\exists y \forall z (z \notin y \wedge y \in x) \wedge \forall y (y \in x \rightarrow \forall z (\forall u (u \in z \leftrightarrow u = y) \rightarrow z \in x))).$$

²⁰See (21, Ch. 1) for the relation between PA and ZF less Infinity and Replacement.

where “ $uu \equiv vv$ ” abbreviates the equivalence “ $\forall z(z \prec uu \leftrightarrow z \prec vv)$.” Plural Law V says that the set of uu is the same as the set of vv if and only if the objects among the uu are the same as the objects among the vv —i.e. uu and vv are coextensional and, thus, the same. PLV and Frege’s Law V are similar, but they differ in two respects. First, unlike Frege’s law V, which abstracts over concepts —it assigns an extension to each concept—, PLV abstracts over pluralities of objects. It assigns a set to each plurality, establishing an intimate connection between these two kinds of extensional entities. Second, PLV is a predicative abstraction principle. At least part of the sets introduced do not belong to the domain where the universal quantifier is interpreted —for instance, this is the case of the Russellian set corresponding to the plurality of all non-self-membered sets. On the contrary, it would be inconsistent for exactly the same reason as Frege’s Law V is.²¹ Consequently, the predicative nature of PLV results in the expansion of the domain of quantification. Besides, this allows us to iterate its application in the expanded domains, generating more objects each time. But, to express its iteration the modal idiom is required.

As explained, PLV performs a double job. First, it introduces new entities: sets. It asserts that for all plurality of objects uu which is coextensional with some plurality of objects, there is a set $\{uu\}$ whose members are exactly the objects among the uu —i.e. the set of uu . But, any plurality is coextensional with itself. Therefore, PLV asserts the existence of the set of uu , for any plurality uu . Additionally, it provides sets with a criterion of identity: two sets are identical if and only if the corresponding pluralities are the same. That is, PLV assigns to each plurality a set, and to different pluralities different sets. Thus, it can be factorized into a criterion of existence and a criterion of identity.

Collapse

$$\forall uu \exists x (\{uu\} = x)$$

PLV_I

$$\forall xz \forall uu vv (x = \{uu\} \wedge z = \{vv\} \rightarrow (x = y \leftrightarrow uu \equiv vv))$$

Finally, using the interpretational modality introduced in 6.2, these criteria are turned into modal principles, and the domain expansions performed by PLV are made explicit.

Collapse[◇]

$$\Box \forall uu \Diamond \exists x (\{uu\} = x)$$

PLV_I[◇]

$$\Box \forall xz \Box \forall uu vv (x = \{uu\} \wedge z = \{vv\} \rightarrow (x = y \leftrightarrow \Box (uu \equiv vv)))$$

According to the abstractionist reading of the modal operators, Collapse[◇] says: no matter where the speakers are in the process of abstraction, for every plurality of objects uu so far introduced, they can abstract as to introduce the set $\{uu\}$. In turn, PLV_I[◇] says: no matter where speakers are in the process of abstraction, two sets are the same if and only if their corresponding pluralities are the same no matter the process of abstraction.

The new principle of Collapse[◇] is the main potentialist engine of set production. It may be useful to compare the role of pluralities with the role played by formulas in Gödel’s constructive hierarchy L.²²

²¹ Just replace concepts with pluralities in the derivation of Russell’s paradox —Chapter 5, 5.4.2.

²² See (21, Ch. 3) for a detailed exposition of L.

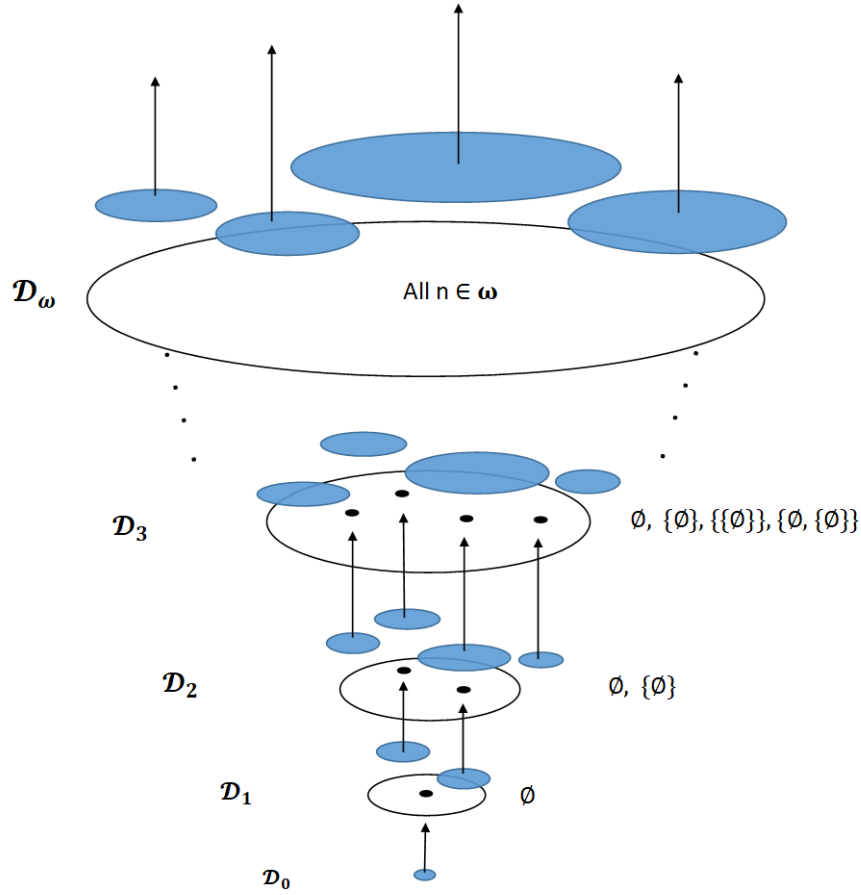


Figure 6.3: The plural hierarchy of sets. The first-order domains $\mathcal{D}_\alpha$ contain sets represented by black dots. Pluralities of sets are represented by blue circles. According to PLV, each plurality is mapped by an arrow to a set. The origin of the hierarchy is the empty plurality. The domain $\mathcal{D}_\omega$ contains all sets generated at finite stages. From there, the hierarchy enters the transfinite.

This is the iterative hierarchy built by stages using $\mathcal{L}_\in$ -formulas —formulas are the main engine of set production. Given a layer of sets, the next layer is formed by adding, for all $\mathcal{L}_\in$ -formulas $\varphi(x)$, a new set: the set of those sets in the previous layer satisfying $\varphi(x)$. Once more, there is no object in stage zero. Also, two sets are the same just in case their corresponding formulas are coextensional when they are relativized to the stage before the first stage where both sets are formed. As can be seen, Gödel built L using a predicative version of Frege’s Law V restricted to those concepts which are expressible by some $\mathcal{L}_\in$ -formula. On the contrary, according to PLV, the role played by formulas is replaced by pluralities: for each one of these, a new set is introduced. Besides, by using a quasi-combinatorial notion of plurality, allowing for non-describable pluralities, PLV ensures the existence of non-describable sets. Thus, the potentialist hierarchy does not collapse into a constructive hierarchy. $\text{Collapse}^\diamond$ establishes a one-to-one mapping from pluralities onto sets, which has two immediate consequences. First, it offers an analysis of the Set of operation: to each plurality uu it assigns the set of uuu . Second, it entails a potentialist reading of the iterative conception. The collection to which every set belongs is not a plurality. If it were, $\text{Collapse}^\diamond$ would assign it a set which —on pain of contradiction— is not among this plurality. As a result, it is not the plurality of all sets.

The process of abstraction must start at some point, and as I noted in 6.2 the origin of the abstraction

process is a delicate matter. To solve the epistemological challenges, the abstraction steps cannot take place on entities that are epistemologically as dubious as mathematical objects themselves —such as concepts or other intensional entities. Abstractionism must rest on a secure basis. Now, pluralities provide this basis. They are no mysterious objects. Indeed, they are not objects beyond the objects forming them—they are just a linguistic device used to talk about these objects. Consequently, if the objects forming a plurality are in good condition, the plurality formed by them will also be in good condition. Besides, the origin of the plural hierarchy of sets is constituted by a special case: the empty plurality, i.e. no objects at all are required for starting the series of domain expansions.²³ Despite its strange metaphysical appearance, the empty plurality is not strange at all. The reason is that it is not an object and does not have a metaphysical nature. It is only a linguistic device.

PLV uses new, powerful and specific resources: modal operators and pluralities. Therefore, the next step will be to specify which logical principles govern interpretational modality, plural reasoning and the interaction between them. See (82, Ch. 3, 12) and (86; 87) for a careful exposition of the modal and plural logic governing potentialism.

6.5.2 Background Modal Logic

Interpretational modality is modelled by the system S4.2, which axiomatizes intuitionistic frames. The rationale for this choice is given by the structure of the series resulting from the iteration of abstraction attempts. First, note that the stages of this process can be modelled as possible worlds —these are represented by lower case “ w ”, adding a subscript “ w_i ” when required. I will use possible world semantics just as a heuristic device, i.e., a convenient device to express some ideas. No special metaphysical interpretation of possible worlds is advocated here —indeed, they can be thought of as mathematical objects. The domain of each possible world is the collection of sets introduced by abstraction so far. Second, the abstraction attempts are represented as an accessibility relation R defined on the possible worlds. As a result of performing a successful abstraction attempt, speakers access a larger domain of mathematical objects. Thus, $w_1 R w_2$ will mean that speakers move from w_1 to w_2 as a result of expanding their domain.

The structure of the series of abstraction attempts forms a partial ordering —thus, R is reflexive, anti-symmetric and transitive.

Partial Ordering R is reflexive, anti-symmetric and transitive.

First, speakers trivially access the stage of the process in which they are —hence, reflexivity. Second, a successful abstraction attempt does expand the domain, it does not narrow it —hence, anti-symmetry. Successful abstraction attempts do not lose objects along the way. Therefore, if speakers expand their domain as a result of an abstraction attempt, abstraction will not take them back to the previous stage. Finally, if speakers expand their domain as a result of a successful abstraction attempt, and then perform

²³If metaontological minimalism —Chapter 5, 5.3— is accepted, this entails one of the assumptions Leitgeb uses to prove that set theory, if true, is necessary [(74)]. Namely, that the empty set necessarily exists. If nothing of the world is required for its existence, it will exist in any metaphysical scenario. Therefore, if his argument is sound, metaontological minimalism entails the necessity of set theory.

a second expansion, the third domain expands not only the second, but also the former one —thus, transitivity. Now, the modal system S4 is sound and complete in Kripke frames which are symmetric and transitive. S4 is the normal modal system whose logical axioms are T —which axiomatizes reflexive frames

T

$$\Box\varphi \rightarrow \varphi$$

and 4 —which axiomatizes transitive frames

4

$$\Box\Box\varphi \rightarrow \Box\varphi.$$

Additionally, R is well-founded: every collection of possible worlds has an R -minimal element.

Well-Foundedness Every collection of possible words has a possible world w such that for any other possible world w_i in this collection, it is not the case that $w_i R w$.

Linnebo (82, p. 206) requires R to be well-founded. However, as Linnebo and Shapiro(86, p. 172) note, nothing relevant depends on this choice. The reason for accepting that R is well-founded is that the iteration process has an origin. As I said, the iteration originates in the empty plurality of objects —i.e. this process does not require any metaphysical assumptions to start. In consequence, well-foundedness appears as a reasonable feature of the hierarchy.

Another feature is convergence. Suppose that speakers have continued the abstraction process to a certain extent. At this point, they have to make a “choice.”²⁴ they have to “choose” which sets are going to be introduced to expand the domain. For example, let’s assume that $\emptyset$ and $\{\emptyset\}$ are in the domain of quantification. Collapse[◇] says that as a result of some domain expansion $\{\{\emptyset\}\}$ will be in the domain. Similarly, it says that as a result of some domain expansion $\{\emptyset, \{\emptyset\}\}$ will be in the domain of quantification too. However, Collapse[◇] does not ensure that the domain is expanded by adding both sets at the same time. Abstraction could be performed in such a way that only $\{\emptyset\}$ or only $\{\emptyset, \{\emptyset\}\}$ were introduced. Either way, this does not mean that the introduction of one of these sets is incompatible with the introduction of the other. When a sequence of abstraction attempts expands the domain until a new domain containing one of them is obtained, another round of abstraction could introduce the other. But, this means that the iteration process is convergent: any two worlds extending a former one exceed a common world.

Convergence If $w_1 R w_2$ and $w_1 R w_3$, then there is a world w_4 such that $w_2 R w_4$ and $w_3 R w_4$.

Convergence entails that “choices” concerning which objects are introduced by abstraction are not relevant. If the domain is extended with a small part of them, the others will be added at later stages. However, to simplify things I will assume the maximality thesis that any set that can be introduced in the domain by abstraction is indeed introduced.



Maximality At a given stage, any set introduced by Collapse[◇] is in the next accessible stage.

²⁴Talk of choice can be taken as a metaphor that simplifies things.

Maximality entails that R is a total order —irreflexive, asymmetric and transitive—, but does not reduce to this. Convergent frames are axiomatized by the logical principle G.

G

$$\Diamond\Box\varphi \rightarrow \Box\Diamond\varphi$$

The modal system resulting from adding G to S4 is known as S4.2. This will be the system used to model interpretational modality.

Interpretational modality varies the domain of quantification. Each successful abstraction attempt expands the domain by adding new sets. Therefore, something must be said about the regularities governing this expansionist process. Its main feature is that domains grow throughout the series of iterations, and never diminish. Successful abstraction attempts do not lose objects along the way. Thus, R is monotonic over the inclusion of domains: the domain of a given world is a subdomain of another world if the former accesses the latter.

Monotonicity If $w_1 R w_2$, then $\mathcal{D}_{w_1} \subseteq \mathcal{D}_{w_2}$,

where $\mathcal{D}_w$ is the domain of objects of the world w . This condition corresponds to the Converse Barcan Formula, which commands that domains increase along the accessibility relation:

CBF

$$\Box\forall x\varphi \rightarrow \forall x\Box\varphi.$$

However, S4.2 entails CBF. Therefore, there is no need to add it as an axiom and everything is in order.

The background modal logic of set-theoretic potentialism will be first-order S4.2, enlarged with an additional logical principle $\neg\Box\forall xz(x \neq z \rightarrow \Box(x \neq z))$ —stating the necessity of distinctness.²⁵

Definition 5.2 (MFOL) *Let $\mathcal{L}^\Diamond$ be a first-order language resulting from adding the modal operators “ $\Box$ ” and “ $\Diamond$ ” to a first-order, classical language $\mathcal{L}$. Then, the modal first-order logic (MFOL) of $\mathcal{L}^\Diamond$ is classical, first-order S4.2—with the usual rules of introduction and elimination of quantifiers, modal operators and the identity predicate—and an additional logical axiom $\neg\Box\forall xz(x \neq z \rightarrow \Box(x \neq z))$ —stating the necessity of distinctness.*

This system captures the principles governing the interpretational modality introduced in 6.2.

6.5.3 Modal Mathematics

Set theory is standardly formulated in non-modal language. On the contrary, $\mathcal{L}^\Diamond$ -sentences are about a series of domain expansions performed by successful abstraction attempts. These are modal sentences stating what is possible or what is necessarily the case throughout the series. However, in this sense, they are similar to $\mathcal{L}_\in$ -sentences. Both are about objects in a hierarchy. While the latter are about the classical hierarchy of sets, the former are about the potential hierarchy resulting from abstraction. This suggests that ZF can be interpreted as a modal theory in $\mathcal{L}^\Diamond$. Then, the potentialist set theory could be used to motivate it in the same way the iterative conception motivates ZF. To do so, it is first necessary to explain

²⁵While S4.2 proves the necessity of identity, it does not prove the necessity of distinctness

what is the relation between $\mathcal{L}_\epsilon$ -formulas and their modal counterparts formulated in $\mathcal{L}_\epsilon^\diamond$ —the result of adding the modal operators “ $\Box$ ” and “ $\Diamond$ ” to $\mathcal{L}_\epsilon$.

To recover set theory in modal idiom it is necessary to find modal sentences performing a similar role to that of classical ones. As a result of going predicative, the classical quantifiers “ $\forall$ ” and “ $\exists$ ” are insufficiently general. They only quantify over the sets formed at a given stage of the hierarchy. However, this limitation can be overcome by using the modal operators. The modalized quantifiers “ $\Box\forall$ ” and “ $\Diamond\exists$ ” are powerful enough to express absolute generality over the potential hierarchy of sets. A sentence of the form $\Box\forall\alpha\varphi$ says that for all objects at all steps of the abstraction process, φ is true. In turn, $\Diamond\exists\varphi$ says that for some object at some step of the abstraction process, φ is true. Consequently, the modalized quantifiers do the same job in the potential hierarchy as the classical quantifiers do in the classical hierarchy. Thus, let the *potentialist translation* of an $\mathcal{L}_\epsilon$ -formula φ be the $\mathcal{L}_\epsilon^\diamond$ -formula $\varphi^\diamond$ resulting from replacing each occurrence of classical quantifiers by the corresponding modalized quantifier.

Definition 5.3 (Potentialist Translation) *Let $\mathcal{L}$ be a first-order, classical language. The potentialist translation of an $\mathcal{L}$ -formula φ is the $\mathcal{L}^\diamond$ formula $\varphi^\diamond$ resulting from replacing every occurrence of “ $\exists$ ” by “ $\Diamond\exists$ ” and every occurrence of “ $\forall$ ” by “ $\Box\forall$ ”. This can be recursively defined as follows:*

$$\begin{aligned} (\tau_1 \in \tau_2)^\diamond &=_{df} \tau_1 \in \tau_2 \\ (\tau_1 = \tau_2)^\diamond &=_{df} \tau_1 = \tau_2 \\ (\neg\varphi)^\diamond &=_{df} \neg\varphi^\diamond \\ (\varphi \wedge \psi)^\diamond &=_{df} \varphi^\diamond \wedge \psi^\diamond \\ (\forall x\varphi)^\diamond &=_{df} \Box\forall x\varphi^\diamond \end{aligned}$$

Secondly, it is not only necessary to find modal copies of $\mathcal{L}_\epsilon$ -sentences. Additionally, the modal counterparts obtained by the potentialist translation must preserve logical entailment. No modal theory will be suitable to interpret ZF if any of its axioms is false, or if it cancels any of the inferences of this classical theory. Fortunately, the potentialist translation of first-order, classical formulas satisfies this requirement. Let $\mathcal{L}$ be a classical, first-order language. Also, let’s say that an $\mathcal{L}$ -formula φ is *stable* if and only if it is necessarily the case that, if φ is true, then it is necessarily true, and it is necessarily the case that if φ is false, it is necessarily false.

$$\text{Stb}_\varphi^+ \quad \Box(\varphi \rightarrow \Box\varphi)$$

$$\text{Stb}_\varphi^- \quad \Box(\neg\varphi \rightarrow \Box\neg\varphi)$$

In the next section, I will show that all predicates involved in the potentialist set theory are stable. Then, the following theorem is provable.

