
SCUOLA NORMALE SUPERIORE
UNIVERSITE PARIS DIDEROT

Tesi di perfezionamento in discipline filosofiche

Thèse de doctorat en Sciences mathématiques

**Contributions towards the
generalization of Forcing Axioms**

Giorgio Venturi

Relatori:

Gabriele LOLLI

Boban VELICKOVIC

Referees:

Anno accademico 2013-2014

Contents

Introduzione	5
0.1 Sommario	7
Introduction	11
0.2 Sommaire	13
Introduction	17
0.3 Summary	19
1 The concept of axiom in Hilbert's thought	23
1.1 Two concepts	25
1.1.1 The first period	26
1.1.2 The second period	32
1.2 Hilbert and intuition	37
1.2.1 First period	38
1.2.2 Completeness	44
1.2.3 Second period	61
1.3 Hilbert's position(s) now	67
1.3.1 Hilbert's program and Gödel's position	68
1.3.2 Hilbert's position and Gödel's program	70
1.3.3 Idea of completeness and contemporary axiomatics	73
2 Set theoretical foundations of mathematics	78
2.1 Set theoretical foundation as unity	79
2.1.1 Hilbertian origins	80
2.1.2 Practical reasons	83
2.1.3 Example of sufficient conditions	85
2.1.4 Examples of equivalence and equiconsistent results	88
2.1.5 Necessary and sufficient conditions	89
2.2 Axioms as explanations	95

2.2.1	Applying Kitcher's account?	98
2.2.2	Kitcher's problem for mathematical explanation	104
2.2.3	Towards a more general distinction	111
2.3	Two different foundational ideas	112
2.3.1	Theoretical foundation	113
2.3.2	Practical foundation	115
3	Naturalness in mathematics	120
3.1	How to deal with this concept?	122
3.2	A tentative methodology	131
3.3	Applying the method	136
3.4	Case studies	140
3.4.1	The concept of set	141
3.4.2	Natural new axioms for ZFC	147
3.4.3	Dynamism and the missing ingredient: normativism	148
3.5	Philosophy in mathematical clothing	150
4	The case of Forcing Axioms	153
4.1	A realist stand	154
4.2	Forcing Axioms as unifying axioms	158
4.3	Forcing Axioms as natural axioms	161
5	The side conditions	164
5.1	Some technical background	164
5.2	The pure side conditions poset \mathbb{M}_θ^1	165
5.3	Strong properness of \mathbb{M}_θ^1	167
6	PFA and some of its consequences	169
6.1	Consequences of PFA	169
6.2	Forcings associated to OCA and PID	171
6.2.1	The poset for OCA	171
6.2.2	The poset for PID	172
6.3	Verification of properness	173
6.4	Density arguments	176
6.5	Some application of PID	177
7	The five element basis theorem	180
7.1	Saturation of Aronszajn trees	181
7.2	Colorings of Aronszajn trees	183
7.3	The forcing $\partial^*(K)$	189

7.4	The main theorem	192
8	Generalized side conditions	194
8.1	The forcing \mathbb{M}	195
8.2	Adding a club in ω_2 with finite conditions	199
8.3	Strong chains of uncountable functions	208
8.4	Thin very tall superatomic Boolean algebras	218
8.5	Souslin trees	222
9	Suslin trees and side conditions	227
9.1	Souslin trees and properness	227
9.2	Preservation of T and countable models	228
9.3	$\text{PFA}(T)$ with finite conditions	231
	Conclusions	239
9.4	A different generalization	240

Introduzione

Vi sono diverse finalità in questa tesi, molte delle quali specifiche e che verranno presentate nel sommario di questo lavoro, ma vi è anche una ambizione di più ampio respiro: tenere unite una riflessione filosofica e una riflessione matematica, senza dover sacrificare il livello di precisione di nessuna delle due. Questa aspirazione non ha motivazioni soltanto personali, ma origina dalla convinzione che gli strumenti matematici e quelli filosofici siano complementari per lo studio della matematica in generale e - in misura anche maggiore - della logica matematica in particolare. Nell'affrontare questo compito, però, vi son problemi di natura sia soggettiva, sia oggettiva.

Per quanto riguarda il lato soggettivo, devo ammettere che non è sempre facile mantenere lo stesso livello di chiarezza in entrambe le discipline e, allo stesso tempo, produrre un lavoro che possa interessare studiosi sia con interessi filosofici, sia matematici. Inoltre, sul versante delle difficoltà oggettive, anche se propongo un approccio interdisciplinare alla logica, sono cosciente che la filosofia e la matematica, per quanto abbiano interessi comuni, hanno tuttavia metodologie differenti. L'aspetto più significativo di questo fenomeno è il diverso stile di argomentare che varia profondamente tra due discipline. Ma potremmo anche citare la rilevanza della letteratura su un certo argomento, o i differenti modi di porsi rispetto all'autonomia della propria disciplina, o la relazione tra un punto di vista locale e uno più globale.

Tuttavia credo che questi aspetti non rappresentino una difficoltà, quanto piuttosto una complementarità della filosofia e della matematica in un ambito, come la logica matematica, in cui si mischiano componenti formali e informali. Infatti sono convinto che la matematica non possa essere ridotta ad un puro gioco di simboli, che ne escluda ogni elemento informale, e nemmeno che la filosofia debba evitare il ricorso a strumenti formali nei suoi argomenti. Tuttavia bisogna evitare che la filosofia segua la matematica troppo da vicino, sul terreno delle componenti formali. Infatti, come alcuni aspetti della filosofia analitica mostrano, questo rischierebbe di restringere eccessivamente l'ambito d'indagine; con la conseguente perdita di alcuni problemi genuinamente filosofici, che pos-

sono sorgere in un ambito formale.

La tensione tra aspetti formali ed informali, in matematica, ha recentemente dato vita alla filosofia della pratica matematica. Questa branca della filosofia ha come presupposto il riconoscimento che nella pratica matematica si possano trovare considerazioni estranee al suo carattere formale e che queste possano essere trattate in maniera filosofica. Quindi non sorprenderà che il lavoro presentato in questa tesi abbia notevoli punti di convergenza con questa posizione filosofica. Tuttavia credo che - almeno in ambito logico - un passo ulteriore sia necessario: da un'accurata analisi del lavoro di un matematico possono emergere problemi genuinamente filosofici. Questa considerazione è motivata dal peculiare carattere della logica contemporanea, che ha sia un carattere deduttivo, sia descrittivo. Il primo permette di formalizzare le nostre regole di inferenza, mentre il secondo dà la possibilità di connettere considerazioni informali e strumenti formali. Di conseguenza, la stessa possibilità di stabilire una connessione tra elementi del discorso presentati in una maniera matematicamente rigorosa e idee informali genera problematiche di non facile soluzione, che necessitano di una trattazione filosofica in un contesto puramente matematico. Come vedremo, questa convinzione distanzia il lato filosofico di questo lavoro da molte forme di naturalismo, sia da un punto di vista teorico, sia da un punto di vista metodologico.

Dal momento che sono convinto che la matematica non si esaurisca nella sua componente formale e che per riconoscere gli aspetti filosofici di un problema matematico sia necessario comprenderne i dettagli, credo dunque che il modo migliore per proporre un'analisi filosofica della matematica sia prendere parte personalmente allo sviluppo di quest'ultima. Di conseguenza questa tesi combina insieme risultati matematici e riflessioni filosofiche, sperando che entrambe le discipline siano favorite dalla reciproca interazione.

Tuttavia sia la filosofia, sia la matematica, in quanto attività umane, sono organizzate compatibilmente con una comunità scientifica: in generali aree di interesse, in gruppi di ricerca e in scuole. Questo aspetto è un'altra problema di cui bisogna tenere conto nel momento in cui si propone una ricerca di carattere interdisciplinare. Nel tentativo di trovare un ambito sufficientemente stabile per sviluppare un tale lavoro, ho cercato quindi un tema classico, che avesse le sue origini nel tempo in cui la filosofia e la matematica non fossero così distinte come oggi. Questo tema è, in generale, la nozione di assioma, in particolare, la nozione di assioma nel contesto della teoria degli insiemi e, ancora più nello specifico, i problemi legati alla giustificazione di nuovi assiomi nel contesto di questa teoria.

Infatti questo tema ha sia un aspetto filosofico, sia un aspetto matematico: per un verso ci si può chiedere quali ragioni abbiamo per accettare un nuovo

assioma, dal momento che la sua stessa natura lo rende indimostrabile; dall'altro ci si può chiedere cosa si possa dimostrare a partire da un nuovo assioma. Come vedremo, gli ambiti di queste due domande non sono così nettamente separati: per quanto riguarda la prima argomenterò in favore di un carattere dimostrativo del problema della giustificazione di un assioma; mentre per quanto riguarda la seconda seguirò un approccio induttivo: dalle conseguenze agli assiomi. Infatti presenterò un metodo generale la cui soluzione uniforme di differenti problemi costituirà un argomento a favore della proposta di un nuovo assioma, capace di catturarne l'aspetto combinatorio.

E' arrivato dunque il momento di rendere più concreto questo discorso e di delineare la struttura della tesi.

0.1 Sommario

Questo lavoro è diviso in due parti. I primi quattro capitoli sono principalmente filosofici, mentre i restanti, insieme alla conclusione, soprattutto matematici. Questa distinzione è fatta in modo da aiutare il lettore. Tuttavia è importante tenere a mente che le due parti fanno parte dello stesso ragionamento, che perderebbe in concretezza senza la parte matematica; mentre, omettendo la parte filosofica, sarebbe difficile capire e contestualizzare i risultati matematici.

Il Capitolo 1 verte sull'origine del concetto di assioma, così come viene inteso nel contesto assiomatico contemporaneo. Quando ho cominciato a lavorare al problema della giustificazione degli assiomi in teoria degli insiemi, ho sentito la necessità di analizzare in dettaglio il concetto stesso di assioma. Infatti l'analisi di quest'ultimo non viene spesso affrontata direttamente, poiché esso viene considerato considerato, per quanto problematico, sufficientemente chiaro. Inoltre credo che una seria riflessione filosofica di un problema coinvolga anche un'indagine delle sue radici storiche. Questa convinzione mi ha quindi portato ad analizzare il concetto di assioma nel pensiero di Hilbert. In questo compito ho incontrato due difficoltà: l'assenza di una vasta letteratura su questo tema specifico - a paragone di quella che esiste su Hilbert e l'assiomatica - e, inoltre, la concezione classica di Hilbert come campione del formalismo. I due problemi sono intimamente legati, poiché il fatto di concentrare l'attenzione sul metodo assiomatico spesso mette in ombra l'origine e la natura della nozione di assioma, mantenendo l'analisi al livello dei sistemi formali e quindi distorcendo la rilevanza che le componenti informali hanno nel lavoro di Hilbert. Il principale risultato di questa analisi sarà che, nel lavoro di Hilbert, si possono trovare almeno due distinte nozioni di assioma e che entrambe sono legate a due diverse nozioni di intuizione. La differenza tra queste ultime consiste nel ruolo

che l'evidenza gioca rispetto al legame che viene instaurato tra gli assiomi e ciò che viene formalizzato. Penso che Hilbert fosse al corrente delle difficoltà di giustificare un legame tra le gli aspetti formali e quelli informali del lavoro matematico e, inoltre, una delle tesi di questo capitolo sarà che il tentativo di trovare una soluzione a questo problema spingerà Hilbert a formulare, prima, l'Assioma di Completezza, e, poi, la teoria della dimostrazione. Anche se l'analisi storica è interessante di per sé, e necessaria per meglio comprendere le riflessioni di Hilbert, mostrerò che è anche istruttiva nel contesto della giustificazione dei nuovi assiomi in teoria degli insiemi. Infatti verso la fine del capitolo riallacerò l'analisi storica con la riflessione teorica, proponendo alcuni criteri per accettare una specifica classe di assiomi, che chiamerò *sufficienti*: assiomi che agiscono come condizioni sufficienti per la formalizzazione di un ambito del sapere.

Nel Capitolo 2 cerco di indagare gli effetti del ruolo fondazionale della teoria degli insiemi sul problema della giustificazione dei suoi assiomi. Nello specifico propongo di considerare la teoria degli insiemi come una teoria capace di caratterizzare la possibilità di effettuare una dimostrazione matematica e quindi capace di caratterizzare e unificare la matematica: una fondazione insiemistica della pratica matematica. Sosterrò questa tesi con diversi esempi da differenti ambiti: algebra, analisi funzionale e teoria degli insiemi. Inoltre argomenterò che questo punto di vista pratico sulla matematica è anche utile per la rilevanza della teoria degli insiemi nel contesto della filosofia della pratica matematica. Infatti credo che alcuni assiomi della teoria degli insiemi possano fornire delle reali spiegazioni di alcuni fatti matematici. Nel contesto di una discussione sulla spiegazione in matematica, analizzerò principalmente la teoria di Kitcher sull'unificazione scientifica, collegando quest'ultima con il ruolo esplicativo che alcuni assiomi della teoria degli insiemi manifestano, in un contesto fondazionale. In particolare discuterò alcuni risultati di Woodin e Viale, dove i temi della spiegazione e della giustificazione sono legati insieme, nel contesto del ruolo unificante della teoria degli insiemi. Questo mi permetterà di proporre un'analisi filosofica della nozione di unificazione, che porterà alla formulazione di un corrispondente criterio per la scelta di un assioma. Verso la fine di questo capitolo amplierò l'analisi, proponendo una distinzione tra due diverse attitudini nei lavori fondazionali: una pratica e una teorica. Questa distinzione servirà a chiarire e distinguere - legittimandoli così entrambe - due diversi approcci fondazionali: una visione che accetta l'esistenza di una pluralità di universi della teoria degli insiemi, opposta a quella che mira a trovare il vero e unico modello della teoria degli insiemi.

Nel Capitolo 3 cercherò di chiarire il concetto di naturalezza. Questa nozione è spesso utilizzata in matematica e interviene abitualmente nei dibattiti sui criteri per la giustificazione dei nuovi assiomi in teoria degli insiemi. Cercherò

di non affrontare direttamente la questione della naturalezza degli assiomi, ma cercherò di comprendere l'uso di questo termine, in un generale contesto matematico. L'analisi partirà dall'evidenza statistica che il ricorso a componenti naturali del discorso matematico è notevolmente aumentato negli ultimi decenni. Di conseguenza, un'altra importante domanda a cui cercherò di rispondere in questo capitolo sarà: “perché i matematici chiamano qualcosa naturale?”; sperando che una risposta a quest'ultima possa gettare luce sull'evidenza statistica riscontrata nell'uso del termine “naturale”. Una grande parte di questo capitolo sarà dedicata a discutere la metodologia di indagine, poiché su questo argomento specifico e con questa prospettiva, non è possibile appoggiarsi su contributi di altri. Un aspetto teorico rilevante di questo studio, che informerà la metodologia proposta, sarà una critica del naturalismo, in particolare quello proposto da Maddy. A differenza di questa forma di naturalismo, sono convinto che esistano genuini problemi filosofici nel lavoro matematico. Inoltre, l'analisi della naturalezza servirà a svelare una scelta realista nel suo uso. Infatti la principale tesi di questo capitolo è che il riferimento alla natura, implicito nel riferimento a componenti naturali della matematica, nasconde un tentativo di rendere concettualmente stabile un fenomeno che, al contrario, ha caratteristiche dinamiche e normative: la relazione tra gli aspetti formali e quelli informali di un lavoro matematico. Una versione più lunga di questo capitolo consiste in due articoli, scritti in collaborazione con Luca San Mauro.

Nel Capitolo 4 riepilogherò tutti i criteri discussi e proposti nei capitoli precedenti e li testerò nel caso particolare degli Assiomi di Forcing. Allo stesso tempo discuterò le idee generali che sono alla base della moderna concezione della teoria degli insiemi. In particolare discuterò la nozione di insieme arbitrario, collegando quest'ultima a quella di insieme generico. In conclusione poi argomenterò in favore della naturalezza degli Assiomi di Forcing.

A partire dal Capitolo 5 inizia la parte matematica di questo lavoro. In questo capitolo presenterò prevalentemente il metodo delle *side conditions*, nella sua forma più nota. Questo metodo consiste nell'utilizzo di modelli della teoria degli insiemi come parte integrante della definizione delle nozioni di forcing ed il suo obiettivo è quello di preservare alcuni cardinali. Ricapitolerò le principali definizioni, preparando il terreno per i capitoli successivi.

Nel Capitolo 6 riporto le note che ho scritto, in collaborazione con Justin Moore, di un breve corso sugli Assiomi di Forcing, da lui tenuto allo “Young set theory workshop” a Raach (vicino a Vienna), nel 2010. Credo che queste note siano una buona e sintetica presentazione di questi assiomi. Inoltre utilizzano il metodo delle *side conditions* in connessione con gli Assiomi di Forcing.

Nel Capitolo 7 include il lavoro *A direct proof of the five element basis theorem* ([180]) di Veličković, poiché da un lato è un esempio interessante dell'uso delle

side conditions in congiunzione con gli Assiomi di Forcing, mentre dall'altro ho contribuito alla sua preparazione. Questo teorema è stato dimostrato per la prima volta da Justin Moore, nel 2006, e poi semplificato da König, Larson, Moore e Veličković. Questo lavoro consiste in un'ulteriore semplificazione della dimostrazione di questo teorema.