Theorem 5.1 (Mirroring) *Let Γ be a set of $\mathcal{L}$ -formulas, and ψ a $\mathcal{L}$ -formula. Let $\vdash^\diamond$ be deducibility in MPFO and add to this the collection of sentences stating that all predicates of $\mathcal{L}$ are stable. Then:*

$$\Gamma \vdash \psi \text{ iff } \Gamma^\diamond \vdash^\diamond \psi^\diamond,$$

where $\Gamma^\diamond = \{\varphi^\diamond / \varphi \in \Gamma\}$.

See (82, p. 74) for a proof.²⁶

6.5.4 Background Plural Logic

Let's turn to the plural logic of set-theoretic potentialism. First of all, the language $\mathcal{L}_P$ of classical plural logic extends the language of first-order classical logic by adding a countable collection of plural variables “ uu ”, “ vv ”, “ zz ”, etc., and the binary predicate “ $\prec$ ” representing the relation *to be among*: “ $x \prec uu$ ” means that x is among the uu . In addition to the usual rules of introduction and elimination of plural quantifiers, plural logic includes an impredicative comprehension scheme:

P-Comp

$$\exists uu \forall z (z \prec uu \leftrightarrow \varphi(z)),$$

where “ uu ” is not free in $\varphi(z)$. The impredicativity of P-Comp means that the plurality denoted by “ uu ” is in the domain of the plural quantifiers of $\varphi(z)$. This reading is motivated by the quasi-combinatorial notion of plurality adopted in the formulation of PLV. According to it, the case of finite pluralities of objects is generalized to any plurality whatsoever—including infinite pluralities. Consequently, when we are given a plurality of objects we are also given all their sub-pluralities. Descriptions are just a tool to highlight them.

Definition 5.4 (PL) *Let $\mathcal{L}_P$ be the language resulting from extending the language of classical, first-order logic with plural variables and the binary predicate “ $\prec$ ”. Plural logic (PL) is the system whose language is $\mathcal{L}_P$ and adds the classical introduction and elimination rules of quantifiers and the plural comprehension scheme (P-Comp) to classical, first-order logic.*

Beyond the usual axioms and rules of classical plural logic,²⁷ additional principles governing the interaction between pluralities and interpretational modality are required. Let $\mathcal{L}_P^\diamond$ be the language resulting from adding the plural vocabulary of $\mathcal{L}_P$ to $\mathcal{L}_P^\diamond$. First, it is natural to accept that pluralities are stable:²⁸

$$\text{Stb}^+ \quad \Box \forall uu \forall z (z \prec uu \rightarrow \Box (z \prec uu))$$

$$\text{Stb}^- \quad \Box \forall uu \forall z (z \not\prec uu \rightarrow \Box (z \not\prec uu)).$$

A plurality is no more than some objects. For example, the plurality of Aristotle and Hypatia is no more than Aristotle and Hypatia. Therefore, if some of these objects are removed, or other, different objects are added, the resulting plurality will be a different one. The plurality of Hypatia is different from the plurality of Aristotle and Hypatia—Aristotle is not one of Hypatia. Similarly, the plurality of Aristotle and Hypatia is different from the plurality of Aristotle, Hypatia and Ada Lovelace—Ada Lovelace is not one of Aristotle and Hypatia. Consequently, necessarily Aristotle is one of Aristotle and Hypatia. In turn, necessarily Ada Lovelace is not one of Aristotle and Hypatia.

²⁶See (86) for a proof of the intuitionist version. (134, p. 170) also proves a Mirroring theorem for his bimodal set theory.

²⁷See (83) for a brief exposition of plural logic.

²⁸Strictly speaking, the formulas “ $x \prec uu$ ” are stable. However, nothing is lost by attributing stability to pluralities.

When evaluated at some possible world w —that is, at some point of the iteration of abstraction attempts— the stability of a plurality uu ensures that the objects in the domain of w which are among the uu necessarily are among them. Also, it ensures that the objects in the domain of w which are not among the uu necessarily are not among them. However, this does not entail that additional objects in the domain of another accessible possible world w_1 —but not in the domain of w — cannot be among the uu . Loosely speaking, stability is consistent with the growth of pluralities. But, this is something that should be forbidden. On the contrary, suppose —making a small effort to imagine that set-generating processes also generate people— that Aristotle and Hypatia, but not Ada Lovelace were in the domain of w . Also, suppose that Ada Lovelace is in the domain of w_1 . It could be the case that the plurality of Aristotle and Hypatia was the same in w_1 as the plurality of Aristotle, Hypatia and Ada Lovelace. But, this contradicts the extensional nature of pluralities. Ada Lovelace is not one of Aristotle and Hypatia both in w and w_1 . In this sense, pluralities are very different from concepts and other intensional objects which are allowed to change their extension without changing their identity. What is required is a transworld principle stating that pluralities are inextensible. This is achieved by relativizing the Barcan Formula²⁹

BF

$$\forall x \Box \varphi \rightarrow \Box \forall x \varphi$$

to the appropriate formula φ as follows —the intended principle is the universal closure of the following scheme prefixed by an occurrence of the square operator:

$$\text{InExt}_\varphi \quad \forall z (\varphi(z) \rightarrow \Box \psi) \rightarrow \Box \forall z (\varphi(z) \rightarrow \psi),$$

where $\varphi(z)$ and ψ are $\mathcal{L}_P^\diamond$ -formulas. Semantically, BF does the opposite job of CBF: it axiomatizes those Kripke frames where the domains do not grow:

$$\text{If } w_1 R w_2, \text{ then } \mathcal{D}_{w_2} \subseteq \mathcal{D}_{w_1}.$$

BF ensures that the domain of a given world is a subdomain of the domain of another world when the latter accesses the former. When interpreted in some world, BF can be understood as imposing a condition on its domain —that it will not grow. Besides, monadic predicates select an “inner domain” inside the domain of possible worlds —the extension of the predicate. Therefore, if BF is relativized to a given predicate, it will state that its extension does not grow across the series of accessible possible worlds. This is precisely what InExt_φ does. Now, take those instances of this scheme where φ is the formula asserting that an object is one among the uu .

$$\text{InExt}_\prec \quad \forall uu (\forall z (z \prec uu \rightarrow \Box \psi) \rightarrow \Box \forall z (z \prec uu \rightarrow \psi)),$$

where ψ is an $\mathcal{L}_P^\diamond$ -formula. $\text{InExt}_\prec$ ensures that if z is one of uu at some accessible possible world, then it is already one of uu in the actual possible world. In conjunction, $\text{Stb}_\prec^+$, $\text{Stb}_\prec^-$ and all the instances of $\text{InExt}_\prec$ ensure that pluralities behave as desired across the sequence of successful abstraction attempts.³⁰

²⁹As CBF, BF is a scheme.

³⁰ $\text{InExt}_\prec$ is inconsistent with some instances of the potentialist translation of P-Comp

$$\text{P-Comp}^\diamond =_{df} \diamond \exists uu \Box \forall z (z \prec uu \leftrightarrow \varphi(z)).$$

$\text{P-Comp}^\diamond$ claims that for some possible plurality of objects uu , necessarily the objects among these are exactly the

The logical system governing the potential, plural hierarchy of sets is the result of combining MFOL, PL and the last three principles.³¹

Definition 5.5. (MPL) Let $\mathcal{L}_P^\diamond$ the language that results of extending $\mathcal{L}^\diamond$ with the plural vocabulary of $\mathcal{L}_P$. Modal plural logic (MPL) is the system whose language combines MFOL with PL, adding $\text{Stb}_\prec^+$, $\text{Stb}_\prec^-$ and all the instances of $\text{InExt}_\prec$ as logical axioms.

6.5.5 Potentialist Set Theory

Finally, I will introduce specific principles characterizing the notion of set. So far, all that has been specified is the logical system appropriate to reason about the potential, plural hierarchy. The axioms and rules of MPL capture the structure of the sequence of successful abstraction attempts, as well as the interaction between pluralities and the modal operators. The principles of Well-Foundedness and CBF correspond respectively to the principles of Plenitude, Origin, and Foundation of the iterative conception. However, this system is suitable for reasoning about any objects exemplifying a similar structure. Thus, additional principles governing the nature of sets must be included. Three features distinguish sets from other kinds of collections. First, sets are collections whose identity is exhausted by the identity of their members. As pluralities, they are extensional collections. But, in contrast to them, sets are objects distinct from their members. Second, because of their similarity, the potential, plural hierarchy establishes a close connection between both kinds of collection. According to $\text{Collapse}^\diamond$, every plurality uu corresponds to a set $\{uu\}$ —the set of the objects among the uu . Third, the elements of a set are prior to it. This means that, if a sequence of domain expansions meets a domain where a given set is, the elements of this set belong to previous domains of the sequence. The first and third characteristics correspond to the principles of Extensionality and Priority of the iterative conception. In what follows, I will explain how these principles are retrieved in the plural and modal context.

Let $\mathcal{L}_{P\in}^\diamond$ be the language that results from adding the binary predicate “ $\in$ ” and a singular-term-forming-operator “ $\{\}$.” When applied to plural terms “ $\alpha\alpha$ ”, this operator forms complex terms “ $\{\alpha\alpha\}$ ” denoting sets. First, the main engine of set production is $\text{Collapse}^\diamond$:

$$\Box \forall uu \Diamond \exists x (\{uu\} = x).$$

This says that no matter where the speakers are in the process of abstraction, for every plurality of objects uu so far introduced, they can abstract as to introduce the set $\{uu\}$. Thus, sets are intimately tied to objects satisfying $\varphi(z)$. Thus, take the instance resulting from the condition “ $z = z$ ”:

$$\Diamond \exists uu \Box \forall z (z \prec uu \leftrightarrow z = z).$$

Suppose the objects among the uu are in the domain of a given world w . By $\text{InExt}_\prec$, the objects among the uu will not grow when we move to accessible possible worlds. In turn, “ $z = z$ ” is satisfied by every object in every possible world. Moreover, successful abstraction attempts perform an expansion of the domain. Consequently, the condition of being self-identical is satisfied by more objects in successive stages of the hierarchy. Therefore, uu and “ $z = z$ ” cannot be coextensive, and full $\text{P-Comp}^\diamond$ is not acceptable in this context.

³¹Note that P-Comp is a theorem of PL. Therefore, by the necessitation rule, MPL entails the necessitation of P-Comp.

pluralities. Second, neither $\text{Collapse}^\diamond$ nor $\text{PLV}_I^\diamond$ guarantee that a plurality uu and its corresponding set $\{uu\}$ are coextensive. More generally, they do not guarantee that a set does not vary its elements across worlds, or that the identity of sets is given by their elements —i.e. they do not guarantee that sets are extensional collections. To solve this, the connection between sets and pluralities must be stronger than a one-to-one mapping. Let's say that a formula $\varphi(x)$ is *extensionally definite* if it is coextensive with a plurality of objects —again, the intended scheme is the universal closure of the following scheme prefixed by an occurrence of the square operator:

$$\text{ED}_\varphi \quad \exists uu \Box \forall z (z \prec uu \leftrightarrow \varphi(z)),$$

where $\varphi(z)$ is an $\mathcal{L}_\in^\diamond$ -formula. ED_φ entails Stb_φ^+ , Stb_φ^- and InExt_φ . The properties of the corresponding plurality of objects —in particular, its extensional nature— are inherited by the extension of φ . Then, the required connection between pluralities and sets is captured by the instance stating that membership is extensionally definite.

$$\text{ED}_\in \quad \Box \forall x \exists uu \Box \forall z (z \prec uu \leftrightarrow z \in x)$$

$\text{ED}_\in$ ensures that the members of a set do not vary at different possible worlds.

The extensional nature of sets is fully retrieved by adding the necessitation of the axiom of Extensionality:

$\Box \text{Extensionality}$

$$\Box \forall xy (x = y \leftrightarrow \forall z (z \in x \leftrightarrow z \in y))$$

According to this, in any given stage two sets are the same when they have the same elements. $\Box \text{Extensionality}$ together with $\text{ED}_\in$ entail that this is the case across stages. Furthermore, $\text{ED}_\in$ follows if membership is defined as follows and non-standard sets are excluded —sets that do not correspond to any plurality:

$$z \in x \equiv_{df} \exists uu (x = \{uu\} \wedge z \prec uu).$$

This definition ensures that any set has the same members as the plurality from which it is introduced. Similarly, given this definition and the previous assumption $\Box \text{Extensionality}$ and $\text{PLV}_I^\diamond$ are equivalent.

As usual, inclusion between sets is defined as follows:

$$z \subseteq x \equiv_{df} \forall y (y \in z \rightarrow y \in x).$$

As in the case of membership, I will require that inclusion is extensionally definite:

$$\text{ED}_\subseteq \quad \Box \forall x \exists uu \Box \forall z (z \prec uu \leftrightarrow z \subseteq x)$$

$\text{ED}_\subseteq$ does not follow from $\text{ED}_\in$. Nevertheless, it is motivated by the Maximality requirement on the accessibility relation R . For, suppose that a set $\{uu\}$ is introduced in the domain of a stage w_2 from a plurality of objects in the domain of another stage w_1 . At this stage all the subpluralities of uu are already available. By Maximality, the sets corresponding to these subpluralities are in the domain of w_2 —by P-Comp the domains of stages are pluralities too, just instantiate the scheme with the condition of being self-identical. But, these sets are precisely the subsets of $\{uu\}$. Therefore, at w_2 there is a plurality coextensive with $\mathcal{P}(\{uu\})$. $\text{Collapse}^\diamond$, $\text{ED}_\in$, $\Box \text{Extensionality}$, and $\text{ED}_\subseteq$ characterize the extensional nature of sets and their relation to pluralities.

It only remains to explain the priority of the elements with respect to sets. Different authors do this in slightly different ways. For instance, (10) introduces an axiom stating this in his stage theory. In turn, (134) can prove the Priority principle using a bimodal logic and Löb's Axiom. These two options come close to accepting the axiom of Foundation

$$\forall x(\exists z(z \in x) \rightarrow \exists z(z \in x \wedge \forall y(y \in x \rightarrow y \notin z)))$$

among the axioms of the iterative conception. Thus, (82) directly accepts this axiom as a natural characteristic of sets. Here, given the proximity between these solutions, the latter solution will be adopted for simplicity. That is, the axiom of Foundation is included among the principles characterizing the nature of sets.

With the addition of this axiom, the description of those principles characterizing the notion of set is finally complete. The *potentialist set theory* NS is the result of adding $\text{Collapse}^\diamond$, $\Box\text{Extensionality}$, $\text{ED}_\in$, $\text{ED}_\subseteq$, and Foundation to the background modal, plural logic MPL. Together all these principles characterize the iterative process by which successive abstraction attempts extend the domain, introducing new sets.

Definition 6 (NS) Let $\mathcal{L}_{p\in}^\diamond$ be the language that results of extending $\mathcal{L}_\in^\diamond$ with the membership predicate “ $\in$ ” and the abstraction operator “ $\{\}$.” The *potentialist set theory* (NS) is the $\mathcal{L}_\in^\diamond$ -theory whose axioms are:

$\text{Collapse}^\diamond$	$\Box\forall uu\Diamond\exists x(\{uu\} = x)$
$\text{ED}_\in$	$\Box\forall x\exists uu\Box\forall z(z \prec uu \leftrightarrow z \in x)$
$\Box\text{Extensionality}$	$\Box\forall xy(x = y \leftrightarrow \forall z(z \in x \leftrightarrow z \in y))$
$\text{ED}_\subseteq$	$\Box\forall x\exists uu\Box\forall z(z \prec uu \leftrightarrow z \subseteq x)$
Foundation	$\forall x(\exists z(z \in x) \rightarrow \exists z(z \in x \wedge \forall y(y \in x \rightarrow y \notin z)))$.

The potentialist set theory NS is powerful enough to partially interpret ZF. This theory proves the potentialist translation of S's axioms—that is, ZF without the axioms of Infinity and Replacement.

Theorem 2 Let $\vdash_{\text{MPL}}$ be provability in MPL. Then, $\text{NS} \vdash_{\text{MPL}} S^\diamond$.

See (80, pp. 217, 220) for a proof.

6.6 REFLECTION PRINCIPLES

The abstractionist approach to mathematics looks promising. Once quantifiers are read predicatively, the abstraction attempts can be iterated. This results in a sequence of domain expansions, introducing new objects. Then, the modal operators are used to express these expansions at the level of the object language. Once this is done, the modal, plural logic MPL is defined, its axioms govern the iterated process of abstraction. Also, the potentialist set theory NS, which characterizes the sets introduced by extending the domain, is defined. Finally, the potentialist theory NS partially interprets in this logic the axioms of ZF—NS proves the potentialist translation of ZF's axioms without Infinity and Replacement.

Consequently, we obtain an abstractionist justification of much of set theory. This is enough to show that abstraction principles allow us to explain serious mathematical reasoning.

However, in this way, abstraction principles are not sufficient to recover the axioms of Infinity and Replacement—the most powerful axioms of ZF. Indeed, without these two axioms, the power of set theory is equivalent to that of arithmetic. Thus, the hierarchy described by NS turns out to be the potentialist hierarchy of hereditarily finite sets.

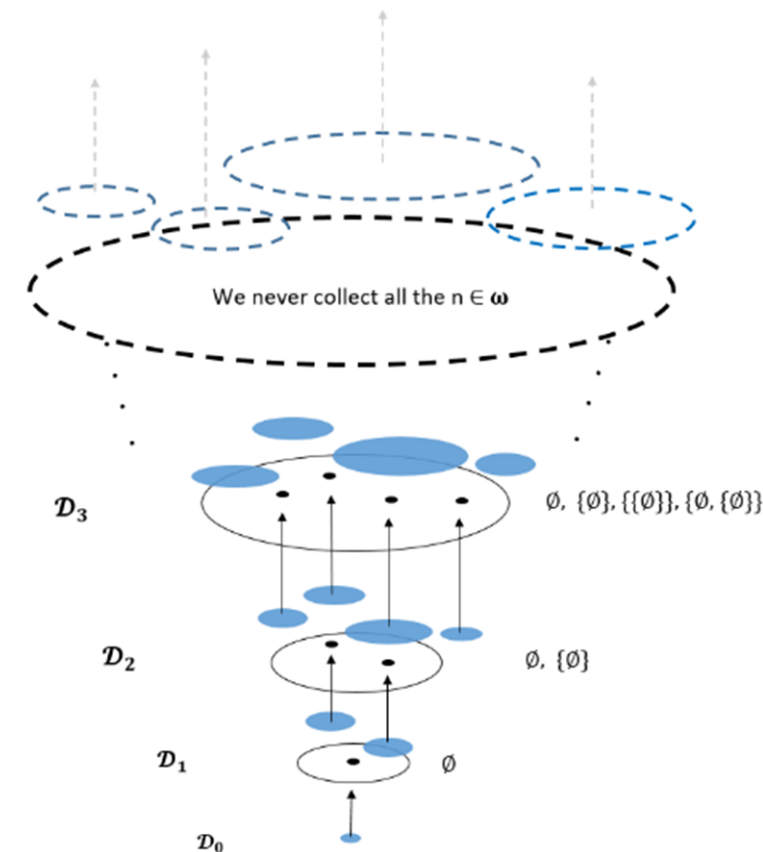


Figure 6.4: The plural hierarchy of finite sets.