Nel Capitolo 8 includo (nelle prime quattro sezioni) le note del seminario *Proper forcing remastered* tenuto da Veličković in occasione dell' "Appalachian Set Theory Workshop", tenuto il 5 ottobre 2011 presso l'University of Illinois at Chicago. Come nel caso del capitolo 7 ho contribuito alla preparazione di queste note. In esse viene presentato ed applicato il metodo delle *side conditions*, in linea con il lavoro di Itay Neeman, che per primo ha proposto, nel 2011, di considerare come *side conditions* delle \in -catene, chiuse per intersezione, di modelli di due cardinalità differenti. Le prime quattro sezioni consistono nella presentazione dell'ordine parziale formato dalle sole *side conditions* di due tipi di modelli e l'applicazione di questo nuovo metodo al problema di come forzare un insieme chiuso e illimitato in ω_2 , con condizioni finite, dell'esistenza di una catena di lunghezza ω_2 , in $(\omega_1^{\omega_1}, <_{Fin})$, dell'esistenza di una *Thin Tall Superatomic Boolean Algebra*. Queste note sono state pubblicate, in una forma leggermente diversa, come [181]. Nella sezione 5, seguendo una proposta di Veličković, applico le *side conditions* al problema dell'esistenza di un albero ω_2 -Suslin.

Nel Capitolo 9, seguendo una proposta di Veličković applico il metodo delle *side conditions* nella dimostrazione di consistenza di $PFA(T)$: l'Assioma di Forcing per la classe di ordini parziali che sono propri e che preservano un albero di Souslin T . La sezione 2 è basata su di un'idea di Veličković, mentre il principale teorema della sezione 3 - dove mostro che forzando $PFA(T)$ non viene cambiato il fatto che T è Souslin - segue da vicino la dimostrazione di consistenza di PFA di Neeman.

Introduction

Cette thèse poursuit plusieurs objectifs, dont la majeure partie sont spécifiques et seront développés dans le sommaire, mais cette thèse poursuit également une ambition plus grande: maintenir unies une réflexion philosophique et une réflexion mathématique, sans sacrifier la précision de l'une ou de l'autre. Cette aspiration n'est pas seulement motivée par une ambition personnelle, mais par la conviction que les instruments mathématiques et philosophiques sont complémentaires dans l'étude des mathématiques en général et - dans une plus grande mesure encore - de la logique mathématique en particulier. Pour mener à bien ce projet, nous affrontons des difficultés aussi bien subjectives qu'objectives.

En ce qui concerne l'aspect subjectif, je dois admettre qu'il n'est pas toujours facile de maintenir le même niveau de clarté dans les deux disciplines et, dans le même temps, produire un travail qui puisse intéresser des spécialistes en mathématiques ou en philosophie. En outre, en ce qui concerne l'aspect objectif, même si je propose une approche interdisciplinaire à la logique mathématique, je suis conscient que la philosophie et les mathématiques, même si elles ont des points communs, conservent des méthodologies différentes. La dimension la plus significative de ce phénomène est le style d'argumentation, qui varie profondément entre les deux disciplines. Mais nous pourrions aussi citer l'importance de la littérature sur un sujet particulier ou les différentes façons de se situer par rapport à l'autonomie d'une des deux disciplines, ou la relation entre un point de vue local et un autre plus global.

Cependant, je crois que ces aspects ne représentent pas une difficulté, mais plutôt une complémentarité entre la philosophie et les mathématiques dans un domaine, comme la logique mathématique, où se mêlent des composantes formelles et informelles. En effet, je suis convaincu que les mathématiques ne peuvent pas être réduites à un pur jeu de symboles, qui exclut tout élément informel, tout comme je suis convaincu que la philosophie ne doit pas s'interdire le recours à des instruments formels dans ses argumentations. Toutefois, il faut éviter que la philosophie ne suive de trop près les mathématiques sur le terrain des composantes formelles. En effet, comme le montrent certains aspects de la

philosophie analytique, cela risquerait de restreindre excessivement le champ de recherche, avec la perte corollaire de certaines problématiques substantielles à la philosophie.

La tension entre des aspects formels et informels, en mathématiques, a récemment donné vie à la philosophie de la pratique mathématique. Cette branche de la philosophie a comme présupposé la reconnaissance que, dans la pratique mathématique, nous puissions trouver des considérations étrangères à son caractère formel et que celles-ci puissent être traitées de façon philosophique. Ainsi, il ne sera pas surprenant que le travail présenté dans cette thèse ait de considérables points de convergence avec cette position philosophique. Cette considération est motivée par le caractère particulier de la logique mathématique contemporaine, qui a soit un caractère déductif, soit descriptif. Le premier permet de formaliser nos règles d'inférence alors que le second donne la possibilité d'établir une connexion entre des éléments du discours présentés d'une manière mathématiquement rigoureuse et des idées informelles. Cette dernière possibilité génère des problématiques, difficiles à résoudre, et qui nécessitent un traitement philosophique dans un contexte purement mathématique. Comme nous le verrons, cette conviction écarte la dimension philosophique de cette thèse de nombreuses formes de naturalisme, d'un point de vue théorique comme d'un point de vue méthodologique.

Dans la mesure où je suis convaincu que la mathématique ne se limite pas à sa composante formelle et que pour reconnaître les aspects philosophiques d'un problème mathématique il soit nécessaire d'en comprendre les détails, je crois que le meilleur moyen pour proposer une analyse en philosophie des mathématiques soit en prenant part personnellement au développement de cette dernière. En conséquence, cette thèse combine des résultats mathématiques et des réflexions philosophiques, en espérant que les deux disciplines bénéficient de cette interaction réciproque.

Cependant, que ce soit la philosophie ou les mathématiques, appréhendées comme des activités humaines, elles sont organisées de façon être compatible avec une communauté scientifique : en aires d'intérêts général, groupes de recherche et écoles. Cet aspect est un autre problème dont nous devons tenir compte à partir du moment où est proposée une recherche de caractère interdisciplinaire. Dans la tentative de trouver un cadre suffisamment stable pour développer un tel travail, j'ai cherché un thème classique, qui ait ses origines dans une période où la philosophie et les mathématiques n'étaient pas distinctes comme elles le sont aujourd'hui. Ce thème est, en général, la notion d'axiome, en particulier, la notion d'axiome dans le contexte de la théorie des ensembles, et encore plus spécifiquement, les problèmes liés à la justification de nouveaux axiomes dans le contexte de cette théorie.

En effet, ce thème a un aspect philosophique comme un aspect mathématique: d'une part, on peut se demander quelles raisons nous avons pour accepter un nouvel axiome, à partir du moment où sa nature le rend indémontrable; d'autre part on peut se demander ce qu'on peut démontrer à partir d'un nouvel axiome. Comme nous le verrons, les domaines de ces deux questions ne sont pas vraiment nettement séparés: en ce qui concerne la première, j'argumenterai en faveur d'un caractère démonstratif du problème de la justification d'un axiome; alors que, en ce qui concerne la seconde, je suivrai une approche inductive: des conséquences à l'axiome. En effet, je présenterai une méthode générale dont la solution uniforme à différents problèmes constituera un argument en faveur de la proposition d'un nouvel axiome, capable d'en capturer l'aspect combinatoire.

Il est temps de rendre plus concret mon propos et de présenter la structure de ma thèse.

0.2 Sommaire

Cette thèse est divisée en deux parties. Les quatre premiers chapitres sont principalement philosophiques, les suivants, ainsi que la conclusion, surtout d'ordre mathématique. Cette distinction est faite pour aider le lecteur. Il est important de garder cependant à l'esprit que les deux parties participent au même raisonnement, qui perdait en concret sans la partie mathématique ; alors que, en omettant la partie philosophique, il serait difficile de comprendre et de contextualiser les résultats mathématiques.

Il Chapitre 1 interroge les origines du concept d'axiome, comme il est compris dans le contexte axiomatique contemporain. Quand j'ai commencé à travailler sur le problème de la justification des axiomes dans la théorie des ensembles, j'ai senti la nécessité d'analyser en détail le concept même d'axiome. En effet, l'analyse de ce dernier n'est pas souvent faite directement, car celle-ci est considérée suffisamment claire, ce qui est en soi problématique. En outre, je crois qu'une réflexion philosophique sérieuse sur un problème doit comprendre également une enquête sur ses racines historiques. Cette conviction m'a porté à analyser le concept d'axiome dans la réflexion d'Hilbert. Dans mon travail, j'ai rencontré deux difficultés : l'absence d'une vaste littérature sur ce sujet ? en comparaison avec celle qui existe sur Hilbert et l'axiomatisme- et, en outre, la conception classique d'Hilbert comme champion du formalisme. Les deux problèmes sont intimement liés, puisque que le fait de concentrer l'attention sur la méthode axiomatique rejette souvent dans l'ombre l'origine et la nature de la notion d'axiome, maintenant l'analyse au niveau des systèmes formels et ainsi distordant l'intérêt que les composantes informelles ont dans le travail d'Hilbert.

Le résultat principal de cette analyse sera que, dans le travail d'Hilbert, on peut trouver au moins deux notions distinctes d'axiome et que les deux sont liées à deux notions différentes d'intuition. La différence entre ces deux dernières tient dans le rôle que l'évidence joue par rapport au lien instauré entre les axiomes et ce qui est formalisé. Je pense qu'Hilbert était au courant des difficultés de justifier un lien entre les aspects formels et informels du travail mathématique et, en outre, une des thèses de ce chapitre traitera de la tentative de trouver une solution à ce problème qui poussera Hilbert à formuler, d'abord, l'Axiome de Complétude, et ensuite, la théorie de la démonstration. Même si l'analyse historique est intéressante en soi, et nécessaire pour comprendre mieux les réflexions d'Hilbert, je montrerai que, il est également instructif dans le contexte de la justification de nouveau axiome dans la théorie des ensembles. En effet, vers la fin du chapitre, je rattacherai l'analyse historique avec la réflexion théorique, en proposant quelques critères pour accepter une classe spécifique d'axiome, que j'appellerai suffisants : axiomes qui agissent comme des conditions suffisantes pour la formalisation d'un environnement du savoir.

Dans le Chapitre 2, je chercherai de comprendre les effets du rôle fondateur de la théorie des ensembles dans la justification de ses axiomes. Plus spécifiquement, je propose de considérer la théorie des ensembles comme une théorie capable de caractériser la possibilité d'effectuer une démonstration mathématique et donc capable de caractériser et unifier les mathématiques, soit un fondement de la pratique mathématique. Je soutiendrai cette thèse avec de nombreux exemples de plusieurs domaines : algèbre, analyse fonctionnelle et théorie des ensembles. En outre, j'argumenterai que ce point de vue pratique sur les mathématiques est aussi utile pour l'importance de la théorie des ensembles dans le contexte de la philosophie de la pratique mathématique. En effet, je crois que certains axiomes de la théorie des ensembles peuvent fournir des explications réelles de certains faits mathématiques. Dans le contexte d'une discussion sur l'explication en mathématique, j'analyserai principalement la théorie de Kicher sur l'unification scientifique, en liant cette dernière avec le rôle explicatif que certains axiomes de la théorie des ensembles manifestent, dans un contexte fondateur. En particulier, je discuterai certains résultats obtenus par Woodin et Viale, où les thèmes de l'explication et la justification sont liés, dans le contexte du rôle unifiant de la théorie des ensembles. Ceci me permettra de proposer une analyse philosophique de la notion d'unification, qui portera à la formulation d'un critère correspondant pour le choix d'un axiome. Vers la fin de ce chapitre, j'amplifierai l'analyse, en proposant une distinction entre les deux attitudes dans les travaux fondateurs : une vision qui accepte l'existence d'une pluralité des univers de la théorie des ensembles, opposées à celle qui vise à trouver le vrai et unique modèle de la théorie des ensembles.

Dans le Chapitre 3, je chercherai à éclaircir le concept de naturalité. Cette notion est souvent utilisée en mathématique et intervient habituellement dans les débats sur les critères pour la justification des nouveaux axiomes dans la théorie des ensembles. Je chercherai à ne pas affronter directement la question de la naturalité des axiomes, mais je chercherai à comprendre l'utilisation de ce terme, dans un contexte général mathématique. L'analyse partira de l'évidence statistique que le recours à des composantes naturelles du discours mathématique a visiblement augmenté dans les dernières décennies. En conséquence, une autre question importante à laquelle je chercherai de répondre dans ce chapitre sera : « pourquoi les mathématiques appellent quelque chose de naturel ? » ; en espérant que la réponse à cette dernière puisse mettre en valeur l'évidence statistique. Une grande partie de ce chapitre sera dédiée à discuter la méthode d'enquête, puisque sur ce sujet spécifique et avec cette perspective, il n'est pas possible de s'appuyer sur les contributions des autres. Un aspect théorique important de cette étude, que la méthodologie proposée informera, sera une critique du naturalisme, en particulier celui proposé par Maddy. À la différence de cette forme de naturalisme, je suis convaincu qu'il existe des problèmes philosophiques intrinsèques dans le travail mathématique. En outre, l'analyse de la naturalité servira à révéler un choix réaliste dans son usage. En effet, la thèse principale de ce chapitre est que la référence à la nature, implicite dans la référence à des composantes naturelles des mathématiques, cache une tentative de rendre conceptuellement stable un phénomène qui, au contraire, a des caractéristiques dynamiques et normatives : la relation entre les aspects formels et ceux informels d'un travail mathématique. Une version plus longue de ce chapitre consiste en deux articles, écrits en collaboration avec Luca San Mauro.

Dans le chapitre 4, je reprendrai tous les critères discutés et proposés dans les chapitres précédents et je les testerai dans le cas particulier des Axiomes du Forcing. Dans le même temps, je discuterai des idées générales qui sont à la base de la notion d'ensemble arbitraire, liant cette dernière à celle d'un ensemble générique. En conclusion, enfin, j'argumenterai en faveur de la naturalité des Axiomes de Forcing.

À partir du chapitre 5, commence la partie mathématique de cette thèse. Dans ce chapitre, je présenterai préventivement la méthode des *side conditions*, dans sa forme la plus connue. Cette méthode consiste en l'utilisation de modèle de la théorie des ensembles comme partie intégrante de la définition des notions de forcing, son objectif est celui de préserver quelques cardinaux. Je récapitulerai les définitions principales, préparant le terrain pour les chapitres suivants.

Dans le chapitre 6, je reporterai les notes que j'ai écrites, en collaboration avec Moore, dans un bref cours sur les Axiomes de Forcing, tenu par lui au

“Young set theory workshop” à Raach (près de Vienne) en 2010. Je crois que ces notes sont une bonne et synthétique présentation de ces axiomes. En outre, ils utilisent la méthode des *side conditions* en connexion avec les Axiomes de Forcing.

Dans le Chapitre 7, j’inclurais le travail *A direct proof of the five element basis theorem* ([180]) de Veličković, parce que d’une coté il représente un exemple intéressant de l’utilisation des *side conditions* en combinaison avec les Axiomes de Forcing, et de l’autre coté j’ai pris part à la préparation de ce papier. Le *five element basis theorem* a été démontré pour la première fois par Moore, en 206, et ensuite simplifié par König, Lason, Moore et Veličković. Ce travail consistait en une simplification ultérieure de la démonstration de ce théorème.

Dans le chapitre 8, je reporterai (dans le quatre premières sections) les notes du séminaire *Proper forcing remastered* donné par Veličković a l’“Appalachian Set Theory Workshop”, qui a eu lieu le 5 Octobre 2011 à l’University of Illinois at Chicago. Comme dans le chapitre 7 j’ai pris part à la préparation des ces notes. Ici est présentée et appliquée une généralisation de la méthode des *side conditions*, en lige avec le travail d’Itay Neeman, qui le premier a proposé en 2011 de considérer comme *side conditions* des \in chaînes, fermés par intersection, de modèles des deux cardinalités différentes. Dans les première quatre sections on trouve la définition de l’ordre partiel formé des seules *side conditions* de deux types de modèles et l’application de cette méthodes généralisée au problème de comment forcer un ensemble fermé et illimité en ω_2 , avec des conditions finies, dans l’existence d’une chaîne de longueur ω_2 in $(\omega_1^{\omega_1}, <_{Fin})$, de l’existence d’une *Thin Tall Superatomic Boolean Algebra*. Ces notes, dans une version presque identique, ont été publiées comme [181]. Dans la section 5, d’après une conseil de Veličković, j’appliquerais les *side conditions* au problème de l’existence d’un arbre ω_2 -Suslin.

Dans le Chapitre 9, d’après une conseil de Veličković, j’appliquerais la méthode des *side conditions* dans la démonstration de consistance de l’axiome $PFA(T)$: l’Axiome de Forcing pour la classe des forcing qui sont propres, et dans le même temps, préservent un arbre de Souslin T . La section 2 utilise une idée de Veličković, et le theorem principal de la section 3 - ou je montre que d’après avoir force $PFA(T)$ l’arbre T est encore Souslin - suit de près celle de Neeman de la consistance de PFA avec des conditions finies.