Although this theory involves some mathematical complexity, abstractionism seems to deliver less than promised. Without the axiom of Infinity, we cannot even guarantee the existence of a set containing infinitely many objects. But, this fact plays a notorious role in many branches of mathematics—such as in the theory of real numbers. Similarly, without the axiom of Replacement, we cannot continue the process of set generation to the highest levels of hierarchy. Thus, without this axiom, we cannot prove that every well-ordered set is isomorphic to an ordinal (142). In consequence, we cannot introduce the ranks and reconstruct the iterative conception of sets with sets!

The limitations of potentialist theories—in general of stage theories—are well known in the literature. Usually, these “shortcomings” are remedied by recourse to principles more powerful than abstraction principles. Thus, the most common way to recover the axioms of Infinity and Replacement is to add a reflection principle³²

³²See (68, Ch. 1, 3) for a survey of reflection principles.

Reflection

$$\varphi^\diamond \rightarrow \diamond \varphi$$

to potentialist set theory. (103, p. 133), (134, p. 721), (82, p. 217) This principle says that if the potentialist translation of a formula is true, then the abstraction process reaches a stage where the classical formula is true. Thus, truths about the potentialist hierarchy are “reflected” inside it. Reflection principles of diverse order are used to recover the stronger axioms of ZF. Therefore, when added to NS, the above principle suffices to obtain the axiom of Infinity.

Theorem 3 *Let $\vdash_{MPL}$ be provability in MPL. Then, $NS + \text{Reflection} \vdash_{MPL} \text{Infinity}^\diamond$.*

See (80, pp. 219-221) for a proof.

Some authors also use reflection principles to recover the corresponding translation of the axiom of Replacement (103; 136). However, other principles have also been used to prove this axiom. For instance, (82, p. 217) uses the following principle stating that if a formula is extensionally definite and is correlated with another formula, the latter is also extensionally definite:

ED-Repl

$$Func(\psi(uv)^\diamond) \rightarrow \Box \forall xx \diamond \exists yy (\forall u \prec xx) (\exists v \prec yy) \psi^\diamond(uv)$$

Theorem 4 *Let $\vdash_{MPL}$ be provability in MPL. Then, $NS + \text{ED-Repl} \vdash_{MPL} \text{Replacement}^\diamond$.*

See (80, pp. 219-221) for a proof.

The modal version of this kind of stronger axioms fits with the bottom-up perspective introduced by the abstractionist approach. For example, reflection principles are often considered to introduce a top-down perspective in classical set theory. However, (82, p. 218) argues that the formulation of Reflection shows that this is not true in the case of potentialist set theory. However, although the use of such principles is perfectly justified for the abstractionist, they show the limits of the abstraction principles. If the latter only recover arithmetic, they will certainly fall far short. Consequently, it is worth exploring: first, whether abstraction principles alone allow us to recover the axioms of Infinity and Replacement, and second, whether there is an alternative way to the more familiar means of proving them. This is what I will do in chapter 7.

6.7 CONCLUSION

In this chapter, I explained how abstraction principles can be used to recover serious mathematical reasoning. Specifically, I explained how the iteration of predicative abstraction principles is powerful enough to recover—a large part of—Zermelo-Fraenkel set theory. First, I explained how predicative abstraction principles allow to iterate abstraction attempts. This leads to a series of domain expansions, generating the required amount of objects. Then, I explained how these expansions can be expressed at the level of the object language using modal operators. The result is a dynamic version of the abstraction process. Second, I contrasted set-theoretic potentialism and actualism, arguing that the former is in a better condition than the latter. Set-theoretic potentialism does not need to explain why some sets are formed

while others do not. In a sense, it simply accepts that all sets that can be formed, are formed. Third, I exposed how the iterated application of abstraction principles explains the two primitive operations of the iterative conception of set: the Set of and the Powerset operations. This explanation entails set-theoretic potentialism. I introduced the modal principles that govern the iteration of abstraction attempts and the expansion of domains. Also, I introduced additional principles that characterize the nature of sets. Together, all these principles form the *potentialist set theory*. Finally, I showed how the potentialist set theory interprets S —ZF without the axioms of Infinity and Replacement. However, to recover these two axioms, the theory must be complemented by additional, more powerful principles.

7 Completing the Hierarchy

Modal Abstraction Principles

7.1 INTRODUCTION

In this chapter, I will explain how —predicative— Abstraction Principles of the form

$$\forall\alpha\beta(f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta),$$

—where remember, $f(\alpha)$ and $f(\beta)$ fall out the range of the quantifier $\forall$ — can be modified so the equivalence relation $\sim$ does not apply inside the domain of classical and ontologically committed quantifiers. Indeed, abstraction will take place over modal facts, not over a collection of objects bearing the equivalence relation to each other. The result is modal, *specificantionless* abstraction: the abstraction principles introduce identity conditions which are metaphysically grounded in modal facts, devoid of existential assumptions. There are two routes to modal abstraction. The first one is inspired by (46) and (80), who suggested that the abstraction process leading to —thin— abstract objects does not require the actual existence of specifications.

There is no requirement on any particular region of spacetime since the witnessing line could be located anywhere. I later argue that the existence of a direction does not require that there must exist a line that instantiates it; it suffices that there *possibly* exists such a line. In general, abstract objects are thinner than concrete objects because they do not make demands on any particular region of spacetime. [*Italics in the original*] ((80), p. 45).

Recently, Studd (136) has outlined how this new form of abstraction could be developed. The basic idea is to take advantage of the modal operators to “see” how the equivalence $\sim$ would behave in circumstances other than the actual ones —employing possible world semantics, to go out of the actual world and see how $\sim$ behaves at other possible worlds. Take again classical predicative abstraction principles

$$\forall\alpha\beta(f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta).$$

They only involve classical quantifiers —the prefix “ $\forall\alpha\beta$ ” as well as all quantifiers present in the definition of $\sim$. Thus, they tell us to look at the domain $\mathcal{D}$ where these quantifiers are interpreted. A domain of existent objects. Then, to many pairs of objects α, β in $\mathcal{D}$ bearing the equivalence $\sim$ to each other, the abstraction principle assigns a new object $f(\alpha)$ —or what is the same $f(\beta)$ — that is not in $\mathcal{D}$. Now,

suppose the language is extended by adding modal operators “ $\Diamond$ ” and “ $\Box$ ” —for now, I will be neutral on the kind of modality these express. They allow us to speak about what is possible or necessary, that is, they allow us to talk about circumstances that go beyond what we can describe here and now, in the current context of speech. Specifically, they allow *modal quantifiers* “ $\Diamond\exists$ ” and “ $\Box\forall$ ” to be formed. Take the latter one. Replace the universal quantifier in the prefix of the classical abstraction principle with it. The result is

AP $^\Diamond$

$$\Box\forall\alpha\beta(f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta),$$

a new kind of abstraction. This modal abstraction principle works similarly to the previous one. For the sake of simplicity, I will continue using the mathematical machinery of possible world semantics. The new abstraction principle tells us to look at those objects “ $\Box\forall$ ” quantifies over. Then, as in the case of classical abstraction, to each pair α, β assigns an object $f(\alpha)$ out of its reach. Furthermore, the modalized quantifier “ $\Box\forall$ ” is read predicatively and $f(\alpha)$ does not fall under its scope. Consequently, for the same reason that predicative, classical abstraction ensures the existence of “new” objects, modal abstraction ensures that $f(\alpha)$ exists —the abstract terms $f(\alpha)$ and $f(\beta)$ are referentially successful and the scope of modal quantifiers is appropriately extended.

The advantage of these new abstraction principles is that they do not impose any requirements on the objects in $\mathcal{D}$ —the domain where classical quantifiers are interpreted. The modalized universal quantifier “ $\Box\forall$ ” allows us to see beyond the actual world, extending the abstraction process to other accessible worlds. Thus, the mere possibility that two objects α and β such that $\alpha \sim \beta$ exist suffices for the existence of another object $f(\alpha)$ —that is, $\Diamond\exists\alpha\beta(\alpha \sim \beta)$ suffices for $f(\alpha) = f(\beta)$. Employing the sufficiency operator $\Rightarrow$ introduced in the last chapter, a *first approximation* to this principle would be as follows:

$$\Diamond\exists\alpha\beta(\alpha \sim \beta) \Rightarrow f(\alpha) = f(\beta).$$

Take the case of directions. According to classical abstraction, if there are two lines a and b in $\mathcal{D}$ such that $a \parallel b$, then there is another object $d(a)$ —or what is the same, $d(b)$ — corresponding to them: their direction. On the contrary, the new abstraction principles do not require the existence of lines for the existence of directions. If it is possible for two parallel lines to exist, there will be a direction here and now corresponding to this modal fact. Again, a *first approximation* to the desired abstraction principle would be as follows:

$$\Diamond\exists xy(x \parallel y) \Rightarrow d(x) = d(y).$$

As a result of abstracting over modal facts, abstraction becomes *specificationless*. However, as written, these formulas do not express the desired facts: the second occurrence of variables is outside the scope of the quantifier. Consequently, the antecedent and the consequent are completely disconnected. While the antecedent picks a precise fact —the possibility that two objects are related to each other—, the consequent expresses no fact: it is an open formula. The aim of the text will be to seek an adequate formalization of these principles. The quantifier must affect the variables of the consequent. At the same time, the consequent must be outside the scope of the modal operator. Thus we can say that the modal fact is sufficient for the existence of the abstracts specified by α and β . Once the right formulation is found, modal abstraction principles can be employed to prove the axiom of Infinity.

The second route to modal abstraction analyzes how modal resources —operators and frames— can be employed to shape *intensional* equivalence relations. This idea can be made precise by employing variable domain semantics together with a background *positive* free logic —allowing to talk about “non-existent” objects. Fix a constant outer domain $\mathcal{D}$ interpreting non-modal quantifiers “ $\exists$ ” and “ $\forall$ ”, and allow inner domains $\mathcal{D}_w$ —that is, the domain of existent objects in w — to vary. Because the background free logic is positive, objects not in $\mathcal{D}_w$ have properties and bear relations to other objects in $\mathcal{D}$ —but out of $\mathcal{D}_w$ — in w . Now, let a unity relation $\sim^\diamond$ to be defined over $\mathcal{D}$ and incorporate it inside abstraction principles in the obvious way

$$\forall \alpha \beta (f(\alpha) = f(\beta) \leftrightarrow \alpha \sim^\diamond \beta).$$

The equivalence $\sim^\diamond$ can hold transworldly between non-coexistent objects. Again, this results in specificationless abstraction —however, note that because of the background positive free logic, this route diverges both epistemologically and metaphysically from the former. For any two objects α and β —existing or not in w_0 —, the abstraction principle ensures the existence —in w_0 — of another object $f(\alpha)$ out of $\mathcal{D}$. Nevertheless, I will not develop here the details concerning this second form of modal abstraction. In particular, I will not analyze how these abstraction principles are affected by the properties of different accessibility relations R , as well as by the growth of domains under the Barcan formula and its converse.

In this chapter, I will only develop the former kind of modal abstraction. The plan goes as follows. In Section 2 I will discuss a variety of technical matters. First, I will explain the background modal and free logic of modal abstraction. Then, I will spell out the details of how modal abstraction allows speakers to “look at those objects “ $\Box\forall$ ” quantifies over” —this is not trivial. In doing this it will become clear why modal abstraction must be predicative. In Section 3, I will modify the inheritance principles turning them into a modal component of modal abstraction. Once this is done, abstraction can be used again to extend the language and secure reference to additional objects. In Section 4, I will use abstraction principles to prove the axiom of infinity. Finally, the requirements and difficulties of generalizing the proposal will be assessed.

7.2 VARIETIES OF MODAL ABSTRACTION

In this section, I will analyze a variety of technical details concerning the correct form of modal abstraction principles. This is not as obvious as it could seem and needs elaboration. First, I will describe the background modal and free logic required to make sense of modal abstraction. To obtain *specificationless* abstraction the former will be a variable domain modal logic. Furthermore, because I am concerned with the explanation of mathematical reference, the free logic will be a negative one. Once this is done, the logical form of modal abstraction principles must be spelt out. I will consider three options. Only the last one does the job. Finally, it will be clear why modal abstraction principles must be predicative.



7.2.1 Background Modal Logic

From a semantic point of view, the background logic must be normal modal logic with *variable possible world semantics*. Variable possible world semantics is needed to make sense of *specificationless* abstraction. Otherwise, the proposal would not be able to make sense of the possibility of some objects α and β being related by the unity relation $\sim$, unless α and β exist in the actual world. Therefore, a fixed domain of objects will collapse modal abstraction principles into the classical ones.

To express this and other ideas I will use the idiom of possible worlds. However, it is important to keep in mind that they are only a useful and convenient communicative device. Especially, no particular metaphysical interpretation is attributed to them. I will denote possible worlds by lowercase “ w ”, adding a subindex “ w_i ” when dealing with several of them. Besides, the domain of a possible world w will be denoted by “ $\mathcal{D}_w$ ”, and R will be the accessibility relation. The key feature of variable world semantics can be expressed as follows: it allows $D_{w_i} \neq D_{w_j}$, if $w_i \neq w_j$. I.e. the domains of different worlds can be different. For the semantics of normal modal logic, see Appendix 7.5.

7.2.2 Background Free Logic

Once a variable domain modal logic is chosen as background logic, the free logic must be specified. (82) employs *negative* free logic —combined with S4.2— in the background logic of potentialist set theory. As a result, those sentences containing terms that refer to objects $a \notin D_{w_i}$ will be immediately false.¹ For instance, “Hamlet is Danish” is false according to negative free logic because “Hamlet” does not refer to any person —indeed, it does not refer at all. Only sentences that refer to objects in D_{w_i} will come out as true in w_i . Other authors, such as (103, p. 116), have proposed *positive* free logic as the correct background logic for dealing with domain expansions. Positive free logic is essential for his project because he proposes a trans-world criterion of identity, telling us when a set a denoted in a world w_i is the same set as b denoted in the world w_j —such that $w_i R w_j$.² To compare both a and b in w_i it is required to say some things about a in this context even if $a \notin D_{w_i}$. Hence, positive free logic seems to be the correct choice.

However, I do not believe that positive free logic is the best option when it comes to explanations of mathematical reference. Payne justifies his choice by saying that, according to his view, positive free logic does not entail the existence of a discourse committed to non-existent objects (103, p. 124) —which would be highly controversial: “The claim that it is possible to refer to objects which do not exist should no longer be so mysterious. It is simply the claim that a just does not fall under the present domain of quantification.” As an example, he considers the case of contextual restrictions of the scope of

¹ Assume that E is a predicate expressing existence. Then the following is a logical truth in negative free logic:

$$\varphi(\alpha_1 \dots \alpha_n) \rightarrow E\alpha_1 \wedge \dots \wedge E\alpha_n.$$

²(103, p. 118) proposes the following modal version of Frege’s Basic Law V to compare cardinalities in different worlds:

BLV $\downarrow$

$$\Box\Box\forall F\forall G(\varepsilon(F) = \downarrow \varepsilon(G) \leftrightarrow \forall x(Fx \leftrightarrow \downarrow \exists y(y = x \wedge Gy))).$$

The operator $\downarrow$ is known as the *backtracking* operator. I will say more about its function below.

quantifiers (103, p. 116): if speakers are talking about the blue tables in the room, then it is reasonable that “everything is a table” comes as true in this context. Consequently, the Moon—which is not a table—does not exist inside such narrow limits. It still exists—for physicians and owls—, but it does not in the present context. It seems to me that this reasoning is correct for cases of *contextual restriction*. Reference to some objects “living outside” the restricted context—as the Moon—has been fixed in advance. When the restriction is performed, speakers simply ignore these objects. But, in doing this they do not lose their ability to refer to them—their memory or linguistic abilities are not erased by the restriction.

On the contrary, things are different when reference itself—not a context restriction—is explained as the result of shifting the interpretation of the quantifiers. According to abstractionism, this is the case of reference to mathematical objects. Until these are fixed as referents of some term τ —which does not refer in the preset context—, nothing can be said using τ . Remember that “term” is a *semantic* category. Therefore, I doubt even that τ can be qualified as such until its referent is fixed—at most, it would be a term by stipulation. For instance, suppose that reference was fixed only to elements in $V_1 = \{\emptyset\}$ —that is, speakers only refer to $\emptyset$. If reference to $\{\emptyset\}$ is only secured through a re-interpretation of the quantifiers which expands the speakers’ domain, they will not have any resources to refer to $\{\emptyset\}$ until such an expansion will be performed. Because of this, I believe that *negative* or *neutral free logic* is preferable in the context of meta-semantic explanations of mathematical language. Specifically, I will employ neutral free logic. According to this, it is neither true nor false in w that objects not in $\mathcal{D}_w$ have properties or bear relations to other objects. More generally, no sentence $\varphi(\alpha_1 \dots \alpha_n)$ is true or false in w if $\alpha_i \notin \mathcal{D}_w$, $1 \leq i \leq n$. Thus, classical quantifiers “ $\forall$ ” and “ $\exists$ ” are interpreted in the inner domains $\mathcal{D}_w$. The choice of neutral free logic affects the shape of modal abstraction principles: there is some risk of collapsing them into an inoperable tool devoid of a definite truth value. As I will show, they are safe as long as the unity relation $\sim$ is inside the scope of modal operators.

7.2.3 Varieties of Modal Abstraction

It is not obvious what the logical form of a modal abstraction principle should be like. The alleged benefit of these principles is that they allow for *specificationless* abstraction. An abstraction step performed at a world w does not take place over objects α and β in $\mathcal{D}_w$ bearing the unity relation $\sim$ to each other. It does not happen inside the inner domains where the classical and ontologically committed quantifiers “ $\exists$ ” and “ $\forall$ ” are interpreted. On the contrary, a modal abstraction step only requires that at a world w it is *possible* for some objects α and β to be related by $\sim$. As a result, modal abstraction takes place over modal facts, not over a collection of existing objects—objects in $\mathcal{D}_w$. The key idea behind specificationless abstraction is that speakers can take advantage of the modalized quantifier “ $\Box\forall$ ” to “go outside” of $\mathcal{D}_w$ —to use a suggestive metaphor—, take a look of what happens with objects in $\mathcal{D}_{w_j}$ for all accessible worlds w_j —worlds such that wRw_j —, and come back to w with an enriched stock of objects—the abstracts $f(\alpha)$, $f(\beta)$, etc.—, resulting from their “travel experience”.



As modal abstraction principles were stated,

$$\Box\forall\alpha\beta(f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta),$$

they will not do the job. Assume this principle is evaluated in some possible world w_0 . The reason is that the operator “ $\Box$ ” drags the formula

$$\psi \equiv \forall \alpha \beta (f(\alpha) = f(\beta) \leftrightarrow \alpha \sim \beta),$$

which features inside its scope, across the net of worlds accessible from w_0 . However, $AP^\diamond$ does not provide any resources to reverse this effect. Speakers will not be able to return to w_0 from their journey. $AP^\diamond$ says: at every w_j , such that $w_0 R w_j$, it is the case that ψ . Thus, it just says that ψ is true at all accessible possible worlds. This has two problems. First, surely speakers are “going out” of $\mathcal{D}_{w_0}$. However, they are not surveying—to abuse the metaphor—the whole range of “ $\Box\forall$ ” from w_0 . Remember that the idea is to consider from w_0 what is possible concerning the unity relation $\sim$, and then to perform an abstraction step over these possible facts. More specifically, what is required is an abstraction principle yielding a correct formalization of the following statement

$$\diamond \exists \alpha \beta (\alpha \sim \beta) \Rightarrow f(\alpha) = f(\beta).$$

The quantifier must affect the variables of the consequent, and the consequent must be outside the scope of the modal operator. Thus we can say that the modal fact is sufficient for the existence of the abstracts specified by α and β .

Second, $AP^\diamond$ is inconsistent when added to NS. Take the instance of $AP^\diamond$ corresponding to Plural Law V (PLV):

PLV $^\diamond$

$$\Box \forall u u s s (\{u u\} = \{s s\} \leftrightarrow \forall x (x \prec u u \leftrightarrow x \prec s s)).$$

Let rr be the Russellian plurality in w_0 —the world where PLV $^\diamond$ is evaluated—satisfying

$$x \prec rr \leftrightarrow x \notin x, \text{ for all } x \in \mathcal{D}_{w_0}.$$

The plurality rr is a subplurality of $\mathcal{D}_{w_0}$ by plural comprehension (P-Comp). Then, *Collapse* entails that the set $\{rr\}$ possibly exists—for some w_j such that $w_0 R w_j$, $\{rr\} \in \mathcal{D}_{w_j}$. But, on pain of contradiction, $\{rr\}$ does not exist in w_0 —i.e. $\{rr\} \notin \mathcal{D}_{w_0}$. For, suppose the contrary and proceed by *reductio*. The characteristic property of rr entails that the equivalence

$$\{rr\} \prec rr \leftrightarrow \{rr\} \notin \{rr\}$$

holds in w_0 . Also, by the definition of $\in$, the following equivalence

$$\{rr\} \in \{rr\} \leftrightarrow \{rr\} \prec rr$$

holds in w_0 . But, from these two equivalences, the following obtains in w_0 :

$$\{rr\} \in \{rr\} \leftrightarrow \{rr\} \notin \{rr\}.$$

A contradiction immediately arises: therefore, $\{rr\} \notin \mathcal{D}_{w_0}$. However, PLV $^\diamond$ entails $\{rr\} = \{rr\}$ in w_0 , and thus $\{rr\} \in \mathcal{D}_{w_0}$. So, a contradiction ensues after all.

One solution consists in using constant domain semantics. Set $\mathcal{D}_{w_i} = \mathcal{D}_{w_j}$, for all worlds w_i, w_j . Then, at w_0 $AP^\diamond$ says that ψ holds at any w_j such that $w_0 R w_j$. Now, suppose that $\alpha \sim \beta$ at some of these w_j . This and $AP^\diamond$ entail that $f(\alpha) = f(\beta)$ in w_j . But, because the use of neutral free logic, $f(\alpha) \in \mathcal{D}_{w_j}$ —that is, $f(\alpha)$ exists in w_j . And, because $\mathcal{D}_{w_j} = \mathcal{D}_{w_0}$, $f(\alpha) \in \mathcal{D}_{w_0}$ —that is, $f(\alpha)$ also exists in w_0 . Indeed, all abstracts will exist in all worlds. This partially solves the first of the aforementioned problems. If identity is necessary, $f(\alpha) = f(\beta)$ would be true in a different world other than w_0 . However, this solution is unsatisfactory for two reasons. First, constant domain semantics invalidates specificationless abstraction. As I argued, this is the key idea behind modal abstraction principles: to perform abstraction on mere possible facts, without requiring the actual existence of the specifications. If the domain is constant, even if facts about $\sim$ vary from world to world—and $\sim$ is not stable—, specifications α and β would exist in the actual context. Second, this solution does not solve the second problem: the principle is still inconsistent with NS. More generally, this “solution” just varies the semantics, but not the formalization of $AP^\diamond$.

Something similar happens if the Brouwerian axiom

B

$$\varphi \rightarrow \diamond \Box \varphi$$

is added to the background modal logic. This axiom axiomatizes symmetric frames: if a world accesses another, the latter also accesses the former. As a consequence of symmetry, accessible worlds have the same domain: if $w_i R w_j$, then $\mathcal{D}_{w_i} = \mathcal{D}_{w_j}$. Therefore, the situation is virtually the same as before. The accessibility relation does not discern between domains of objects, and this allows the same reasoning used in the previous case. Thus, this is unsatisfactory for the same reasons it was. First, Specificationless abstraction collapses into classical abstraction. If the accessible domains are the same, the modalized quantifier “ $\Box \forall$ ” ranges over the same objects as the classical universal quantifier. Second, $AP^\diamond$ remains inconsistent with NS. The problem is that B, as constant domain semantics, modifies the semantics of modal abstraction, but not the form of the abstraction principle.

Instead of modifying or adding restrictions to the background modal logic, what is required is an alternative formalization of $AP^\diamond$. In what follows I will explain what this formalization looks like. First, I will use the *backtracking* operator $\downarrow$ to go back to w_0 —the world where $AP^\diamond$ is applied. That is, to “return” from the travel across the net of accessible worlds to the origin. As desired, the operator allows speakers to take into account the whole range of the modalized quantifier “ $\Box \forall$ ” from there. Second, I will extend the language of the modal logic by adding two additional modal operators $\blacksquare$ and $\blacklozenge$. This is required to avoid contradiction when the appropriate version of $AP^\diamond$ is combined with NS. As a result, $AP^\diamond$ turns out to be a *predicative* modal abstraction principle mimicking predicative classical abstraction.

7.2.3.1 Option 1: Scope Operators

The *backtracking operator* $\downarrow$ is an operator introduced by (59) as a device to make travel across worlds more flexible. The operator can be understood both semantically or syntactically. From a semantic standpoint, $\downarrow$ makes the formula φ in its scope to be evaluated at the possible world where the last modal operator affecting $\downarrow \varphi$ was evaluated. Metaphorically speaking, when evaluating a formula and we find

the expression $\downarrow \varphi$, the operator $\downarrow$ gives instructions³ to travel back to the last visited world and to evaluate φ at this world. It is useful to compare the backtracking operator with the actuality operator $@$. The actuality operator causes a formula φ in its scope to be evaluated at a privileged possible world, the *actual* world. In turn, the backtracking operator does not identify a fixed, salient possible world where φ must be evaluated independently of the sentential context where it features. This operator identifies such a world as a function of the sentential context. Alternatively, these operators can be interpreted purely syntactically. From a syntactic point of view, the actuality operator exempts φ from the scope of all modal operators. On the contrary, the backtracking operator exempts φ from the scope of the innermost modal operator affecting $\downarrow \varphi$. Thus, the following sentences come out as equivalent:

$$\varphi \equiv \Box \downarrow \varphi \equiv \Box \Box \downarrow \downarrow \varphi \equiv \Box \downarrow \Box \downarrow \varphi.$$

According to the syntactic reading, the status of $\downarrow$ would be akin to that of parentheses.

The backtracking operator not only applies to formulas. It can also be applied to other expressions, such as names. Indeed, these offer an example of how the operator is used in natural language. The function of the backtracking operator is revealed in those contexts where speakers make modal assertions which require varying the circumstances of evaluation while keeping fixed the extension of some expressions. For example, take the case of temporal statements about objects in the present context. Think about the statement “it does not matter how much money I earn, I will still be poor”. This sentence must be evaluated in some future circumstances different from the actual ones —circumstances in which foreseeable I will earn more money than now. However, the “I” keeps fixed its reference —me— even if circumstances are different. I am talking about what would happen to me —I will still be poor. Recently, Payne (2013) has used the operator to provide abstractionism with trans-world abstraction principles — which offer trans-world criteria of identity for abstracts.⁴ Similarly, $\downarrow$ can be used to solve the first of the aforementioned problems: to explain how speakers can take stock of the scope of the modalized quantifier “ $\Box \forall$ ” from the point of view of a particular world as well as to evaluate the identity in $AP^\Diamond$ at a different world from the one in which the unity relation is evaluated.

Let be the following statement a first approximation to the desired formalization of $AP^\Diamond$:

³I take the metaphor from (104, p. 238).

⁴Working in the context of providing sets with abstraction principles, Payne proposes the following modal version of Frege’s Law V: (103, p. 118)

BLV $\downarrow$

$$\Box \forall F \forall G (\varepsilon(F) = \downarrow \varepsilon(G) \leftrightarrow \forall x (Fx \leftrightarrow \downarrow \exists y (y = x \wedge Gy))),$$

where F and G range over concepts, ε is an singular-term-forming operator and $\varepsilon(X)$ is an abstract term denoting the extension of X —at some possible world. Here, the backtracking operator is applied both to a term and to a formula. Suppose the formula

$$\Box \forall F \forall G (\varepsilon(F) = \downarrow \varepsilon(G) \leftrightarrow \forall x (Fx \leftrightarrow \downarrow \exists y (y = x \wedge Gy)))$$

—BLV $\downarrow$ without the first square— is evaluated at w_0 . Then, the backtracking operator is applied to the abstract term $\varepsilon(G)$. This allows to keep fixed the object it denotes in w_0 —the extension of G in w_0 — when the formula inside the scope of the square is evaluated at some w_1 such that $w_0 R w_1$. Reference is thus preserved across worlds. Similarly, the unity relation is modified to hold between objects inhabiting different worlds. Therefore, BLV $\downarrow$ reads: the object denoted by $\varepsilon(G)$ in w_0 is the same as the object denoted by $\varepsilon(F)$ in w_1 if and only if for all objects $x \in \mathcal{D}_{w_1}$: Fx in w_1 if and only if there is an object $z \in \mathcal{D}_{w_0}$ such that $x = z$ and Gz in w_0 .

$AP_1^\diamond$

$$\Box \forall \alpha \beta (\downarrow [f(\alpha) = f(\beta)] \leftrightarrow \alpha \sim \beta).$$

Suppose $AP_1^\diamond$ is evaluated in w_0 . The modalized quantifier “ $\Box \forall$ ” ranges over all the $\mathcal{D}_{w_i}$ such that w_i is accessible from w_0 . Thus, when two specifications $\alpha, \beta \in \mathcal{D}_{w_i}$ are related by the unity relation $\sim$ in w_i , the previous statement ensures that the identity $f(\alpha) = f(\beta)$ holds. Moreover, as the identity is inside the scope of $\downarrow$ —we have $\downarrow [f(\alpha) = f(\beta)]$ —it holds in w_0 : the world where “ $\Box$ ” is evaluated. As a result of this and the background negative free logic, $f(\alpha) \in \mathcal{D}_{w_0}$ —the abstracts $f(\alpha), f(\beta)$, etc. exist in w_0 . Modal abstraction guarantees the existence of abstracts just by taking stock of what is possible about $\sim$: *abstraction grounds existence in modal facts*.

$AP_1^\diamond$ needs further development, and the idea of grounding actual existence on mere modal facts is interesting enough—this form of abstraction will be subject to the kind of modality we are working with. However, $AP_1^\diamond$ does not solve the second of the aforementioned problems: it is inconsistent when added to the potentialist set theory NS. Remember that the background modal logic of the latter is S4.2, and its accessibility relation is reflexive. Thus, $AP_1^\diamond$ ranges over objects in the domain of the world where it is evaluated. Now, consider the following instance of $AP_1^\diamond$ corresponding to PLV:

$PLV_1^\diamond$

$$\Box \forall uvv (\downarrow [\{uu\} = \{vv\}] \leftrightarrow \forall z (z \prec uu \leftrightarrow z \prec vv)).$$

Suppose that it is evaluated in w_0 and, again, let rr be the Russellian plurality of w_0 . $AP_1^\diamond$ entails that $\{rr\} \in \mathcal{D}_{w_0}$ — $\{rr\}$ exists in w_0 . But, by the same reasoning as in Section 2.3 this leads to a contradiction.⁵ As Studd (2023) points out, the problem with $PLV_1^\diamond$ is that it turns out to be an *impredicative* abstraction principle as soon as T

$$\Box \varphi \rightarrow \varphi$$

—which axiomatizes reflexive frames—is assumed in the background logic. The reason is that the modalized quantifier “ $\Box \forall$ ” ranges over objects in $\mathcal{D}_{w_0}$. The objective of developing modal abstraction is to use it to prove the axiom of Infinity when combined with NS. Therefore, I will not analyze here what happens when T is removed from the background modal logic of abstraction. Instead, I will add a second modification to $AP^\diamond$, turning it into a *predicative* abstraction principle.

7.2.3.2 Option 2: Predicative Modal Abstraction

What does it take for modal abstraction to succeed when it is combined with NS, and S4.2 is assumed as the background logic? $PLV_1^\diamond$ must be transformed into a predicative abstraction principle. This will ensure that the identity $\{uu\} = \{ss\}$ holds in a world other than the world where $PLV_1^\diamond$ is evaluated. Furthermore, such a world must be outside the range of the modal operators $\Box$ and $\Diamond$. Let W be the collection of possible worlds used to provide $\Box$ and $\Diamond$ with truth conditions. The reason is that Russell’s paradox can be replicated as soon as the Russellian set $\{rr\}$ of any world $w \in W$ is considered. Suppose that $PLV_1^\diamond$ is evaluated at w_0 . Suppose the left side of $PLV_1^\diamond$ —that is, identities of the form $\{uu\} = \{vv\}$ —

⁵Recall that abstraction principles do not settle mixed identities between “old” and “new” abstracts. However, according to the piecemeal solution to the Caesar Problem as well as to NS, old and new abstracts are identified by extensionality.

is evaluated at some $w_i \in W$ other than w_0 . Then, the Russellian plurality of sets in w_i is enough to generate a contradiction in the same way as in Section 2.3. Abstracts must not only be outside of D_{w_0} , they must also be out of reach of “ $\Box\forall$ ”.

A predicative version of $AP_1^\diamond$ must perform a twofold job. First, it must take stock of all equivalences $\alpha \sim \beta$ which are the case at all w_k accessible from w_0 . Second, it must assign an abstract $f(\alpha)$ to each of these equivalences, locating it in world⁶ $w_\omega \notin W$. Thus, the predicative version of $PLV_1^\diamond$ will say—at a world w_0 —something along the following lines:

- (i) there is a possible world $w_\omega \notin W$
- (ii) such that for all w_i accessible from w_0 , and for all $\alpha, \beta \in \mathcal{D}_{w_i}$: $\alpha \sim \beta$ is true in w_i if and only if
- (iii) the abstracts $f(\alpha), f(\beta), \dots$ exist in w_ω — $f(\alpha), f(\beta), \dots \in \mathcal{D}_{w_\omega}$.

As they were stated, (i)-(iii) belong to the meta-language —these sentences indulge in possible-world talking. Thus, some expressive resources are required to replicate them at the object-language level. Let $\mathcal{L}^\diamond$ be the language of the background logic, and let $\mathcal{L}_\downarrow^\diamond$ be the language that results from adding $\downarrow$ to the background modal language. To express (i), I will extend this language by adding two modal operators $\blacksquare$ and $\blacklozenge$. These operators go beyond the space covered by the operators $\Box$ and $\Diamond$, expressing new possibilities. In this case, the possibility that all abstracts exist in the same world. Consequently, the modal abstraction becomes predicative. Classical, predicative abstraction expands the domain where the quantifiers $\forall$ and $\exists$ are interpreted. Now, modal abstraction performs a similar expansion affecting the modalized quantifiers $\Box\forall$ and $\Diamond\exists$. The result of adding these operators is a bimodal language $\mathcal{L}_\downarrow^{\blacklozenge}$. Strictly speaking, this is not a two-sorted language. These black operators combine with the other expressions in $\mathcal{L}_\downarrow^\diamond$ generating well-formed formulas. Nevertheless, the resulting picture of abstraction has—for the moment—a similar flavour to the two-sorted framework used in Linnebo (2018) as a temporary device to introduce predicative abstraction. Similarly, the bimodal language is just a temporary device used to introduce modal abstraction.

The possibility represented in (i) by w_ω will be covered by $\blacklozenge$ in the object language. Next, (ii) requires to obviate the modality in (i), and to consider what is possible at w_0 . What is required is to go back to a previously visited world. Then, the backtracking operator $\downarrow$ works perfectly here. It exempts the expressions inside of its scope from the scope of the innermost modal operator affecting them: in the present case, the black diamond in (i).

Finally, (iii) we have to travel again to w_ω . But, neither the black modal operators nor the backtracking operator perform this function. Therefore, we need additional resources to do this job. At this point I will introduce another operator $\uparrow$ —see Appendix 6.B for how its semantics could be developed. The idea is that the *unbacktracking operator* $\uparrow$ would reverse the previous actions of $\downarrow$. From a semantic point of view, the unbacktracking leads us back to the last world where $\downarrow$ was applied. And it does so regardless of what other worlds we have visited due to the application of other modal operators. To illustrate its

⁶I call this world “ w_ω ” because the reason to introduce modal abstraction is to prove the modal translation of the axiom of Infinity.

action, consider the following equivalent formulas:

$$\diamond \downarrow \uparrow \varphi \equiv \diamond \varphi \equiv \diamond \downarrow \square \square \uparrow \varphi.$$

The unbacktracking operator allows us to express (iii). If the backtracking operator tells us to ignore the $\diamond$, going back to w_0 , and the square “ $\square$ ” takes us ahead to another accessible w_i , the new $\uparrow$ should be able to reset all these operations placing us in the original w_ω .

Let $\mathcal{L}_{\downarrow \uparrow}^\diamond$ denote the language resulting from adding $\uparrow$ to $\mathcal{L}_{\downarrow}^\diamond$. If this richer language is in good condition, then the following would be a formalization of (i)-(iii), thus a formalization of the predicative version of $\text{PLV}_1^\diamond$:

$\text{PLV}^\diamond$

$$\diamond \downarrow \square \forall uvv (\uparrow [\{uu\} = \{vv\}] \leftrightarrow \forall z (z \prec uu \leftrightarrow z \prec vv)).$$

—see Appendix 6.B for a proof of the relative consistency of $\text{AP}^\diamond$. This principle states: it is *black-possible*, that *from now on* it is *always* the case that for any objects xx and any objects yy , the set of xx will be equal to the set of yy if and only if the xx are the same objects as the yy . Therefore, it has the following structure:

It is *black possible* ($\diamond$) that *from now on* ($\downarrow$) it is *always* ($\square$) the case that: ψ will occur ($\uparrow$) iff γ .

This generates a whole family of abstraction principles of the form:

$\text{AP}^\diamond$

$$\diamond \downarrow \square \forall \alpha \beta (\uparrow [f(\alpha) = f(\beta)] \leftrightarrow \alpha \sim \beta).$$

This principle reads: it is *black-possible*, that *from now on* it is *always* the case that for any specification α and any specification β , the abstract $f(\alpha)$ will be equal to the abstract $f(\beta)$ if and only if $\alpha \sim \beta$. However, recovering this kind of modal abstraction principles in natural language is cumbersome. Alternatively, remember that AP can be factorized into the simpler principles *Collapse* and AP_I . So, let's do the same here. Furthermore, instead of the abstraction operator “ $\{\}$ ” I will introduce a set-predicate “ $\text{Set}(xx, y)$ ” —this means that y is the set of the xx .⁷ Then, $\text{PLV}^\diamond$ can be “factorized” in the following two parts:

Collapse $^\diamond$

$$\diamond \downarrow \square \forall uu \uparrow \exists y \text{Set}(uu, x)$$

Extensionality $^\diamond$

$$\blacksquare \forall xz \downarrow \square \forall uvv (\text{Set}(uu, x) \wedge \text{Set}(vv, z) \rightarrow (z = z \leftrightarrow \forall y (y \prec uu \leftrightarrow y \prec vv)))$$

Each of these parts performs a different role. For instance, *Collapse* $^\diamond$ reads: it is *black-possible* that *from now on* it is *always* the case that for any objects xx it will exist the set of these objects. The structure of this sentence is:

It is *black possible* ($\diamond$) that *from now on* ($\downarrow$) it is *always* ($\square$) the case that φ will occur.