Introduction

This thesis has many goals, many of which are specific and I will survey them in the summary, but there is also a global ambition that is pursued here: to keep together a philosophical and a mathematical reflection, without sacrificing the precision of none of the two. This concern is not only personal, but originates from the conviction of the complementarity of philosophical and mathematical tools in mathematics in general, and - to a great extent - in logical investigations in particular. When engaged in such a effort, one has to face both subjective and objective difficulties.

On the subjective side I have to acknowledge that it is sometimes hard to keep the same standard of clarity in both domains and to produce an interesting work for both philosophically and mathematically minded researchers. Moreover, on the objective side, even if I am proposing an interdisciplinary approach to logic, I am well aware of the fact that philosophy and mathematics share common interests but have different methodologies. The tip of the iceberg of this phenomenon is the style of argumentation that varies deeply between the two disciplines, but we can also list the relevance of the literature on a topic, or the different attitudes with respect to the autonomy of the discipline, or the relationship between a local and a global point of view.

Far from being a weakness, these different aspects of mathematics and philosophy point to their complementarity role in a field, like logic, that mixes together formal and informal sides of knowledge. More in general I believe that mathematics cannot be reduced to a pure game of symbols, excluding any informal component, nor that philosophy should avoid any use of formal tools in its argumentation. However, we should be careful not to make philosophy chasing mathematics on a too formal ground. This mistake, as it can be seen in some aspects of the analytic philosophy tradition, would restrict excessively the scope of an enquiry, missing some genuine philosophical problems that can arise in it.

The tension between the formal and the informal side of mathematics has recently favored the birth of the philosophy of mathematical practice. This branch of philosophy incorporates the attitude towards mathematics that finds in the

practice of doing mathematics the emergence of non mathematical problems and that tries to tackle them in a philosophical way. Thus it is not surprising that the work presented in this thesis shares some convergences with this philosophical posture. However, I believe that - at least in the logical matters - a step forward is needed: from a very close analysis of the work of a practitioner of a discipline can emerge some truly philosophical problems. This belief is motivated by the peculiar character of logic, that has both a descriptive and a deductive character. The latter allows to mathematize our rules of inference, while the former to connect formal tools and informal considerations. Then, the main possibility of a matching between perfectly rigorous pieces of mathematics and informal ideas rises some non trivial problems, that call for a general philosophical treatment in a purely mathematical context. As we will see, this conviction distances the philosophical side of my thesis from many forms of naturalism both at a theoretical, and at a methodological level.

Since I believe that mathematics is not exhausted by its formal side, and that in order to recognize some philosophical aspects of some problems one needs to understand also the technical details, in my opinion the best way to perform a philosophical analysis of mathematics is by taking part to its development. This is why this thesis combines mathematical results and philosophical reflections, hoping to favor both sides by the interaction with the other discipline.

However, both philosophy and mathematics, as all human activities, are organized in accordance with a scientific community: in general areas of interest, groups of research, and schools. This is also another concern for the development of a interdisciplinary research. In order to find a sufficiently stable ground for the development of such a work I looked back at a classical subject that has its roots in the period when philosophy and mathematics were not so disentangled, as in recent times. This subject is, in general, the notion of axiom, in particular, that of axiom in set theory, and even more specifically, the problems of the justification of new axioms in set theory.

Indeed this subject has both philosophical and mathematical aspects: on the one hand we can ask which are the reasons to accept a new axiom, since being an axiom implies the impossibility of its proof; on the other hand we can ask what we can prove from a new axiom. As we will see, the scopes of both questions are not so neatly separated. As a matter of fact, for what concerns the former question, I will argue in favor of a demonstrative character of its answer, while, for what concerns the latter, I will favor a more inductive approach: from consequences to axioms. Indeed I will present a method whose uniform treatment of different problems will call for a unifying axiom able to capture and justify the combinatorial aspects arising in the mathematical solution of the problems.

It is now time to make this discourse more precise and to delineate the structure of this thesis.

0.3 Summary

This work is divided in two parts. The first four chapters are mainly philosophical, while the rest of the thesis, including the conclusions, are mainly mathematical. This distinction is made in order to help the reader, but it should be kept in mind that the two parts pertain to the same reasoning, that would lack concreteness, if one skips the mathematical part; whereas it would be difficult to understand the aim and the sense of the mathematical results, if one forgets the philosophical context.

Chapter 1 deals with the origin of the concept of axiom in contemporary axiomatics. When I started to work on the justification of the axioms in set theory, I felt the need to clear the concept of axiom itself. Indeed this concept does not often receive a direct treatment, since it is normally assumed, although problematic, as sufficiently clear. Moreover, I believe that a serious philosophical analysis of a problem needs to investigate its historical roots. Then, this conviction led me to investigate the concept of axiom in Hilbert's thought. In doing so I had to face two difficulties: the absence of a vast literature on this specific topic - in comparison to the huge number of works on Hilbert's axiomatic - and the standard view that considers Hilbert's as the champion of formalism. The two problems are deeply related, because the major attention to the axiomatic method often obscures the origin and the nature of the notion of axiom, keeping the analysis at the level of formal systems and thus distorting the relevance of the informal components that can be found in Hilbert's work. The main outcome of this analysis will be that we can find two different notions of axiom in Hilbert's thought and both are linked to different conceptions of intuition. The difference between the latter notions is the role that evidence plays in respect to the relationship between axioms and the subject matter they formalize. I believe that Hilbert was well aware of the difficulties in matching formal and informal aspects of mathematics and one of the main theses of this chapter is that the attempt to find a solution to this problem was among the reasons for his proposal of the Axiom of Completeness and for the invention of the proof theory. Even if the historical analysis is interesting in its sake, and necessary to understand better Hilbert's reflection, I will argue that it is also instructive for the problem of the justification of new axioms in set theory. Indeed in the end of this chapter I will tie together the historical inquiry and the theoretical reflection, proposing some useful criteria for a specific class of axioms, that I will

call *sufficient*: axioms that act like sufficient conditions for the formalization of a specific domain of knowledge.

In Chapter 2 I will consider the effects of the foundational role of set theory on the problem of the justification of its axioms. In particular I will propose to look at set theory as a theory able to characterize the possibility of a proof of a mathematical fact and, so, able to characterize and to unify mathematics: a set theoretical foundation of mathematical practice. I will sustain this thesis with some examples taken from different fields: algebra, functional analysis and set theory. Moreover I will argue that this practical way of considering a set theoretical foundation is useful in order to understand the relevance of set theory for the context of a philosophy of mathematical practice. Indeed I will argue that some axioms in set theory can be seen as explanations of some mathematical phenomena. In this respect I will mainly discuss Kitcher's account on scientific explanation, in terms of unification, connecting his ideas with the explanatory role that some axioms of set theory manifest, with respect to a foundational setting. In particular, I will discuss two set theoretical results, by Woodin and Viale, where the matter of explanation and that of justification are tied together, in the context of the unifying role of set theory. This will allow me to propose a philosophical analysis of the notion of unification, that will lead to the formulation of a corresponding criterion for new axioms. In the end of this chapter I will broaden the analysis proposing the distinction between two different attitudes in the foundational enterprise: one theoretical and one practical. This general distinction will make clearer and distinguish - while legitimating both - two contemporary attitudes in the foundation of set theory: a multiverse view, opposed to the search of the true model for set theory.

In Chapter 3 I will try to clear the concept of naturalness. This notion is often used in modern mathematics and it normally intervenes in the debate on the criteria of justification for new axioms in set theory. I will not analyze the use of naturalness specifically in this debate, but I will try to understand the meaning of this term, in a more general mathematical context. The analysis will start from the statistical evidence that the appeal to natural component of the mathematical discourse has noteworthily increased in the last decades. Another important question that I will try to investigate in this chapter will be: "why mathematicians call something natural?", in order to account for the discovery of the growth of the use of the term "natural". A large part of this chapter will be devoted to discuss the methodology of this analysis, because on this particular subject, and in this perspective, it is not possible to rely on others' work. One important theoretical aspect of this investigation, that will inform the methodology proposed, will be a criticism of naturalism, in particular of Maddy's. Contrary to this form of naturalism, I believe that there are genuine

philosophical problems in some mathematical works. Moreover, the analysis of naturalness will help to unveil a realistic posture in its use. Indeed the main outcome of this chapter will be that the reference to nature, implicit in the reference to natural components of mathematics, hides an attempt to make conceptually stable a phenomenon that manifests, on the contrary, a dynamical and a normative component: the interplay between the formal and the informal side of mathematics. A longer version of this chapter consists in two joint papers, written with San Mauro.

In Chapter 4 I will resume all the criteria discussed and proposed in the previous chapters and I will test them in the particular case of the Forcing Axioms. In doing so I will discuss the general ideas that lay behind the modern conception of set theory. In particular I will discuss the notion of arbitrary set, linking the latter to the notion of generic set. In the end I will argue in favor of the naturalness of Forcing Axioms.