⁷The choice of the abstraction operator or the set-predicate is irrelevant, as both are interdefinable. Here I will work with the set-predicate to facilitate the reading of the principles.

Thus, Collapse $\blacklozenge$ has a double task. The modalized first-order quantifier $\downarrow \Box \forall$ encloses all the objects that will come into existence from now on. Then, the black-diamond $\blacklozenge$ and the existential quantifier $\exists$ collect all of them at some future point. Thus, we get an infinity of objects.

Two comments are in order regarding AP $\blacklozenge$ —equally, Collapse $\blacklozenge$ and Extensionality $\blacksquare$. First, note that all terms are quantified and that all quantifiers are inside the scope of a modal operator. Thus, the choice of negative or neutral free logic as the background logic does not cause AP $\blacklozenge$ to be devoid of truth value by the mere fact that abstracts $f(\alpha)$, $f(\beta)$, etc or specifications α , β ,... do not exist in the world where AP $\blacklozenge$ is evaluated. Second, I want to raise a worry concerning the metaphysical commitments of the principle. Specificationless, modal abstraction aims to ground identities between abstracts on possible facts concerning the unity relation $\sim$. Using the approximate —but ill-formed— formulation of the idea

$$\blacklozenge \exists \alpha \beta (\alpha \sim \beta) \Rightarrow f(\alpha) = f(\beta).$$

According to the usual semantics for “ $\blacklozenge$ ”, $\exists \alpha \beta (\alpha \sim \beta)$ is true just in case at some accessible world there is an α and a β such that $\alpha \sim \beta$. This operator does not make a finer classification among the worlds where there are objects related by the unity relation $\sim$. Thus, if modal abstraction takes place over the mere possibility that $\sim$ relates two objects between them, then it would collapse into the introduction of a unique identity: the identity corresponding to the modal fact expressed by $\blacklozenge \exists \alpha \beta (\alpha \sim \beta)$. On the contrary, AP $\blacklozenge$ takes into account all the domains $\mathcal{D}_w$ and makes the required distinctions to introduce an infinity of abstracts in w_ω . Then, the following question arises: is modal abstraction metaphysically committed to mere possible objects (*possibilia*)?

7.3 REFURBISHING ABSTRACTION

Once the abstraction principle AP $\blacklozenge$ is in place, the *Inheritance* principle must also be accommodated to the new framework. Remember how a whole language committed to directions was introduced using a successful abstraction attempt —see Chapter 5, 5.2.0.1 for details. The abstraction principle

Dir

$$d(l_1) = d(l_2) \leftrightarrow l_1 \parallel l_2$$

says that two lines l_1 and l_2 specify the same direction $d(l_1)$ if and only they are parallel to each other. Thus, this principle guarantees the existence of directions $d(l_i)$ just in case there is a line l_i entering the parallelism relation $\parallel$ — $l_i \parallel l_i$. Now, using Dir the base language of abstraction $\mathcal{L}_1$ —a language that is only committed to lines— can be extended by adding new expressions about directions. Consider all those formulas $\varphi(\alpha_1 \dots \alpha_n)$ satisfying

Congruence

$$l_1 \parallel l'_1 \wedge \dots \wedge l_n \parallel l'_n \rightarrow (\varphi(l_1 \dots l_n) \leftrightarrow \varphi(l'_1 \dots l'_n))$$

and



$$\varphi(l_1 \dots l_n) \rightarrow l_1 \parallel l_1 \wedge \dots \wedge l_n \parallel l_n.$$

That is, take all those formulas $\varphi(l_1 \dots l_n)$ such that: (i) the relation of parallelism $\parallel$ is a congruence with respect $\varphi(l_1 \dots l_n)$ and (ii) their field is a subset of the field of $\parallel$ —they are only satisfied by lines. Then, for each one of these formulas a new n -ary predicate $\varphi^*(d(l_1) \dots d(l_n))$ can be added satisfying the following inheritance principle

Inheritance

$$\varphi^*(d(l_1) \dots d(l_n)) \leftrightarrow \varphi(l_1 \dots l_n).$$

For instance, the relation of orthogonality holding on directions $\perp^*$ is “inherited” in this way from the corresponding orthogonality relation between lines $\perp$. Parallelism is a congruence concerning $\perp$ —thus, this latter predicate satisfies *Congruence*. Besides, $\perp$ only holds on lines —and thus, it satisfies *Field*. Consequently, the language $\mathcal{L}_1$ can be extended by adding the predicate $\perp^*$ whose assertibility and truth conditions are governed by *Inheritance*. This procedure can be generalized to cover all abstraction principles of the form of Dir. As a result, *Inheritance* is regarded as a *necessary* condition for an abstraction step to be satisfactory.

A similar inheritance principle must be added to perform the corresponding extension of the base modal language $\mathcal{L}^\diamond$ when dealing with modal abstraction. As a result, a successful abstraction attempt will provide speakers with a whole language to talk about the abstracts. Recall that the aim is to apply modal abstraction to NS potentialist set theory to recover the axiom of Infinity. Therefore, I will restrict my attention to the case in which the background modal logic is S4.2 furnished with neutral free logic —the possibility of generalising the proposal to other systems remains open. Indeed, a system satisfying G

$$\diamond \Box \varphi \rightarrow \Box \diamond \varphi$$

—which, remember, axiomatizes convergent frames— would do the job. The reason why G must be added to the background logic will be apparent in a moment. Before, remember that all n -ary predicates P in $\mathcal{L}^\diamond$ are stable:

$$\begin{aligned} \text{Stb}_P^+ & \quad \Box \forall \alpha_1 \dots \alpha_n (P(\alpha_1 \dots \alpha_n) \rightarrow \Box P(\alpha_1 \dots \alpha_n)) \\ \text{Stb}_P^- & \quad \Box \forall \alpha_1 \dots \alpha_n (\neg P(\alpha_1 \dots \alpha_n) \rightarrow \Box \neg P(\alpha_1 \dots \alpha_n)) . \end{aligned}$$

In particular, the unity relation $\sim$ featuring in $\text{AP}^\diamond$ is stable:

$$\begin{aligned} \text{Stb}_{\sim}^+ & \quad \Box \forall \alpha \beta (\alpha \sim \beta \rightarrow \Box \alpha \sim \beta) \\ \text{Stb}_{\sim}^- & \quad \Box \forall \alpha \beta (\alpha \not\sim \beta \rightarrow \Box \alpha \not\sim \beta) . \end{aligned}$$

The requirement of adding G to the background modal logic can be understood now: if G fails, then the stability of the predicates P and the unity relation $\sim$ will not be enough to prevent radical branching in the models and, consequently, that $\sim$ or the P vary its extension across worlds in an “inconsistent” way.

Let v be a function that —among other things— assigns to each pair formed by an n -ary predicate P and a world w a subset of the n -ary Cartesian product of the domain of w (the extension of P in w): $v(P)_w \subseteq D_w^n$ —see Appendix 6.B for the details. Now, think about a non-convergent S4 model in which:

- (i) some possible world w_s accesses two different worlds w_j and w_k — $w_s R w_j$ and $w_s R w_k$,

- (ii) for all $w_r \in W$ such that $w_j R w_r$, then it does not happen that $w_k R w_r$, and for all $w_r \in W$ such that $w_k R w_r$, then it does not happen that $w_j R w_r$ —all those worlds accessible from w_j are inaccessible from w_k , and all those worlds accessible from w_k are inaccessible from w_j ,

and:

- (iii) $\mathcal{D}_{w_j}$ and $\langle \alpha, \beta \rangle \in v_{w_j}(\sim)$, but $\alpha, \beta \in \mathcal{D}_{w_k}$ and $\langle \alpha, \beta \rangle \notin v_{w_k}(\sim)$ —both α and β exist in w_j and w_k , but while $\alpha \sim \beta$ is true in w_j , it is false in w_k

or

- (iv) $\langle \alpha_1, \dots, \alpha_n \rangle \in \mathcal{D}_{w_j}^n$ and $\langle \alpha_1, \dots, \alpha_n \rangle \in v_{w_j}(P)$, but $\langle \alpha_1, \dots, \alpha_n \rangle \notin \mathcal{D}_{w_k}^n$ and $\langle \alpha_1, \dots, \alpha_n \rangle \notin v_{w_k}(P)$ —the $\alpha_1, \dots, \alpha_n$ exist both in w_j and in w_k , but while $P(\alpha_1 \dots \alpha_n)$ is true in w_j , it is false in w_k .

This situation would be fatal for modal abstraction. Any model satisfying (i)-(iv) is consistent with the stability of $\sim$ and the predicates P . However, (i)-(iv) are inconsistent with $\text{AP}^\diamond$. Because $\alpha \sim \beta$ is true in w_j , $\text{AP}^\diamond$ entails that $f(\alpha) = f(\beta)$ is true in w_ω —the possible world where the initial $\diamond$ is instantiated. But, because $\alpha \sim \beta$ is false in w_k , $\text{AP}^\diamond$ entails that $f(\alpha) = f(\beta)$ is false in w_ω . The axiom G together with the stability of those predicates in the base language $\mathcal{L}^\diamond$ prevents this situation and, thus, prevents the failure of $\text{AP}^\diamond$. Specifically, G entails —modulo the stability of the predicates P , among which is $\sim$ — that

if $\alpha \sim \beta \in v_{w_j}(\sim)$ for some $w_j \in W$, then for all $w_k \in W$ such that $\alpha, \beta \in \mathcal{D}_{w_k}$, $\alpha \sim \beta \in v_{w_k}(\sim)$

and

if $\langle \alpha_1, \dots, \alpha_n \rangle \in v_{w_j}(P)$ for some $w_j \in W$, then for all $w_k \in W$ such that $\alpha_1, \dots, \alpha_n \in \mathcal{D}_{w_k}$, $\langle \alpha_1, \dots, \alpha_n \rangle \in v_{w_k}(P)$

That is, if $\alpha \sim \beta$ is true at some possible world, then as soon as α and β exist in any possible world, $\alpha \sim \beta$ is true in this world. And, the same happens with the P s. As a result, the form of modal abstraction explored here only applies to systems satisfying convergence.⁸ Thus, it will be safe to apply it to the potentialist set theory NS, whose background logic combines a plural logic with S4.2, and whose negative free logic is replaced by the neutral option adopted here —see Chapter 6, 6.5, for details.

The next step is to determine which expressions of the base language $\mathcal{L}^\diamond$ are apt to introduce new predicates about abstracts satisfying the corresponding inheritance principle. Which properties must satisfy these expressions? Again, take the case of non-modal abstraction. First, a non-modal abstraction step selects those formulas in the base language with respect to which the unity relation is a congruence. Now, it is reasonable to expect that formulas in the base modal language involved in a modal abstraction step also satisfy a similar requirement. Those formulas should not discriminate between elements related by the unity relation $\sim$. However, unlike non-modal abstraction, the new modal framework demands to go beyond the actual context. All the accessible contexts must be taken into account. Therefore, take all those formulas $\varphi(\alpha_1, \dots, \alpha_n)$ for which the unity relation in $\text{AP}_{\text{NS}}^\diamond$ is *necessarily* a congruence:

⁸It applies to what is known as *non-branching* potentialism.

Congruence[◇]

$$\Box \forall \alpha_1 \dots \alpha_n [\varphi(\alpha_1 \dots \alpha_n) \rightarrow \Box \forall \beta_1 \dots \beta_n [\alpha_1 \sim \beta_1 \wedge \dots \wedge \alpha_n \sim \beta_n \rightarrow \varphi(\beta_1 \dots \beta_n)]].$$

The conditional form is required because the congruence condition —the consequent— holds under the assumption that $\varphi(\alpha_1 \dots \alpha_n)$ is the case in some situations. Formulas satisfying *Congruence*[◇] necessarily do not discriminate between elements related by $\sim$. This is the analogous condition to *Congruence* when the outer domain $\mathcal{D}$ is considered. Consequently, those formulae satisfying *Congruence*[◇] are apt to be “inherited” by the language committed to abstracts. In the case of $\mathcal{L}_{\text{NS}}$ and PLV[◆], atomic formulae —involving the predicates⁹ “ $\prec$ ”, “ $\in$ ” and “ $=$ ” — as well as the potentialist translation of set-theoretic formulas are stable —see Lemma 3.1 in (82, p. 74) — and, in consequence, they satisfy *Congruence*[◇].

Second, non-modal abstraction demands formulae which are only satisfied inside the field of the unity relation. This can be easily adapted to the new modal framework. Simply consider all those formulas which *necessarily* are satisfied inside the field of the unity relation $\sim$ in AP[◆]:

Field[◇]

$$\Box \forall \alpha_1, \dots, \alpha_n (\varphi(\alpha_1 \dots \alpha_n) \rightarrow \alpha_1 \sim \alpha_1 \wedge \dots \wedge \alpha_n \sim \alpha_n).$$

This condition ensures that abstracts inherit their properties from those properties which only their specifications have. Again, in the case of PLV[◆] these are properties about pluralities expressed by atomic formulas containing “ $\prec$ ” and “ $=$ ” as well as other formulas of plural logic.

Congruence[◇] and *Field*[◇] perform in the new modal framework the same job as *Congruence* and *Field* did in the case of non-modal abstraction. Abstracts involved in modal abstraction inherit their properties and relations from those properties and relations expressed by formulas satisfying both *Congruence*[◇] and *Field*[◇]. Thus, we can pick each one of these formulas φ and extend the base language of abstraction $\mathcal{L}^\diamond$ by adding a new predicate φ^* committed to abstracts, whose assertibility and truth conditions are governed by the following modal version of the inheritance principle:

Inheritance[◆]

$$\Box \blacksquare \downarrow \Box \forall \alpha_1 \dots \alpha_n (\uparrow \varphi^*(f(\alpha_1) \dots f(\alpha_n)) \leftrightarrow \varphi(\alpha_1 \dots \alpha_n)).$$

—see Appendix 6.c for a proof of the relative consistency of AP[◆] and *Inheritance*[◆]. The sequence of operators $\Box$, $\blacksquare$, $\downarrow$ and $\uparrow$ mimics the arrangement of these same operators in AP[◆]. Let W^* be a collection of worlds —interpreting the black modal operators— extending the original collection W —interpreting the white modal operators. The white square $\Box$ indicates that *Inheritance*[◆] holds in any world $w \in W$. Then, the black square $\blacksquare$ says that the predicates $\varphi^*(f(\alpha_1) \dots f(\alpha_n))$ are governed by AP[◆] in any $w \in W^*$. This entails that all abstracts $f(\alpha)$ corresponding to some possible specification α exist in all possible worlds in W^* — $f(\alpha) \in \mathcal{D}_w$, for all $w \in W^*$. However, this is a welcomed result when modal abstraction is applied to the potentialist set theory NS. Consider the empty plurality. This is a subplurality of the domain of all worlds interpreting the white modal operators —i.e. it is a subplurality of $\mathcal{D}_w$, for all $w \in W$. Similarly, modal abstraction ensures that the plurality of all abstracts arising from this modality is a subplurality of the domain of all the new worlds —i.e. it is a subplurality of $\mathcal{D}_w$, for all $w_r \in W^*$. This is in tune with the iterative conception of sets. Here, the plurality of all natural numbers —which

⁹Recall that “ $\in$ ” can be taken as a primitive expression.

would correspond to the abstracts generated by NS— is the domain of V_ω and it is a subplurality of the domain of all V_γ , $\omega < \gamma$. Modal abstraction allows gathering all the abstracts generated by NS forming a plurality analogous to V_γ in the case of potentialist set theory.

7.4 INFINITY

Modal abstraction takes place over modal facts. Let's place speakers in a particular world of an S4.2 model. Then, the modalized quantifier " $\Box\forall$ " allows them to go outside this particular point, travelling across the whole chain of accessible worlds. Throughout their journey, the speakers' experience is largely enriched, as they can see what happens in these worlds. Thus, they see what is possible from where they are. Specifically, they see what is possible concerning the unity relation $\sim$. Similarly to non-modal abstraction, modal abstraction principles provide abstracts with identity conditions stated in terms of this relation. The difference is that the modal case does not demand that objects α and β exist in the world for the unity relation to hold between them. The mere possibility of this fact suffices for abstraction to take place. Then, whenever it is possible for two specifications α and β to enter this relation —i.e. whenever it is possible that $\alpha \sim \beta$ —, modal abstraction guarantees that an abstract $f(\alpha)$ corresponding to them exists. As a result, a successful modal abstraction step can extend the language using new abstract terms $f(\alpha)$, $f(\beta)$, etc. and a whole c of new predicates ϕ^* inherited from expressions of the base, modal language only committed to specifications.

As a result of its broader scope, modal abstraction principles can break the ceiling of the finite. When $\text{PLV}^\Diamond$

$$\Box\Diamond\downarrow\Box\forall uvv(\uparrow[\{uu\} = \{vv\}] \leftrightarrow \forall z(z \prec uu \leftrightarrow z \prec vv)).$$

is added to the potentialist set theory NS, modal abstraction enhances the hierarchy of hereditarily finite sets by introducing a countably infinite plurality. Thus, modal abstraction introduces an “actual” infinity of objects, proving the axiom of Infinity and interpreting Z. The idea behind this result is quite simple.¹⁰ When evaluated at some possible world $w \in W$, $\text{PLV}^\Diamond$ takes stock of all possible pluralities of sets placed at accessible worlds. Then, for each possible plurality coextensional with itself —thus, for all possible pluralities— it ensures the existence of an abstract in another world $w_\omega \in W^*$. Therefore, $\text{PLV}^\Diamond$ introduces infinite many objects in $\mathcal{D}_{w_\omega}$.

Let $\mathcal{L}_{\downarrow\uparrow\in}^\Diamond$ be the bi-modal language resulting from extending the language $\mathcal{L}_\in^\Diamond$ of NS by adding the modal operators $\Diamond$, $\downarrow$ and $\uparrow$. The modality expressed by these operators is *interpretational* —see Chapter 6, 6.2. The modal plural logic ($\text{MPL}^\Diamond$) of this language is the S4.2, plural system of MPL —see Chapter 6, 6.5. Let “ $\Diamond$ Infinite” denote the formula that results of preceding the axiom of Infinity

$$\exists x(\emptyset \in x \wedge \forall y(y \in x \rightarrow \{y\} \in x))$$

with an occurrence of $\Diamond$. Finally, let “ $\vdash^\Diamond$ ” denote deducibility in $\text{MPL}^\Diamond$.