From Chapter 5 on starts the mathematical part of this work. In this chapter I will present briefly the method of forcing with side conditions, in its most known form. This method amounts to use models of set theory as part of the definition of a forcing notion, in order to guarantee the preservation of some cardinals. I will recall the main definitions, preparing the ground for the subsequent chapters.

In Chapter 6 I will report the notes that I wrote, in collaboration with Justin Moore, after a short-course on Forcing Axioms given by him at the “Young set theory workshop”, held in Raach (near Vienna), in 2010. I believe that these notes are a good and synthetic presentation of these axioms. Moreover, they make use of the method of side condition in connection with Forcing Axioms.

In Chapter 7 I will include the work *A direct proof of the five element basis theorem* ([180]) by Veličković, because on the one hand it is an interesting example of the use of the side conditions in combination with Forcing Axioms and, on the other hand, I helped in the preparation of this paper. The five element basis theorem was first proved by Justin Moore, in 2006, and then simplified by König, Larson, Moore and Veličković. This work amounts in a further simplification of the proof of this theorem.

In Chapter 8 I will report (in the first four sections) the notes of the seminar *Proper forcing remastered* given by Veličković at the “Appalachian Set Theory Workshop”, held on October 15th 2011 at the University of Illinois at Chicago. As for chapter 7 I helped in the preparation of these notes. Here is presented and applied a generalization of the method of the side conditions, in the line of the work of Neeman, who first proposed, in 2011, to consider \in -chains of models of two different cardinalities, closed under intersection, as side conditions. The first four sections consist in the presentation of the pure side condition poset and in the application of this method to the problem of adding a club set in ω_2 ,

with finite conditions, of the existence of a chain of length ω_2 , in $(\omega_1^{\omega_1}, <_{Fin})$ and of the existence of a Thin Tall Superatomic Boolean Algebra. These notes, in a slightly different form, have been published as [181]. In section 5, following a suggestion of Veličković, I will apply the pure side conditions to the problem of the existence of a ω_2 -Suslin tree.

In Chapter 9, following a suggestion of Veličković, I will apply the pure side conditions in the consistency proof of $\text{PFA}(T)$: the Forcing Axiom for the class of posets that are proper and preserve a tree T being Souslin. Section 2 elaborates on an idea of Veličković, while the main theorem of section 3 - where I show that forcing $\text{PFA}(T)$ does not change the fact that T is Souslin - follows closely Neeman's consistency proof of PFA.

Chapter 1

The concept of axiom in Hilbert's thought

In this chapter we will try to analyze the notion of axiom in contemporary axiomatics at its roots. It is important, since the beginning, to make clear the aim of our inquiry and the direction of our research. The main problem of this work is roughly speaking the following: which argument can we give to accept a new axiom in set theory? However this question is too vague and too general to be tackled directly. We then try to make it more precise, limiting its scope, and addressing a preliminary question: what is an axiom in modern set theory? Of course set theory has its peculiarities, but in order to understand correctly this problem it can be useful to widen our point of view, so that it becomes possible to embrace the right context where to place our question and to look for its answer. Then our more general question becomes: what is an axiom in a formal system in contemporary axiomatics?

When trying to understand how to tackle this concern, one can ask different philosophers of mathematics or ask mathematicians about their theories, ideas or feelings on this subject. However it is difficult to find a neutral or sufficiently wide reflection: a starting point from which it is possible to clear one's mind and to form a more refined and consistent opinion. Nevertheless, even if we suspend our judgment we need a starting point. Luckily history is gentle enough to point clearly in one direction, when asking "where our concept of axiom in modern axiomatic comes from?" Indeed, even if many different traditions, influences and ideas concur in forming one person's idea, the case of David Hilbert is easier to treat and his work on the foundation of mathematics is a good approximate answer to our question. This is true for many reasons: Hilbert is without any doubt one of the most influential mathematicians of the last century

- maybe the most. Even if he has relied on the work of other people, his thinking has always been autonomous, independent and capable of changing in a very original and structured way any fields he worked on. He has always combined deep and technical mathematical results - or better milestones - and theoretical and philosophical reflections, being aware of the importance, the depth and the possible developments of his work. Moreover he is universally recognized as the father of the modern axiomatics. These are all good reasons to turn to Hilbert's conception of axiom in order to understand its origins and maybe to have some hint on contemporaneity. So, rephrasing again our question we could ask: what is Hilbert's conception of axiom? Hoping, then, to gain some useful insight for our original question on the justification of the axioms in modern set theory. For these reasons our aim is not just historical, but philosophical in a broad sense: we want to understand the origin of a concept, in order to be able to deal with it in modern times, with a more conscious attitude. Indeed our historical analysis will always be directed to the philosophical problems that constitute the context of our inquiry. We rely on the historical development of this concept in order to discover which are the aspects we give for granted, while they are in need for a better analysis. Indeed we believe that sometimes a more distant point of observation may help in distinguishing things more clearer.

If we undertake the difficult task of clarifying the ideas of an author far from us in time, then some methodological precautions are necessary. First of all we must avoid the use of contemporary conceptual results in anachronistic contexts. As a matter of fact, understanding the genesis of concepts means going back to the time when those ideas were neither clear, nor completely understood. A careless historical analysis, although precise and competent, risks to obscure not only the intentions of those who took an active role in the development of the events, but also the scope and the extent of the ideas that are investigated. So, the analysis we would like to pursue here aims at contextualizing the choices made by Hilbert, with respects to the foundations of mathematics, without altering the originality of those ideas. We therefore propose to go to the root of the problems that Hilbert addressed, trying to understand the mathematical choices that confronted him and also to unveil the philosophical ideas that motivated them.

We assume as our methodological stance that concepts do not proceed in a straight line of reasoning, but they get more and more clear once they are used in solving problems. In this way, ideas and conceptions, at first vague, are modeled on solutions given to problems. These concepts then become indispensable tools for the discipline that implies them, so that they cannot be disentangled from its subject matter. In exact sciences the historical process is easily mystified in two forms: firstly, a retrospective look tends to discover a linear progression

of knowledge, and secondly the narrative of a discipline often proceeds in the opposite direction to the one that led to its formation.

In this chapter we will then analyze the key concept of Hilbert's axiomatic method, namely that of axiom. We will find two different concepts: the first one from the period of Hilbert's foundation of geometry and the second one at the time of the development of his proof theory. Both conceptions are linked to two different notion of intuition and show how Hilbert's ideas are far from a purely formalist conception of mathematics. Indeed the principal thesis of this chapter is that the main problem that Hilbert encountered in his foundational studies is the possible relationship between formalization and intuition. We will then show that Hilbert's solution is given, from a theoretical point of view, in the first period by means of the Completeness Axiom, while in the second period thanks to his proof theory. We will argue in details that the way in which Hilbert's Axiom of Completeness find an accordance between the formal and the informal sides of mathematical knowledge is by offering, at a logical level, necessary and sufficient conditions for a good formalization of Geometry. In the end we will argue that, due to the phenomenon of incompleteness, the solution to the problem of matching formalization and intuition, proposed in the second period, is not tenable anymore. On the contrary, the conceptual framework, given by the Axiom of Completeness, not only does apply to other branches of mathematics - indeed its role can be assimilated to that of the Axiom of Induction, for arithmetic, and of Church-Turing thesis, for computability theory - but it can also be recovered in the context of our original problem of justifying new axioms in set theory.

1.1 Two concepts

Hilbert's philosophical papers on the foundation of mathematics can be divided into two periods though not neatly separated. Their content differs in what is meant to be the nature of the axioms and in the methods used to achieve certainty and rigor in mathematics. We do not claim here to give an exhaustive account of Hilbert's conceptions of axiom, but at least to show that it changes through time and how it evolves. Strange as it may be, it is not possible to find in the literature a precise study of this particular subject. The role of axiomatization has always obscured the role played by the axioms in Hilbert's thought, and the latter is very instructive in order to understand the philosophical ideas that were behind Hilbert's proposals¹.

¹On the philosophical importance of Hilbert's foundational work see [42] and [140].

The first period centers around his work in geometry and his sketched attempt to prove the consistency of a weak form of arithmetic². In the works of this period, Hilbert's concept of axiom is linked to a “*deepening of the foundations* of the individual domains of knowledge”³. Indeed the axiomatization of a theory is gained by making explicit the logical structure of the corresponding domain of knowledge. In 1917 we can still find traces of this attitude in Hilbert's talk *Axiomatichen Denken*.

When we are engaged in investigating the foundations of a science, we must set up a system of axioms which contains an exact and complete description of the relations subsisting between the elementary ideas of that science⁴.

The second period starts in the early Twenties, after Hilbert has resumed the study of the foundations of mathematics. In this period's works we can see an effort to build the whole of mathematics on few axioms. These axioms were supposed to gain their legitimacy from the new *proof theory*, that, following Hilbert, “make[s] a protocol of the rules according to which our thinking actually proceeds”⁵. As a matter of fact, these rules are the *a priori* component of any form of mathematical knowledge. Indeed

also [...] mathematical knowledge in the end rests on a kind of intuitive insight [*anschaulicher Einsicht*] of this sort, and even that we need a certain intuitive *a priori* outlook for the construction of number theory⁶.

Hence, considering the whole of the mathematics as a formal system, the choice of its axioms rests on the individuation of the *a priori* principles that governs our conceptual knowledge and our mathematical experience.

1.1.1 The first period

In the first period Hilbert succeeds in completing the axiomatization of geometry and of the theory of real numbers. As a consequence the example of geometry is fundamental in Hilbert's reflections. In this period the enquiry about axioms is seen as a way to delve into the logical relationship among the theorems of

²See [70].

³[73], p. 1109 in [25].

⁴[57], p. 1104 in [25].

⁵[79], p. 475 in [55].

⁶[80], p. 1161 in [25].

a theory. Here ‘theory’ is not used in the formal sense, but it refers to any mathematical field of research that features only one subject of enquiry and homogeneous methods. In Hilbert’s words:

On the contrary I think that wherever, from the side of the theory of knowledge or in geometry, or from the theories of natural or physical science, mathematical ideas come up, the problem arises for mathematical science to investigate the principles underlying these ideas and so to establish them upon a simple and complete system of axioms, that the exactness of the new ideas and their applicability to deduction shall be in no respect inferior to those of the old arithmetical concepts⁷.

The analysis of the basic principles of a theory on the one hand leads to the choice of the axioms, and on the other hand it defines the concepts and relations in play. In this first period Hilbert has a precise idea of what axioms are: they are *implicit definitions*.

The axioms so set up are at the same time the definitions of those elementary ideas⁸.

Axioms define basic concepts and relations of a theory⁹. Moreover, he maintains that the process of formalization is complete¹⁰ only when the definitions of the concepts are properly given, through the axioms, and no other characteristic note can be added. However, it is important to notice that in order to be able to define something we need to have a pre-formal grasp of it. Indeed, even if one of the central novelty of Hilbert’s use of the axiomatic method is the separation between the truth of the axioms and their meaning¹¹, it must be explained how it is possible to match axioms and meaning. This explanation is needed if we

⁷[57], p. 1100 in [25].

⁸[57], p. 1104 in [25].

⁹We see here an implicit use of a sort of principle of comprehension that Hilbert states in this form: “the fundamental principle that a concept (a set) is defined and immediately usable if only it is determined for every object whether the object is subsumed under the concept or no.” in [70] p. 130.

¹⁰At this point, Hilbert has not handled the problem of the formalization of logic yet, nor Russell and Whitehead have written the *Principia mathematica*. For this reason it is clear that this ‘completeness’ Hilbert talks about has nothing to do with the completeness of logic or of the deductive methods. However we will come back later on the idea of completeness and show its relevance for Hilbert.

¹¹Indeed the axiomatic method is used to analyze the meaning of the axioms. Then the latter does not originate from the fact the axioms are true of the subject matter of the theory.

want to avoid the circularity that comes from the fact that axioms, as far as implicit definitions, express the concepts we are trying to analyze with them¹². Even if we accept the idea that truth is not a precondition for the meaning of an axiom, we need to accept, in some way or another, that this latter is a correct axiomatization with respect to the subject matter of a particular theory. In other words, one of the main problems of a formal treatment of a theory that Hilbert encounters is to explain why the axiomatic system so constructed should be a good formalization of the intended intuitive theory.

This is exactly the content of an objection raised by Frege.

Your system of definitions is like a system of equations with several unknowns, where there remains a doubt whether the equations are soluble and, especially, whether the unknown quantities are uniquely determined. If they were uniquely determined, it would be better to give the solutions, i.e. to explain each of the expressions ‘point’, ‘line’, ‘between’ individually through something that was already known. Given your definitions, I do not know how to decide the question whether my pocket watch is a point. The very first axiom deals with two points; thus if I wanted to know whether it held for my watch, I should first have to know of some other object that is was a point. But even if I knew this, e.g. of my penholder, I still could not decide whether my watch and my penholder determined a line, because I would not know what a line was¹³.

The objection is justified on the basis of Frege’s studies on the foundations of geometry. Indeed, he acknowledged that axioms were self-evident propositions and that geometrical objects were abstractions of empirical objects. Frege’s critic, however, is easily rebutted by Hilbert¹⁴. In fact he argues that that was exactly the strength of his method: to establish a formal system able to define an abstract concept, which would respond only to the requirements imposed by the axioms.

This is apparently where the cardinal point of the misunderstanding lies. I do not want to assume anything as known in advance; I

¹²A similar point has been raised by W.Tait in the notes from his talk “Dialectic and logic: the truth of axioms”.

¹³Letter from Frege to Hilbert January 6th, 1900; in [36], p. 45.

¹⁴Or at least this is what Hilbert would have answered, because he chose not to reply to Frege’s letter of January 6th, 1900. Anyway next quote is from Hilbert’s previous letter; and we can assume that if Hilbert did not write back to Frege is because he had already made his point.

regard my explanation in sec. 1 as the definition of the concepts point, line, plane - if one adds again all the axioms of groups I to V as characteristic marks. If one is looking for another definitions of a 'point', e.g. through paraphrase in terms of extensionless, etc., then I must indeed oppose such attempts in the most decisive way; one is looking for something one can never find because there is nothing there¹⁵.

The problem with Hilbert's reply is that it just points to a distinction of levels but does not give an explanation to the problem implicit in Frege's objection. We will call it Frege's problem and we formulate it as follows: why is the axiomatic system presented by Hilbert in the *Grundlagen der Geometrie* to be considered an axiomatization of Geometry? In other words, if the axioms formalize the fundamental ideas of a theory and they are what allows the most important geometrical facts to be proved, what are the criteria that make possible to identify the class of theorems we are interested in as theorems of Geometry? And finally: in Hilbert's view, what is the definition of Geometry once the axiomatic method has cut off the link between formalization and spatial intuition?

These are ones of the major concerns in Hilbert's foundational reflections and we will see how this notion of correctness is phrased in this period and how intuition is considered as a way out from this circularity.

In 1899-1900 Hilbert explicitly mentions the role of both intuition and axioms.

These axioms may be arranged in five groups. Each of these groups expresses, by itself, certain related fundamental facts of our intuition¹⁶.

Also:

The use of geometrical signs as a means of strict proof presupposes the exact knowledge and complete mastery of the axioms which underlie those figures; and in order that these geometrical figures may be incorporated in the general treasure of mathematical signs, there is necessary a rigorous axiomatic investigation of their conceptual content. [...] so the use of geometrical signs is determined by the axioms of geometrical concepts and their combinations¹⁷.

¹⁵Letter from Hilbert to Frege December 29th, 1899; in [36], p. 39.

¹⁶[81], p. 1.

¹⁷[57], pp. 1100-1101 in [25]. Here the notion of sign is wider than in the second period and geometrical figures are considered as signs.

In the passages quoted above is clearly outlined the way the axiomatic method proceeds: it analyzes theorems and concepts of a mathematical theory and isolates the basic principles that correspond to intuitive ideas. Then these principles are formalized in the form of axioms.

The remarks presented so far undermine the claim that Hilbert is a champion of formalism¹⁸, at least in two respects. First, the axiomatic method does not require a formalized logic, not even symbols. Moreover formalization is not meaningless, but it represents the basic concepts of a mathematical theory by means of symbols that have an intuitive content. Before formalization Hilbert sees an *historical* development¹⁹, after which an axiomatization is possible, and moreover the knowledge conveyed by a developed theory is the source of the *meaning* of signs.

In what comes next we will discuss what Hilbert means by “intuitive content of the symbols”. We will also examine how the intuitive content of the symbols links the demonstrative use of a formal system to the meaning of the concepts of a theory.

The analysis of these issues will provide a better understanding of the reason why a consistency proof is, according to Hilbert, the only possible way to establish the truth of the axioms. Moreover Hilbert argues that axioms, insofar implicit definitions, they guarantee the existence of the defined entities, modulo consistency.

Granting that consistency and existence are so strongly linked, we could ask which role the axioms play in this relationship. Hilbert acknowledges that if the axiomatization of a theory is complete, then the entities a theory is about are defined and uniquely determined by the axioms. But, for Hilbert, any set of axiom defines something, and so the process of extending a set of axioms gives rise to a sequence of different definitions: “every axiom contributes something to the definition [of the concept], and hence every new axiom changes the concept”²⁰. Nevertheless the intended model for a theory is what makes coherent all the different definitions: “[...] the definition of the concept point is not complete till the structure of the system of axioms is complete”²¹. On the other hand the

¹⁸Against this old conception see also [85], [101] and [156].