Definition 1. Let $\text{NS}^\Diamond$ be the $\mathcal{L}_{\downarrow\uparrow\in}^\Diamond$ -theory that adds to NS the following axioms:

¹⁰It has also recently been noted by Studd (2023).

Collapse $^\diamond$	$\diamond \downarrow \Box \forall uu \uparrow \exists y(y = \{uu\})$
Collapse $_2^\diamond$	$\blacksquare \forall uu \diamond \exists x(x = \{uu\})$
Extensionality $^\diamond$	$\blacksquare \forall xz \downarrow \Box \forall uvv(x = \{uu\} \wedge z = \{vv\} \rightarrow (z = z \leftrightarrow \forall y(y \prec uu \leftrightarrow y \prec vv)))$
PLV $^\diamond$	$\blacksquare \forall xz \blacksquare \forall uvv(x = \{uu\} \wedge z = \{vv\} \rightarrow (x = y \leftrightarrow \blacksquare(uu \equiv vv)))$.

Theorem 1 $NS^\diamond \vdash^\diamond \diamond \text{Infinity}$.

Proof Sketch. NS proves $\diamond \exists uu \forall x(x \neq uu)$. Also, it proves $\Box \forall uu \diamond \exists vv \Box \forall z(z \prec vv \leftrightarrow z \prec uu \vee z = \{vv\})$. By Collapse $^\diamond$, $\diamond(\exists y(y = \emptyset) \wedge \forall y \exists z(z = \{y\}))$. Recall that $\blacksquare$ also obeys the necessitation rule. Thus, instantiate $\diamond$ P-Comp with the condition of being self-identical: $\blacksquare \exists uu \forall z(z \prec uu \leftrightarrow z = z)$. It follows from this and the above that $\diamond \exists uu(\emptyset \prec uu \wedge \forall y(y \prec uu \leftrightarrow \{y\} \prec uu))$. Finally, Collapse $_2^\diamond$ entails $\diamond \exists x(\emptyset \in x \wedge \forall y(y \in x \rightarrow \{y\} \in x))$. $\square$

PLV $^\diamond$ supplies us with an infinite plurality of objects. As soon as the classical *Collapse* is extended to apply to pluralities of objects in the worlds in W^* , a countable set appears. As it was done in the case of the abstraction steps performed by PLV, the new abstracts $\{uu\}$, $\{vv\}$, etc. can be identified with the old ones in a natural way using extensionality. Thus, for all old sets x formed from a plurality vv just stipulate that

$$x = \{uu\} \text{ iff } \forall y(y \prec vv \leftrightarrow y \prec uu).$$

This is also in tune with *piecemeal* solution to the Caesar Problem advocated in Chapter 5. 5.4.1.5 The status of mixed identities between abstracts and other objects is determined in the way identities between abstracts: by laying down additional stipulations which constrain speakers' conceptual resources.

Modal abstraction is a powerful device. It can take that which for the hierarchy of finite hereditary sets is a proper class, and “turn it” into a plurality. Metaphorically speaking, it can be compared with condensation. When a gaseous substance condenses, it changes state and becomes a liquid. Similarly, the modal abstraction principles explored here describe the process by which some collection that hangs in the air —the proper classes from the perspective of a particular hierarchy of sets— changes its state, condensing into a plurality. If the process is taken further, the plurality solidifies into a set. Thus, when abstraction is pushed to its limits, speakers can turn what is possible and intensional into an extensional collection of objects. .

7.4.1 Limits of the Proposal

The modal abstraction framework sketched in the previous sections needs to be generalized to achieve a unified abstractionist account of set theory. Abstraction principles can be thought of as explaining the basic operations composing the iterative conception of sets. The classical Collapse provides an analysis of the Set-of operation when applied at successor stages. Given a plurality of objects uu , Collapse guarantees the existence of the set $\{uu\}$ whose members are exactly those objects among the uu . Thus, starting with the empty plurality of objects, Collapse guarantees the existence of the empty set $\emptyset$, which constitutes the first stage of the hierarchy: $V_0 = \emptyset$. Then, the operation can be applied again. Given a stage V_n , Collapse

guarantees the existence of sets corresponding to subpluralities of this stage. The next stage V_{n+1} collects all these sets. This follows in NS if a maximality principle is added to the theory —see Chapter 6, 6.5. The result of the successive application of Collapse starting with the empty plurality is the hierarchy of finitely hereditary sets described by the following equations:

$$\begin{aligned} V_0 &= \emptyset \\ V_{n+1} &= \mathcal{P}(V_n), \end{aligned}$$

for all $n < \omega$. This hierarchy only contains finite sets. No matter how many times Collapse is applied to finite pluralities the result will always be a collection of finite sets.

In turn, Collapse $^\blacklozenge$ can be thought of as providing an analysis of the *union* operation. This operation takes each V_n , for all $n < \omega$, and pour its sets into a new stage V_ω , the first infinite, *limit* stage of the hierarchy.¹¹ Hence, this new stage collects together all the sets placed in the finite V_n .

$$V_\omega = \bigcup_{n < \omega} V_n$$

This is precisely what Collapse $^\blacklozenge$ does. It browses the potential hierarchy of sets and collects all the finite sets inside a new stage of the hierarchy. The first infinite stage. Now, once V_ω is introduced, Collapse can be applied again forming additional sets, and further stages $V_{\omega+1} = \mathcal{P}(V_\omega)$, $V_{\omega+2} = \mathcal{P}(V_{\omega+1})$,... In particular, the application of Collapse to the whole domain of V_ω yields the set ω , the first infinite set, in $V_{\omega+1}$. A whole round of applications gives all the successor stages $V_{\omega+n}$. Then, the union operation enters the picture again forming the limit stage $V_{\omega+\omega}$. The process continues endlessly, alternating rounds of successive applications of the set-of operation with applications of the union operation.

Modal abstraction must be generalized by developing a precise modal theory that describes the potentialist hierarchy arising when whole rounds of Collapse are combined with applications of Collapse $^\blacklozenge$. Furthermore, if this process were taken far enough, abstraction would support not only the axiom of Infinity but also the axiom of Replacement. As it is known, what is needed to prove this axiom is that the hierarchy —in this case, the potentialist hierarchy— is high enough.¹² Indeed, if the process of abstraction reaches a fixed point, absolute quantification over the whole potentialist hierarchy of sets could be retrieved. As a result, the combination of modal and non-modal abstraction principles would provide a unified foundation for ZF.

However, the options for finding this general theory run into serious problems. First of all, the combination of both Collapse and Collapse $^\blacklozenge$ must be expressed in the object language. This requires two modifications to the framework developed so far. First, the bimodal language $\mathcal{L}_{\downarrow\uparrow\in}^\blacklozenge$ must be replaced by the uni-modal language $\mathcal{L}_{\downarrow\uparrow\in}^\diamond$ of NS. This contains the operators $\downarrow$ and $\uparrow$, but the black modal operators are merged with the white ones. Second, Collapse $^\blacklozenge$ must be expressed in the new modal framework. The most immediate option is to replace $\blacklozenge$ by $\diamond$:¹³

¹¹The first limit stage is $V_\emptyset$.

¹²See (10) for a proof of Replacement from an axiom encoding a special property of the iterative conception of sets. See(91) for an analysis of the evidence employed to motivate Replacement.

¹³Aslo, the strange models $\mathcal{M}^* = \langle \mathcal{D}, \mathcal{D}^*, W, W^*, R, R^*, v^* \rangle$ introduced in Appendix 1.B must be replaced by the usual quadruples $\mathcal{M}^+ = \langle \mathcal{D}^+, W^+, R^+, v^+ \rangle$. This should be done by combining in the right way each element x in $\mathcal{M}^*$ with the corresponding x^* .

$\text{Collapse}_2^\diamond$

$$\diamond \downarrow \Box \forall uu \uparrow \exists x(x = \{uu\}).$$

If necessary, additional principles should be added to specify how this modal abstraction principle relates to PLV. But, here is where the problems arise: $\text{Collapse}_2^\diamond$ entails a contradiction when added to NS. More specifically, the problem is the modal axiom 4 —i.e. the transitivity of the accessibility relation.

$$4 + \text{NS} + \text{Collapse}_2^\diamond \vdash^\diamond \perp$$

Together, 4 and $\text{Collapse}_2^\diamond$ reintroduce —once more— the Russell’s paradox. For, transitivity allows the modalized quantifier $\Box \forall$ to range over the whole potential hierarchy of sets. Then, $\text{Collapse}_2^\diamond$ introduces in some world every set corresponding to a plurality. Consequently, the domain of some world will contain all sets, what is inconsistent in NS.

Therefore, we face a dilemma. Or 4 —transitivity— is abandoned, or a different formalization of $\text{Collapse}^\diamond$ is found. Both options are unattractive. Take the first horn of the dilemma. The predicative nature of $\text{Collapse}^\diamond$ speaks against transitivity and *a fortiori* against absolute quantification. The reason for introducing the black modal operators was to expand the logical space covered by the modalized quantifiers in the first place. However, if this expansion can be performed whenever the modal operators $\Box$ and $\diamond$ are introduced, there is little hope of expressing absolute generality by using the modalized quantifiers $\Box \forall$ and $\diamond \exists$. At the level of the object language, this traduces into the failure of transitivity: at a given world the modalized quantifiers do not range over the transitive closure of the accessibility relation. At a given world w , there will be other worlds w_1 and w_2 such that wRw_1 and w_1Rw_2 , but not wRw_2 . Furthermore, the failure of transitivity implies the failure of the potentialist translation. This depends on the modalised quantifiers expressing some form of absolute generality —they must range over the whole hierarchy of sets. Therefore, if the latter fails, so does the former. As a result, ZF cannot be interpreted in the modal language. Predicativism is a beneficial stance for dealing with paradoxes. However, it is only fully satisfactory when it can be combined with absolute quantification.

Alternatively, suppose that $\text{Collapse}_2^\diamond$ is formalized differently. This would be a satisfactory solution. If we find a principle that performs the same function as A and respects transitivity, then it could be applied to NS without any problem. However, so far I have not found such a formalisation. Consequently, the second horn of the dilemma remains as unattractive as the former. Until a suitable way is found to develop either alternative, we will remain trapped in this dilemma. And, as long as this is the case, we will not be able to generalise the modal abstraction principles into a unified and simple system capable of interpreting ZF. If this is not possible and absolute generality is incompatible with modal abstraction principles, this is a point against them.

7.5 CONCLUSION

In this chapter, I have developed modal abstraction principles of the form $\text{AP}^\diamond$. The main benefit of modal abstraction is that it is a form of *specificationless* abstraction. The abstraction process does not take place over objects related by the unity relation in the present world. Instead, it takes place over possible facts

concerning this relation. The possibility that two objects are related by it is enough for abstraction to introduce the corresponding abstract specified by them. First, I have shown some unsatisfactory ways of stating principles. Then, I extended the language of the potentialist set theory NS by introducing the black modal operators, the backtracking operator and the unbacktracking operator. Using these resources, modal abstraction principles can be formalized as $AP^\blacklozenge$. As a consequence, it has become clear that modal abstraction must be predicative, strengthening the power of the modalized quantifiers. Thirdly, I modified Inheritance by using the same resources. As a result, classical abstraction is fully recovered in the new modal context. In fourth place, I have shown how $AP^\blacklozenge$ interprets the axiom of Infinity when added to the potentialist set theory NS. Thus, abstractionism can explain how speakers grasp the notion of infinity. I have pointed out some of the difficulties in generalising and simplifying the proposal. To work with a single type of modal operator, either absolute generality is rejected or $AP^\blacklozenge$ is rejected. However, neither option is attractive. Consequently, if no solution to this dilemma is found, it could be a reason against modal abstraction.

APPENDIX 7.A

To ensure that modal abstraction principles are safe, in these appendices I will present a relative consistency test. Let us start with models for normal modal logics. Let $\mathcal{L}^\diamond$ be a first-order modal language including the operators $\diamond$ and $\Box$. First-order models will be ordered quadruples $\mathcal{M} = \langle \mathcal{D}, W, R, v \rangle$ where:

- (i) $\mathcal{D}$ is a non-empty collection of objects—the *outer domain* of objects
- (ii) W is a non-empty set of possible worlds
- (iii) $R \subseteq W^2$ is the accessibility relation between worlds

and

- (iv) v is a function such that: to each world $w_i \in W$, v assigns its *inner domain* of objects— $v(w) = D_w$ —; to each term τ of $\mathcal{L}^\diamond$ an object d in D — $v(\tau) = d$ —; and to each n -ary predicate P of $\mathcal{L}^\diamond$, and each world $w \in W$ a subset of the n -ary Cartesian product of $\mathcal{D}_w^n$ — $v_w(P) \subseteq \mathcal{D}_w^n$.

Truth conditions are as expected. Otherwise, additional domains to interpret higher-order variables could be added as usual.

APPENDIX 7.B

Let $\mathcal{L}_{\downarrow\uparrow}^\blacklozenge$ be the language that results when $\mathcal{L}^\diamond$ is extended by adding the operators $\blacklozenge$, $\downarrow$ and $\uparrow$. The use of this variety of novel resources calls for explanation and a precise definition in the form of model-theoretic semantics or proof-theoretical rules. Unfortunately, I cannot stop here to give them the attention they require. Instead, I want to say a few words about how the acceptance of these resources would affect the models introduced in the previous appendix.

First, to interpret the provisional $\blacklozenge$ the first-order models should be modified by adding to $\mathcal{M}$ a new domain of objects, a new domain of possible worlds and a new accessibility relation on these new worlds. Thus, models are tuples $\mathcal{M}^* = \langle \mathcal{D}, \mathcal{D}^*, W, W^*, R, R^*, v^* \rangle$ where $\mathcal{D}$, W , and R are the same as in $\mathcal{M}$ —see Appendix 1.B—, and

- (i) $\mathcal{D}^*$ is a non-empty collection of objects, such that possibly $\mathcal{D} \subseteq \mathcal{D}^* \text{ — } \mathcal{D} \cup \mathcal{D}^*$ can be thought of as the *outer domain*
- (ii) W^* is an additional stock of possible worlds such that $W \cap W^* = \emptyset$ and $w_\omega \in W^*$
- (iii) R^* is the accessibility relation defined on $W \cup W^*$ such that $\{ \langle w_i, w_j \rangle \mid w_i \in W \wedge w_j \in W^* \} \subseteq R^*$ —that is, all worlds in W access all worlds in W^*
- (iv) v^* is a variable assignment extending v in the obvious way: to each world $w_i \in W \cup W^*$ it assigns its *inner domain* of objects — $v_{w_i} = \mathcal{D}_{w_i}$ —, to each term τ of $\mathcal{L}_{\text{NS}}^\blacklozenge$ an object $d \in \mathcal{D} \cup \mathcal{D}^* \text{ — } v(\tau) = d$ —, and to each n -ary predicate P_k of $\mathcal{L}_{\text{NS}}^\blacklozenge$ and each world $w_i \in W \cup W^*$ a subset of the n -ary Cartesian product of $\mathcal{D}_{w_i}^n \text{ — } v_{w_i}(P_k) \subseteq \mathcal{D}_{w_i}^n$.

As before, additional domains to interpret higher-order terms can be added as usual. The semantic clauses corresponding to the black operators $\blacklozenge$ and $\blacksquare$ are the same as those for $\lozenge$ and $\square$, replacing W^* and R^* by W and R respectively. These are strange models. Nevertheless, they are only a *provisional* device used to expose modal abstraction. Consequently, the resulting situation is halfway between a two-sorted and a one-sorted framework: while the language $\mathcal{L}_{\downarrow\uparrow}^\blacklozenge$ is one-sorted, models $\mathcal{M}^*$ look like two-sorted models endowed with several domains. This artificial situation needs correction. But, as I will argue, such a correction involves more difficulties than expected —see Section 4.1.

Second, the semantics of $\downarrow$ require further modifications on the definition of *satisfiability* and *validity*. From a semantic point of view, the intended role of the backtracking operator is to go back in our journey across worlds, arriving at the world where the innermost operator affecting it was evaluated. The problem is that the truth condition of the operator can not be given in terms of the inverse R^{-1} of the accessibility relation. For, suppose $\lozenge \downarrow \varphi$ is evaluated at w_0 —the same happens if $\blacklozenge \downarrow \varphi$ is considered instead. Moreover, suppose the diamond is instantiated by w_i —what is at stake is the truth value of $\downarrow \varphi$ in w_i . Besides, suppose that:

- (i) $w_k R w_i$
- (ii) φ is true at w_k
- (iii) φ is false at w_0 .

If the truth condition corresponding to $\downarrow$ were the same as the usual clause corresponding to $\lozenge$, but substituting R^{-1} by R , then the formula $\downarrow \varphi$ would be true in w_i because of (ii) —thus, $\lozenge \downarrow \varphi$ would be true in w_0 . However, both formulas should be false because (iii): $\downarrow$ aims to take us to w_0 , where φ is false. To deal with this kind of difficulty, Hodes provides truth conditions for the modal language in such a way that satisfiability is defined relatively to *sequences of worlds*

$\langle w_1, \dots, w_n \rangle$, where w_i accedes w_{i+1} for each $0 < i < n$

not to a world, as usual, —see (58; 104). These sequences of worlds can be thought of as a record of each “application” of the truth condition corresponding to the modal operators $\Diamond$ and $\Box$ when the truth value of a sentence is assessed. The clauses are the following — v is a variable assignment in $\mathcal{M}^*$ and $R^+ = R \cup R^*$:

$$\begin{aligned}
& \langle \langle w_1 \dots w_n \rangle, v \rangle \models P_i(\alpha_1 \dots \alpha_n) \text{ iff } \langle v(\alpha_1), \dots, v(\alpha_n) \rangle \in v_{w_n}(P_i) \\
& \langle \langle w_1 \dots w_n \rangle, v \rangle \models \neg \phi \text{ iff } \langle \langle w_1 \dots w_n \rangle, v \rangle \not\models \phi \\
& \langle \langle w_1 \dots w_n \rangle, v \rangle \models \phi \vee \psi \text{ iff } \langle \langle w_1 \dots w_n \rangle, v \rangle \models \phi \text{ or } \langle \langle w_1 \dots w_n \rangle, v \rangle \models \psi \\
& \langle \langle w_1 \dots w_n \rangle, v \rangle \models \exists x \phi \text{ iff there is an } a \in \mathcal{D}_{w_n} \text{ such that } \langle \langle w_1 \dots w_n \rangle, v^{a/\alpha} \rangle \models \phi \\
& \langle \langle w_1 \dots w_n \rangle, v \rangle \models \Diamond \phi \text{ iff } \langle \langle w_1 \dots w_n, w_{n+1} \rangle, v \rangle \models \phi \text{ such that } w_n R w_{n+1} \\
& \langle \langle w_1 \dots w_n \rangle, v \rangle \models \blacklozenge \phi \text{ iff } \langle \langle w_1 \dots w_n, w_{n+1} \rangle, v \rangle \models \phi \text{ such that } w_n R^* w_{n+1} \\
& \langle w_1 \dots w_n \rangle \models \downarrow \phi \text{ iff } \langle w_1 \dots w_{n-1} \rangle \models \phi \text{ such that } w_{n-1} R^+ w_n.
\end{aligned}$$

As I said, each w_i , $0 < i \leq n$ can be thought of as representing a point where modal operators are met when evaluating a formula.¹⁴ During this process, each subformula is evaluated both globally, relatively to the world sequence $\langle w_1 \dots w_n \rangle$, and locally, relatively to w_n —the latest point we passed through. Thus, the clause corresponding to $\downarrow$ gives the desired result: it sends us back to w_{n-1} , the last world where the innermost modal operator affecting $\downarrow$ was evaluated. Besides, w_{n-1} is uniquely determined in $\langle w_1 \dots w_n \rangle$ and the aforementioned problem is solved. Now, validity can be defined as follows:

$$\Gamma \models \phi \text{ iff for all } \langle \langle w_1 \rangle, v \rangle \text{ and all } \psi \in \Gamma, \text{ if } \langle \langle w_1 \rangle, v \rangle \models \psi, \text{ then } \langle \langle w_1 \rangle, v \rangle \models \phi,$$

where v is a variable assignment. To be precise about all these points: the whole semantics of the potentialist set-theory NS should be modified by adding this machinery.

Finally, the unbacktracking operator $\uparrow$ needs a precise semantics. One option is to add a record of the successive applications of $\downarrow$ as follows. The role of $\uparrow$ is to take us back to a world where $\downarrow$ was applied. For instance, suppose we are evaluating a formula and —according to the semantics for $\downarrow$ — a sequence

$$\langle w_1 \dots w_n \rangle$$

of possible worlds is created as a record of the applications of $\Box$ and $\Diamond$. Then, suppose that $\downarrow$ is met. When this happens two operations are performed on the sequence:

1. The last world w_n is erased, as demanded by the truth condition of $\downarrow$: $\langle w_1 \dots w_{n-1} \rangle$
2. A copy of the whole list is added on the right: $\langle \langle w_1 \dots w_{n-1} \rangle, \langle w_1 \dots w_n \rangle \rangle$.

Truth conditions of modal operators and $\downarrow$ will apply to $\langle w_1 \dots w_{n-1} \rangle$, erasing and adding worlds. Then, when $\uparrow$ is met, the leftmost list is removed and the rightmost list $\langle w_1 \dots w_n \rangle$ is put in its place. Consequently, the truth conditions for $\uparrow$ would mimic those of the backtracking operator.

¹⁴Indeed, given a formula ϕ , Hodges’ truth conditions “select” a submodel of $\mathcal{M}^*$ where ϕ is evaluated. This can be thought of as the “portion” of $\mathcal{M}^*$ making ϕ true.

APPENDIX 7.C

To ensure that modal abstraction is a safe form of abstraction, I will sketch how a model can be constructed to prove that $AP^\diamond$ and $Inheritance^\diamond$ are consistent relative to S4.2 with a background neutral free logic. Let $\mathcal{L}^\diamond$ be a modal language with a predicate $\sim$. Furthermore, assume that its background logic is S4.2. Now, let $\mathcal{M} = \langle \mathcal{D}, W, R, v \rangle$ be a first-order model —as described in Section 2.1— of this system such that:

- (i) $\mathcal{M}$ satisfies $Stb_{P_i}^+$ and $Stb_{P_i}^-$ for all predicates P_i in $\mathcal{L}^\diamond$ —in particular, it satisfies $Stb_\sim^+$ and $Stb_\sim^-$,
- (ii) $v_w(\sim)$ is a partial equivalence relation on $\mathcal{D}_w$, for all $w \in W$,

and

- (iii) $\mathcal{M}$ satisfies the Barcan formula.

Define the equivalence relation $\sim^+$ on the outer domain of objects $\mathcal{D}$ as follows:

$$\alpha \sim^+ \beta \text{ iff for some } w \in W, \langle \alpha, \beta \rangle \in v_w(\sim), \text{ for all } \alpha, \beta \in \mathcal{D}.$$

It is easy to check that $\sim^+$ is a partial equivalence relation on $\mathcal{D}$ —that is, a relation that is symmetric and transitive in $\mathcal{D}$. For, assume $\alpha \sim^+ \beta$. Then, —by the definition of $\sim^+$ — there is a world $w \in W$ such that $\langle \alpha, \beta \rangle \in v_w(\sim)$. But, because of (ii), $\langle \beta, \alpha \rangle \in v_w(\sim)$. Therefore, —by definition again— $\alpha \sim^+ \beta$. Hence, $\sim^+$ is transitive. Besides, assume $\alpha \sim^+ \beta$ and $\beta \sim^+ \delta$. By definition, there are worlds $w_1, w_2 \in W$ such that $\langle \alpha, \beta \rangle \in v_{w_1}(\sim)$ and $\langle \beta, \delta \rangle \in v_{w_2}(\sim)$. By convergence, there is a world w_3 such that $w_1 R w_3$ and $w_2 R w_3$. Moreover, by (iii) $\alpha, \beta, \delta \in \mathcal{D}_{w_3}$. And, by $Stb_\sim^+$ and (ii): $\langle \alpha, \delta \rangle \in v_{w_3}(\sim)$. Therefore, it follows —once more by definition— that $\alpha \sim^+ \delta$. Furthermore, if (ii) is strengthened as to claim

- (ii)' $v_w(\sim)$ is an equivalence relation on $\mathcal{D}_w$, for all $w \in W$,

it follows that $\sim^+$ is reflexive, and thus it is an equivalence relation on $\mathcal{D}$. Let $[\alpha]_{\sim^+}$ the equivalence class of α under $\sim^+$.

Now, let $\mathcal{M}^* = \langle \mathcal{D}, \mathcal{D}^*, W, W^*, R, R^*, v^* \rangle$ be a first-order model of $\mathcal{L}_{\downarrow\uparrow}^\diamond$ —as described in Section 2.4— where $\mathcal{D}, W$ and R are the same as in $\mathcal{M}$. Besides, let:

- 1 $\mathcal{D}^* = \mathcal{D} \cup \{[\alpha]_{\sim^+} / \alpha \in \mathcal{D}\}$
- 2 $W^* = \{w^*\}$
- 3 $R^* = \{\langle w, w^+ \rangle / w \in W\}$
- 4 $v^*(w^*) = \mathcal{D}_{w^*} = \mathcal{D}^*$ and $v_{w^*}(f(\alpha)) = [\alpha]_{\sim^+}$.

Note that

$$[\alpha]_{\sim^+} = [\beta]_{\sim^+} \text{ iff for some } w \in W : \langle \alpha, \beta \rangle \in v_w^*(\sim).$$

From this it follows that $f(\alpha) = f(\beta)$ is true in w^* iff for some $w \in W$, $\alpha \sim \beta$ is true in w , for all $\alpha, \beta \in \mathcal{D}$. Using this fact —together with the semantics of $\diamond, \downarrow, \uparrow$ sketched in Section 2.4— it can be proved that all sequences $\langle w_1 \rangle$ of possible worlds such that $w_1 \in W$ satisfy $AP^\diamond$. Thus, $\mathcal{M}^*$ is a model of $AP^\diamond$.

Finally, assume that the formula $\varphi(\alpha_1 \dots \alpha_n)$ of the base language $\mathcal{L}^\diamond$ satisfies *Congruence* $^\diamond$ and *Field* $^\diamond$. Let F_φ be the subset of $\mathcal{P}(\{[\alpha]_{\sim+} / \alpha \in \mathcal{D}\}^n)$ satisfying

$$\langle [\alpha_1]_{\sim+}, \dots, [\alpha_n]_{\sim+} \rangle \in F_\varphi \text{ iff for some } w \in W : \langle \alpha_1 \dots \alpha_n \rangle \in v_w^*(\varphi(\alpha_1 \dots \alpha_n)).$$

Furthermore, let $v_{w^*}(\varphi^*) = F_\varphi$. From the definition of $\mathcal{M}^*$ it can be proved that $\varphi^*(f(\alpha_1) \dots f(\alpha_n))$ is true in w^* iff $\varphi(\alpha_1 \dots \alpha_n)$ is true in some $w \in W$, for all $\alpha_1, \dots, \alpha_n \in \mathcal{D}$. Employing this fact—together with the semantics of $\blacklozenge$, $\downarrow$ and $\uparrow$ sketched in Section 2.4—it can be proved that all sequences $\langle w_1 \rangle$ of possible worlds such that $w_1 \in W$ satisfy *Inheritance* $^\blacklozenge$. Therefore, $\mathcal{M}^*$ is a model of *Inheritance* $^\blacklozenge$.

8 Conclusion

The main objective of this research has been to recover a minimum of non-logical objectivity for mathematical knowledge while defending its status as a science of abstract objects, independent from us. This intuitive and amenable characterization of mathematics has been challenged by Benacerraf's problem and the emergence or limitation theorems at the level of meta-mathematics. As we have seen, the former questions the use of strong, counterfactual relations to explain mathematical knowledge and mathematical objectivity. In turn, limitation theorems have been used to question the use of weaker relations based on the imposition of logical and theoretical constraints on our theories to explain mathematical reference. This situation has led to the emergence of new proposals for rethinking the nature of mathematics. Among the most promising are pluralist philosophies, such as fictionalism and multiversim, that reduce mathematical knowledge to logical knowledge [thus, mathematical objectivity is replaced by logical objectivity. As a result, these proposals deny the existence of a single, objectively correct mathematical theory: all consistent mathematical theories are correct from the point of view of semantics and metaphysics.

However, in this research, I have argued that this kind of pluralist account of mathematical knowledge has serious problems. Mainly, they fail to provide a solution to the Language Acquisition Challenge. Logical knowledge —i.e. knowledge of consistency and logical entailment— presupposes the linguistic competence of speakers. Speakers must first understand a language to make inferences inside it. However, neither fictionalism nor multiversism can explain how mathematical vocabulary acquires its content, and how mathematicians understand it. Consequently, an alternative explanation is required. In this research, the chosen alternative has been abstraction principles.

In Chapter 2, I characterized the objectivity of empirical knowledge as the conjunction of two theses: Independence and Responsiveness. Independence claims that the content of those beliefs, judgements, or statements involved in objective knowledge is independent of agents' epistemic responses. In turn, Responsiveness claims that objective knowledge involves a strong, counterfactual relation between agents and the world. After that, I characterized mathematical platonism as the philosophical thesis resulting from the conjunction of Independence and two additional theses: Existence and Abstractness. Together, they claim that mathematical, abstract objects exist with independence of us. According to platonism, mathematics is the science that studies those objects. Then, I explained Benacerraf's challenge. This claims that platonism cannot offer a suitable epistemology for mathematics. But, the challenge is only conclusive if mathematical knowledge satisfies Responsiveness. And, this is unjustified: once this requirement is abandoned, there is plenty of room for explaining mathematical knowledge. Finally, argued

that platonism does not involve any metaphysical thesis about the structure of the world or any particular theory of truth. This satisfies objective 1.

Chapter 3 was devoted to the achievement of objective 2. In this chapter, I analyzed fictionalism. According to this, while mathematical language should be taken at face value, mathematical theories are untrue fictions. Therefore, mathematics is no longer the science of abstract objects existing independently of us. Instead of truthfully describing the mathematical realm, mathematicians create stories and study what happens in them. Thus, fictionalism reduces mathematical knowledge to logical knowledge: knowledge about the consistency and consequences of those fictions. I challenged fictionalism for three reasons. First, I argued that under two reasonable assumptions, fictionalism cannot explain mathematical reliability. After that, I tried to do better, arguing that even if these assumptions fail, fictionalism cannot explain mathematical reliability. In consequence, it cannot solve Benacerraf's reinforced challenge. Second, I distinguished two types of fictionalism: *full fictionalism* and *algebraic fictionalism*. The former claims that mathematical language is a meaningful language about abstract objects. In turn, the latter does not attribute any content to mathematical language. I argued against full fictionalism, showing that if it can explain how mathematical vocabulary acquires its content and how mathematicians acquire their linguistic competence, then so can platonism. Since the failure of platonism is the best argument for fictionalism, this is fatal for it. Finally, I argued that algebraic fictionalism is in no better position. The reduction of mathematical knowledge to logical knowledge presupposes the linguistic competence of mathematicians. But, if algebraic fictionalism tries to explain this, then it reduces to full fictionalism.

Chapter 4 was devoted to multiversism. Multiversism is very similar to fictionalism. Instead of fictions, it rescues mathematical truth by accepting the existence of several universes. Thus, Multiversism contends that there is no single universe of mathematical, abstract objects, but many of them. It is a new, pluralist form of mathematical platonism: the original mathematical realm is replaced by the multiverse. Moreover, the close connection between ontology and logic that it establishes allows it to reduce mathematical knowledge to logical knowledge. I explained that there are four main types of multiversism: on one side, height and width multiversism, on the other, realist and anti-realist forms of multiversism. The most extreme, combining high, width, and realist multiversism are Hamkinsian multiversism and full-blooded platonism: they impose no limits on the extent of the multiverse. This is why they can reduce mathematical knowledge to logical knowledge. Also, I exposed the two main arguments in favour of multiversism | apart from its ability to solve Benacerraf's challenge: the argument from independence and the argument from categoricity. As in the case of fictionalism, I challenged these two forms of multiversism with the Language Acquisition Challenge. First, I argued that Hamkinsian multiversism cannot explain how mathematical vocabulary acquires its meaning, contrary to what it says. Second, I argued that full-blooded platonism lacks an appropriate notion of consistency for explaining mathematical knowledge. The reduction of mathematical knowledge to logical knowledge presupposes the linguistic competence of mathematicians. However, full-blooded platonism fails to explain how mathematical vocabulary acquires its meaning and how mathematicians acquire their linguistic competence. Moreover, I argued that on the face of a realist ontology, the reduction of logical knowledge to mathematical knowledge imposes unreasonable restrictions on mathematical knowledge. This fulfils the objective 3.

Given the failure of fictionalism and the more extreme forms of multiversism in explaining mathemat-

ical knowledge, Chapter 5 has been devoted to the prosecution of objective 4: to develop an alternative of linguistic competence and mathematical knowledge. Thus, in this chapter, I explained how the abstractionist approach to mathematical objects advocated by (81; 82) solves the Language Acquisition Challenge and supplies us with a successful explanation of mathematical knowledge. First, I explained what abstraction principles are. These principles are equivalences linking two different domains of objects: the abstracts and their specifications. Consequently, these principles introduce a reductionist perspective in metasemantics that can be used to explain how a community of speakers enrich their linguistic resources without attributing to them any relation to the objects the language is about. As a result of a successful abstraction attempt, a community obtains a whole new language to speak about the abstracts. In this way, abstraction principles can solve the Language Acquisition Challenge without presupposing any counterfactual or logical relation between mathematicians and mathematical facts. To make sure that these principles are in good condition, I have analysed the two main problems they face: the Caesar problem and the Bad Company problem. Then, I have shown how both can be solved using the piecemeal solution and a predicate reading of quantification. Finally, I showed how abstraction principles are appropriate to explain mathematical knowledge. The abstraction process by which speakers extend their language also provides them with knowledge of a collection of truths [and not only logical truths. Thus, as (79; 82) and Rayo (116) claim, mathematical knowledge comes with the understanding of language.

Finally, the objective 5 has been achieved through the exposition of Chapters 6 and 7. In Chapter 6 I explained how abstraction principles recover enough mathematical objects to perform serious mathematical reasoning. The iteration of abstraction attempts supplies us with an explanation of the iterative hierarchy of sets. Furthermore, this theory partially interprets the potentialist version of Zermelo-Fraenkel set theory. Nevertheless, it is not powerful enough to prove the potentialist translation of the axioms of Infinity and Replacement. To recover these, additional stronger principles are required. An alternative way to these principles has been explored in Chapter 7 using abstraction principles only. Thus, the abstractionist explanations developed so far were improved. A new form of modal abstraction principle has been introduced in this chapter, showing how they are sufficient to prove the axiom of Infinity and, thus, to explain how speakers acquire the concept of actual infinite. However, the predicativist nature of modal abstraction makes the proposal difficult to generalize. Thus, the Replacement axiom cannot yet be recovered by exclusively abstractionist means.

Bibliography

- [1] Assadian, B. (2023). Linnebo on reference by abstraction. *Analytic Philosophy*, 2.
- [2] Balaguer, M. (1998). *Platonism and Anti-Platonism in Mathematics*. Oxford University Press.
- [3] Balaguer, M. (2001). A theory of mathematical correctness and mathematical truth. *Pacific Philosophical Quarterly*, 82:87–114.
- [4] Balaguer, M. (2009). Fictionalism, theft, and the story of mathematics. *Philosophia Mathematica*, 17(131-162).
- [5] Balaguer, M. (2018). Fictionalism in the philosophy of mathematics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, <https://plato.stanford.edu/entries/fictionalism-mathematics/>.
- [6] Barton, N. (2016). Multiversism and concepts of set: how much relativism is acceptable? In Boccuni, F. and Sereni, A., editors, *Objectivity, Realism, and Proof: FilMat Studies in the Philosophy of Mathematics*, pages 189–209. Springer.
- [7] Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy*, 70:661–679.
- [8] Berry, S. (2022). *A Logical Foundation for Potentialist Set Theory*. Cambridge University Press.
- [9] Boccuni, F. and Woods, J. (2020). Structuralist neologicism. *Philosophia Mathematica*, 28(3):296–316.
- [10] Boolos, G. (1971). The iterative conception of set. *Journal of Philosophy*, 68(8):215–231.
- [11] Boolos, G. (1990). The standard of equality of numbers. In Boolos, G., editor, *Meaning and Method: Essays in Honor of Hilary Putnam*, pages 261–278.
- [12] Bourgress, J. (2005). Being explained away. *The Harvard Review of Philosophy*, 13(2):41–56.
- [13] Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

- [14] Brogaard, B. and Salerno, J. (2019). Fitch's paradox of knowability. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, <https://plato.stanford.edu/entries/fitch-paradox/>.
- [15] Bueno, O. (2005). Dirac and the dispensability of mathematics. *Studies in History and Philosophy of Mod-ern Physics*, 36:465–490.
- [16] Bueno, O. (2009). Mathematical fictionalism. In Linnebo, Ø. and Bueno, O., editors, *New Waves in the Philosophy of Mathematics*, pages 59–79. Palgrave Macmillan.
- [17] Bueno, O. (2018). Putnam's indispensability argument revisited, reassessed, revived. *THEORIA. An International Journal for Theory, History and Foundations of Science*, 33(2):201–218.
- [18] Burgess, J. P. and Rosen, G. (1997). *A Subject with no Object. Strategies for Nominalistic Interpretations of Mathematics*. Oxford University Press.
- [19] Carlin Antos, Sy-David Friedman, R. H. C. T. (2015). Multiverse conceptions in set theory. *Synthese*, 192(8):2463–2488.
- [20] Carnap, R. (1950). Empiricism, semantics, and ontology. *Revue Internationale de Philosophie*, 4(11):20–40.
- [21] Cohen, P. (2008). *Set Theory and the Continuum Hypothesis*. Dover Publications Ink.
- [22] Cook, R. and Ebert, P. (2005). Abstraction and identity. *Dialectica*, 59(2):121–139.
- [23] Daston, L. (1992). Objectivity and the scape from perspective. *Social Studies of Science*, 22(4):597–618.
- [24] Daston, L. and Galison, P. (1992). The image of objectivity. *Representations*, 40:81–128.
- [25] Daston, L. and Galison, P. (2007). *Objectivity*. Princeton University Press.
- [26] Davidson, D. (1967). Truth and meaning. *Synthese*, 17(3):304–323.
- [27] Davidson, D. (1977). Reality without reference. *Dialectica*, 31(3):247–258.
- [28] Dedekind, R. (2013). *¿Qué son y para qué sirven los números? Y otros escritos sobre los fundamentos de la matemática*. Alianza.
- [29] Dorr, C. (2007). There are no abstract objects. In John Hawthorne, T. S. and Zimmerman, D. W., editors, *Contemporary Debates in Metaphysics*, pages 32–64. Blackwell.
- [30] Dummett, M. (1975). The philosophical basis of intuitionistic logic. In Dummett, M., editor, *Truth and Other Enigmas*, pages 215–247. Harvard University Press.
- [31] Echeverría, J. (2003). *La revolución tecnocientífica*. Fondo de Cultura Económica.

- [32] Eklund, M. (2006). Neo-fregean ontology. *Philosophical Perspectives*, 20(1):95–121.
- [33] Feferman, S. (2005). Predicativity. In Shapiro, S., editor, *Oxford Handbook of Philosophy of Mathematics and Logic*, pages 590–624. Oxford University Press.
- [34] Feferman, S. (2009). Conceptions of the continuum. *Intellectica*, 51:169–189.
- [35] Ferreirós, J. (2016). *Mathematical Knowledge and the Interplay of Practices*. Princeton University Press.
- [36] Ferreirós, J. (2020). La herencia oscura del logicismo. *Metatheoria*, 10(2):19–30.
- [37] Ferreirós, J. (2022a). Conceptual structuralism. *Journal for General Philosophy of Science*, 54:125–148.
- [38] Ferreirós, J. (2022b). Degrees of objectivity? mathemata and other objects. *Topoi*, 42(1):199–209.
- [39] Field, H. (1980). *Science Without Numbers*. Princeton University Press.
- [40] Field, H. (1989). *Realism, Mathematics, and Modality*. Basil Blackwell.
- [41] Fine, K. (2005). Our knowledge of mathematical objects. In Gendler, T. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, pages 89–109. Clarendon Press.
- [42] Fine, K. (2006). Relatively unrestricted quantification. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*, pages 20–44. Oxford University Press.
- [43] Fine, K. (2012). Mathematics: discovery of invention? *Think*, 11(32):11–27.
- [44] Fitch, F. (1963). A logical analysis of some value concepts. *Journal of Symbolic Logic*, 28(2):135–142.
- [45] Francesco Berto, Mark Jago, R. F. G. P. D. R. (2019). Counterimpossible conditionals. In Berto, F. and Jago, M., editors, *Impossible Worlds*, pages 267–290. Oxford University Press.
- [46] Frege, G. (1980). *The Foundations of Arithmetic*. Northwestern University Press.
- [47] García-Carpintero, M. (2022). The semantics of fiction. *Mind and Language*, 38(2):604–618.
- [Gödel] Gödel, K. Remarks before the princeton bicentennial conference on problems in mathematics. In Feferman, S., editor, *Collected Works, Volume II: Publications 1938-1974*, pages 150–153. Oxford University Press.
- [49] Gödel, K. (1947). What is cantor’s continuum problem? *The American Mathematical Monthly*, 54(9):515–525.
- [50] Gödel, K. (1995). Conference at göttingen. In Feferman, S., editor, *Collected Works, Volume III: Unpublished Essays and Lectures*, pages 127–155. Oxford University Press.

- [51] Hale, B. (2000). Reals by abstraction. *Philosophia Mathematica*, 8(2):100–123.
- [52] Hale, B. and Wright, C. (2001a). *The Reasons Proper Study*. Oxford University Press.
- [53] Hale, B. and Wright, C. (2001b). To bury caesar. In Hale, B. and Wright, C., editors, *The Reasons Proper Study*, pages 335–396.
- [54] Hamkins, J. (2011). The set-theoretic multiverse. *Review of Symbolic Logic*, 5(3):416–449.
- [55] Hellman, G. (1989). *Mathematics Without Numbers. Towards a Modal-Structural Interpretation*. Clarendon Press.
- [56] Hellman, G. and Bell, J. (2006). Pluralism and the foundations of mathematics. *Minnesota Studies in the Philosophy of Science Series*, 19:64–79.
- [57] Hilbert, D. (1984). On the infinite. In Benacerraf, P. and Putnam, H., editors, *Philosophy of Mathematics*, pages 183–201. Cambridge University Press.
- [58] Hodes, H. (1984a). Logicism and the ontological commitments of arithmetic. *The Journal of Philosophy*, 81(3):123–149.
- [59] Hodes, H. (1984b). On modal logics which enrich first-order s5. *Springer*, 13(4):423–454.
- [60] Horgan, T. and Timmons, M. (2002). Conceptual relativity and metaphysical realism. *Philosophical Issues*, 12(1):74–96.
- [61] Horsten, L. and Leitgeb, H. (2008). How abstraction works. In Hieke, A. and Leitgeb, H., editors, *Reduction, Abstraction, Analysis: Proceedings of the 31th International Ludwig Wittgenstein-Symposium in Kirchberg*, pages 2017–2026. de Gruyter.
- [62] Horwich, P. (1991). On the nature and norms of theoretical commitment. *Philosophy of Science*, 58:1–14.
- [63] Horwich, P. (1998a). *Meaning*. Oxford University Press.
- [64] Horwich, P. (1998b). *Truth*. Oxford University Press.
- [65] Isaacson, D. (2010). The reality of mathematics and the case of set theory. In Novak, Z. and Simonyi, A., editors, *Truth, Reference, and Realism*, pages 1–76. Central European University Press.
- [66] Khlentzos, D. (2021). Challenges to metaphysical realism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, <https://plato.stanford.edu/entries/realism-sem-challenge/>.
- [67] Kocurek, A. (2021). Counterpossibles. *Philosophy Compass*, 16(11).
- [68] Koellner, P. (2003). *The Search for New Axioms*. PhD thesis, Department of Linguistics and Philosophy, MIT.

- [69] Koellner, P. (2006). On the question of absolute undecidability. *Philosophia Mathematica*, 3(14):153–188.
- [70] Koellner, P. (2009). Truth in mathematics: The question of pluralism. In Bueno, O. and Linnebo, Ø., editors, *New Waves in the Philosophy of Mathematics*, pages 80–116. Palgrave Macmillan.
- [71] Kreisel, G. (1967). Informal rigor and completeness proofs. In Lakatos, I., editor, *Problems in the Philosophy of Mathematics*, pages 138–157. North-Holland.
- [72] Kripke, S. (1991). *Naming and Necessity*. Wiley-Blackwell.
- [73] Leitgeb, H. (2016). Abstraction grounded. a note on abstraction and truth. In Ebert, P. and Rossberg, M., editors, *Abstractionism: Essays in Philosophy of Mathematics*, pages 269–282. Oxford University Press.
- [74] Leitgeb, H. (2020). Why pure mathematical truths are metaphysically necessary? a set-theoretic explanation. *Synthese*, 197(7):3113–3120.
- [75] Leng, M. (2007). What’s there to know? a fictionalist approach to mathematical knowledge. In Mary Leng, A. P. and Potter, M., editors, *Mathematical Knowledge*, pages 84–108. Oxford University Press.
- [76] Leng, M. (2009). “algebraic” approaches to mathematics. In Bueno, O. and Linnebo, Ø., editors, *New Waves in the Philosophy of Mathematics*, pages 117–134. Palgrave Macmillan.
- [77] Lewis, D. (1984). Putnam’s paradox. *Australasian Journal of Philosophy*, 62(3):221–236.
- [78] Lewis, D. (1986). *On the Plurality of Worlds*. Basil Blackwell.
- [79] Linnebo, Ø. (2006). Epistemological challenges to mathematical platonism. *Philosophical Studies*, 129(3):545–574.
- [80] Linnebo, Ø. (2011). Some criteria for acceptable abstraction. *Notre Dame Journal of Formal Logic*, 52(3):331–338.
- [81] Linnebo, Ø. (2012). Metaontological minimalism. *Philosophy Compass*, 7(2):139–151.
- [82] Linnebo, Ø. (2018). *Thin Objects. An Abstractionist Approach*. Oxford University Press.
- [83] Linnebo, Ø. (2022). Plural quantification. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, <https://plato.stanford.edu/entries/plural-quant/>.
- [84] Linnebo, Ø. (2023a). Ontology and the algebraic conception of logic. In de Donato, X. and Martínez, C., editors, *Deflationist Conceptions of Abstract Objects*. Springer.

- [85] Linnebo, Ø. (2023b). Platonism in the philosophy of mathematics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, <https://plato.stanford.edu/entries/platonism-mathematics/>.
- [86] Linnebo, Ø. and Shapiro, S. (2012). Actual and potential infinity. *Noûs*, 53(1):160–191.
- [87] Linnebo, Ø. and Shapiro, S. (2020). Potentiality and indeterminacy in mathematics. *Journal of Applied Logics*, 7(6):1199–1219.
- [88] Linnebo, Ø. and Uzquiano, G. (2009). Which abstraction principles are acceptable? some limitative results. *British Journal for the Philosophy of Science*, 60(2):239–252.
- [89] Linsky, B. and Zalta, E. N. (1995). Naturalized platonism vs. platonized naturalism. *The Journal of Philosophy*, 17(10):525–555.
- [90] MacBridge, F. (2006). The julius caesar objection: More problematic than ever. In MacBridge, F., editor, *Identity and Modality*, pages 174–202.
- [91] Maddy, P. (1988a). Believing the axioms i. *Journal of Symbolic Logic*, 53(2):481–511.
- [92] Maddy, P. (1988b). Believing the axioms ii. *Journal of Symbolic Logic*, 53(3):736–764.
- [93] Maddy, P. (1990). *Realism in Mathematics*. Clarendon Press.
- [94] Maddy, P. (2011). *Defending the Axioms. On the Philosophical Foundations of Set Theory*. Oxford University Press.
- [95] Mancosu, P. (2006). *Abstraction and Infinity*. Oxford University Press.
- [96] Manzano, M. (1996). *Extensions of First-Order Logic*. Cambridge University Press.
- [97] Matteo Plebani, L. S. M. and Venturi, G. (2023). Thin objects are not transparent. *Theoria*, 89(3):314–325.
- [98] Melia, J. (2000). Weaseling away the indispensability argument. *Mind*, 109:455–479.
- [99] Parsons, C. (1965). Frege’s theory of number. In Black, M., editor, *Philosophy in America*, page 182–210. Cornell University Press.
- [100] Parsons, C. (1977). What is the iterative conception of set? In Robert E. Butts, J. H., editor, *Logic, Foundations of Mathematics, and Computability Theory*, pages 335–367. Springer.
- [101] Parsons, C. (1983). *Sets and modality*, pages 298–341. Cornell University Press.
- [102] Paseau, A. (2009). Reducing arithmetic to set theory. In Bueno, O. and Linnebo, Ø., editors, *New Waves in the Philosophy of Mathematics*, pages 35–55. Palgrave Macmillan.
- [103] Payne, J. (2013). *Expansionist Abstraction*. PhD thesis, University of Sheffield.

- [104] Payne, J. (2015). Natural deduction for modal logic with a backtracking operator. *Journal of Philosophical Logic*, 44(3):237–258.
- [105] Priest, G. (2008). *Introduction to Non-Classical Logic. From If to Is*. Cambridge University Press.
- [106] Putnam, H. (1967). Mathematics without foundations. *Journal of Philosophy*, 64(1):5–22.
- [107] Putnam, H. (1975). Is semantics possible? In *Mind, language and reality. Philosophical Papers, Vol. 2*, pages 139–152. Cambridge University Press.
- [108] Putnam, H. (1979a). The meaning of ‘meaning’. In *Philosophical Papers, Vol. 2: Mind, Language and Reality*, pages 215–271. Cambridge University Press.
- [109] Putnam, H. (1979b). Truth and necessity in mathematics. In *Mathematics, Matter and Method*, pages 1–11. Cambridge University Press.
- [110] Putnam, H. (1979c). What is mathematical truth? In *Mathematics, Matter, and Method*, pages 60–78. Cambridge University Press.
- [111] Putnam, H. (1980). Models and reality. *The Journal of Symbolic Logic*, 45(3):464–482.
- [112] Putnam, H. (1983). Vagueness and alternative logic. In *Philosophical Papers, Volume 3*, pages 271–286. Cambridge University Press.
- [113] Putnam, H. (2001). *The Threefold Chord. Mind, Body, and World*. Columbia University Press.
- [114] Quine, W. V. (1986). *Philosophy of Logic*. Harvard University Press.
- [115] Rayo, A. (2009). *Towards a trivialist account of mathematics*, pages 239–260. Palgrave-Macmillan.
- [116] Rayo, A. (2013). *The Construction of Logical Space*. Oxford University Press.
- [117] Rayo, A. (2014). Nominalism, trivialism, logicism. *Philosophia Mathematica*, 23(1):65–86.
- [118] Reiss, J. and Sprenger, J. (2020). Scientific objectivity. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, University of Stanford, <https://plato.stanford.edu/entries/scientific-objectivity/>.
- [119] Resnik, M. (1997). *Mathematics as a Science of Patterns*. Oxford University Press.
- [120] Rosen, G. (2010). Metaphysical dependence: grounding and reduction. In Hale, B. and Hoffmann, A., editors, *Modality: Metaphysics, Logic, and Epistemology*, pages 109–135. Oxford University Press.
- [121] Rosen, G. (2011). The reality of mathematical objects. In Polkinghorne, J., editor, *Meaning in Mathematics*, pages 113–131. Oxford University Press.

- [122] Rosen, G. and Yablo, S. (2020). Solving the caesar problem -with metaphysics. In Miller, A., editor, *Logic, Language, and Mathematics: Themes from the Philosophy of Crispin Wright*. Oxford University Press.
- [123] Sam Baron, Mark Colyvan, D. R. (2017). How mathematics can make a difference. *Philosophers' Imprint*, 17(3):1–29.
- [124] Sam Baron, Mark Colyvan, D. R. (2020). A counterfactual approach to explanation in mathematics. *Philosophia Mathematica*, 28(1):1–34.
- [125] Shapiro, S. (2000). *Foundations without Foundationalism. A case for Second-Order Logic*. Oxford University Press.
- [126] Shelah, S. (2003). Logical dreams. *Bulletin of the American Mathematical Society*, 40(2):203–228.
- [127] Shoenfield, J. (1967). *Mathematical Logic*. Addison-Wesley.
- [128] Skolem, T. (1967). Some remarks on axiomatized set theory. In van Heijenoort, J., editor, *From Frege to Gödel*, page 290–301. Harvard University Press.
- [129] Smith, P. (2013). *An Introduction to Gödel's Theorems*. Cambridge University Press.
- [130] Speaks, J. (2019). Theores of meaning. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, <https://plato.stanford.edu/entries/meaning/>.
- [131] Stalnaker, R. (1987). *Inquiry*. The MIT Press.
- [132] Steel, J. R. (2014). *Gödel's program*, pages 153–179. Cambridge University Press.
- [133] Stillwell, J. (2019). *A Concise History of Mathematics for Philosophers*. Cambridge University Press.
- [134] Studd, J. (2013). The iterative conception of set: A (bi-) modal axiomatization. *Journal of Philosophical Logic*, 42(5):697–725.
- [135] Studd, J. (2023a). The caesar problem. *Philosophia Mathematica*, 31(2):236–267.
- [136] Studd, J. (2023b). Linnebo's abstractionism and the bad company problem. *Theoria*, 89(3):366–392.
- [137] Tait, W. (1990). The iterative hierarchy of sets. *The Jerusalem Philosophical Quaterly*, 39:65–79.
- [138] Tait, W. (2005). *The Provenance of Pure Reason. Essays in the Philosophy of Mathematics and its History*. Oxford University Press.
- [139] Takeuti, G. (2003). *Memoirs of a Proof Theorist*. World Scientific Publishing Company.

- [140] Tarski, A. (1956). The concept of truth in formalized languages. In Tarski, A., editor, *Logic, Semantics, and Metamathematics*, pages 152–278. Oxford University Press.
- [141] Thomason, A. (2018). *Ontology Made Easy*. Oxford University Press.
- [142] Úbeda, J. P. (2019). Teoría de conjuntos.
- [143] Weir, A. (2003). Neo-fregeanism: an embarrassment of riches. *Notre Dame Journal of Formal Logic*, 44(1):13–48.
- [144] Wiggins, D. (1980). What would be a substantive theory of truth. In Straaten, Z. V., editor, *Philosophical Subjects*, pages 189–221. Oxford University Press.
- [145] Williamson, T. (1992). Inexact knowledge. *Mind*, 101(402):217–242.
- [146] Williamson, T. (1995). Does assertibility satisfy the s4 axiom? *Critica*, 81(3):3–25.
- [147] Williamson, T. (1996a). Knowing and asserting. *The Philosophical Review*, 105(4):489–523.
- [148] Williamson, T. (1996b). Unreflective realism. *Philosophy and Phenomenological Research*, 56(4):905–909.
- [149] Wittgenstein, L. (1953). *Philosophical Investigations*. MacMillan.
- [150] Wright, C. (1983). *Frege’s Concept of Number as Object*. Aberdeen University Press.
- [151] Wright, C. (1994). *Truth and Objectivity*. Harvard University Press.
- [152] Wright, C. (2001). On the philosophical significance of frege’s theorem. In Hale, B. and Wright, C., editors, *The Reasons Proper Study*, page 272–306. Oxford University Press.
- [153] Yablo, S. (2012). Explanation, extrapolation, and existence. *Mind*, 121:1007–1029.
- [154] Zermelo, E. (1986). *On Boundary Numbers and Domains of Sets: New Investigations in the Foundations of Set Theory*, pages 1219–1233. Oxford Clarendon Press.

List of Figures

Fig. 6.1	Domain expansion: the domain $\mathcal{D}_1$ expands the domain $\mathcal{D}_0$ by adding abstracts specified by the elements in this domain.	179
Fig. 6.2	The iterative hierarchy of sets.	185
Fig. 6.3	The plural hierarchy of sets. The first-order domains $\mathcal{D}_\alpha$ contain sets represented by black dots. Pluralities of sets are represented by blue circles. According to PLV, each plurality is mapped by an arrow to a set. The origin of the hierarchy is the empty plurality. The domain $\mathcal{D}_\omega$ contains all sets generated at finite stages. From there, the hierarchy enters the transfinite.	189
Fig. 6.4	The plural hierarchy of finite sets.	199

The objectivity of mathematics has been questioned for two reasons: Benacerraf's challenge and the emergence of several limitation theorems in mathematics. This has led to the search for new explanations of mathematical knowledge and alternatives to universalist platonism. Mainly, mathematical knowledge has been reduced to logical knowledge. Against this, this dissertation argues that mathematical knowledge cannot be reduced to logical knowledge. The two best proposals defending this reduction, fictionalism and multiversism, run into serious problems. Both will be challenged to explain how mathematical language acquires its content. Then, it will be argued that they do not satisfy this test. Faced with this problem, an alternative explanation of mathematical language and mathematical knowledge will be developed using abstraction principles. As a result, mathematical knowledge will be explained through linguistic competence. Finally, the proposal will be applied to ZF set theory.