¹⁹Remember Hilbert’s description of the development of science: “The edifice of science is not raised like a dwelling, in which the foundations are first firmly laid and only then one proceeds to construct and to enlarge the rooms. Science prefers to secure as soon as possible comfortable spaces to wander around and only subsequently, when signs appear here and there that the loose foundations are not able to sustain the expansion of the rooms, it sets about supporting and fortifying them. This is not a weakness, but rather the right and healthy path of development.” In [70], p. 102.

²⁰Letter from Hilbert to Frege December 29th, 1899; in [36], p. 40.

²¹Letter from Hilbert to Frege December 29th, 1899; in [36], p. 42.

problem of the possible existence of different compatible formalizations of the same theory does not concern Hilbert. As a matter of fact, Hilbert acknowledges that the question whether different axiomatic systems can both be legitimate is theoretically interesting. However he does not explain how it is possible that two different theories can talk about the same things, since different axioms define different concepts. It is possible, but we are now entering the realm of speculation, that Hilbert would have replied that there are different ways to define the same concept, and so its syntactical presentations can differ.

In these observations we can see a weak form of realism in Hilbert's ideas, since the existence of abstract entities does not depend on the symbols we use to express them. Nevertheless, in Hilbert's works we cannot find an answer to this problem, but it is reasonable to think that Hilbert believed in the existence of a complete formalization of any mathematical concept²². This idea is coherent with his confidence that every mathematical problem has a definite solution.

On this respect there is another question that is left open by Hilbert: what kind of entities are defined by the implicit definitions of the axioms? This problem was explicitly raised by Frege in the letter to Hilbert:

The characteristic marks you give in your axioms are apparently all higher than first-level; i.e., they do not answer to the question "What properties must an object have in order to be a point (a line, a plane, etc.)?", but they contain, e.g., second-order relations, e.g., between the concept *point* and the concept *line*. It seems to me that you really want to define second-level concepts but do not clearly distinguish them from first-level ones²³.

Indeed Hilbert is not precise in saying what the axioms define, sometimes they seem to define mathematical objects:

I regard my explanation in sec. 1 as the definition of the concepts point, line, plane - if one adds again all the axioms of groups I to V as characteristic marks.²⁴

²²In the end of this chapter we will see that even if the existence of a concept is stable, the development of a formal system and the choice of the implicit definitions given by its axioms are a dynamical process and potentially an open one. In this respect it would be interesting to analyze the relationship between Hilbert and Husserl. There is indeed a phenomenological flavor in this idea of formalization as an open process, but there is not a corresponding theory of concepts.

²³Letter from Frege to Hilbert January 6th, 1900; in [36], p.46.

²⁴Letter from Hilbert to Frege December 29th, 1899; in [36], p. 39.

Sometimes Hilbert says that axioms define the relations between mathematical objects. As a matter of fact, in the same letter to Frege, he says that the axioms of the *Grundlagen der Geometrie* can also define the concept of “between”. In defense to Hilbert it is worth saying that at that time there was no clear distinction between first order and, if not second order, stronger logics. In the following we will see that Hilbert will abandon this careless ontological commitments and will consider the axioms just as defining the relations between mathematical concepts, or as “images of thoughts”. Even if Hilbert chose not to reply to the objections of Frege, we could argue that he indeed carefully considered them.

In this first period Hilbert thinks that it is also necessary to prove the independence of the axioms - or of the group of axioms. The proof, once obtained, would be a point in favor of the adequacy of the choice of the axioms, rather than in favor of their truth. Indeed if the search for axioms is an analysis of the basic principles of a theory, an independence proof would mean that the analysis has been accurate and has been able to single out the right basic principles of the theory. In other words, independence is deeply linked with the notion of completeness.

This need of an independence proof, unlike that of a consistency proof, will be dropped in the second period. This is another hint of a change in the concept of axiom.

1.1.2 The second period

Hilbert’s second period begins, publicly, in the early Twenties, but its roots can be traced back to the last years of the previous decade²⁵. Different reasons concur in indicating a change of opinions and attitudes towards the foundations of mathematics. Hilbert abandons the confidence in the systematization of logic proposed by Russell and Whitehead and expounded in the *Principia mathematica*. Hilbert then starts the most original contribution to the study of logic, in order to improve its formalization. Moreover, in that period the debate around intuitionism became more and more controversial. This leads Hilbert to often intervene in the debate against the irrational pushes that animated the mathematical community in those years.

In the lectures *Neubergründung der Mathematik. Erste Mitteilung* (1922) and *Die logischen Grundlagen der Mathematik* (1923) Hilbert outlines a new analysis of the concept of axiom. These two works can be regarded as belonging to a transition period in between the first and the second, not only chronologically

²⁵For a detailed description of this transition period see [156].

but also from a conceptual point of view. The work that clearly marks that a change has happened is *Über das Unendliche* (1925).

Consider the following definition of axiom that reflects the effect of this shift and shows how the transition between the first and the second period is continuous.

In order to investigate a subfield of a science, one bases it on the smallest possible number of principles, which are to be as simple, intuitive, and comprehensible as possible, and which one collects together and sets up as axioms. Nothing prevents us from taking as axioms propositions which are provable, or which we believe are provable²⁶.

In the above quotation the axioms, although are still basic principles, can just be provable propositions, or maybe-provable propositions.

In the same paper, later on, the concept of axiom is defined in the following ways:

The continuum of real numbers is a system of things which are linked to one another by determinate relations, the so-called axioms²⁷.

Next definition can also be found in *Die Grundlagen der Mathematik* (1928) and in *Über das Unendliche*,

Certain formulas which serve as building blocks for the formal structure of mathematics are called *axioms*²⁸.

First of all we need to notice that the axioms do not define any kind of mathematical objects, but just their relations²⁹. On the other hand the difference between axioms and other formulas begins to be less marked. Hilbert says:

The axioms and provable theorems [...] are the images of the thoughts that make up the usual procedure of traditional mathematics; but they are not themselves the truth in any absolute sense. Rather, the absolute truths are the insights that my proof theory furnishes into the provability and the consistency of these formal systems³⁰.

²⁶[76], p. 1119 in [25].

²⁷[76], p. 1118 in [25].

²⁸[76], p. 1125 in [25] and [78].

²⁹As we noted in the previous section this change in Hilbert's conception of axioms may be traced back to his correspondence with Frege.

³⁰[77], p. 1138 in [25].

In the above quotation one can see not only that axioms and provable propositions have the same relevance, as far as they are “images of the thoughts”, but also that axioms are deprived of their truth³¹ and keep just an operational character for the “usual procedure of traditional mathematics”. It is also possible to read the beginning of a separation between the concepts of consistency and truth, as it will then become apparent later, considering the further development of logic in the Thirties.

During the Twenties Hilbert’s proof theory was born. Consequently the axiomatic method becomes a tool that influences the principles of the whole mathematics in its formalized presentation. Recall that, in the earlier period, the axiomatic method consisted in the analysis of a field of mathematical knowledge, in order to isolate its principles and to make it a formal theory.

In this new perspective Hilbert defines a new kind of axiom.

This program already affects the choice of axioms for our proof theory³².

We recall them, without discussion, just to give a general insight of Hilbert’s ideas.

I. Axioms of implication

$$A \rightarrow (B \rightarrow A)$$

(Adjunction of a presupposition)

$$(A \rightarrow (A \rightarrow B)) \rightarrow (A \rightarrow B)$$

(Deletion of a presupposition)

$$(A \rightarrow (B \rightarrow C)) \rightarrow (B \rightarrow (A \rightarrow C))$$

(Exchange of a presupposition)

$$(B \rightarrow C) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$$

(Elimination of a statement)

II. Axioms of negation

$$A \rightarrow (\bar{A} \rightarrow B)$$

(Law of contradiction)

$$(A \rightarrow B) \rightarrow ((\bar{A} \rightarrow B) \rightarrow B)$$

(Principle of *tertium non datur*)

³¹Not that Hilbert considered axioms as absolutely true but previously he assigned them the function of defining concepts.

³²[77], p. 1138 in [25].

III. Axioms of equality

$$a = a$$

$$a = b \rightarrow (A(a) = A(b))$$

IV. Axioms of number

$$a + 1 \neq 0$$

$$\delta(a + 1) = a$$

V. Transfinite axiom

$$A(\tau A) \rightarrow A(a)$$

VI. Axiom defying the all- and existence-signs

$$A(\tau A) \rightarrow (a)A(a)$$

$$(a)A(a) \rightarrow A(\tau A)$$

$$A(\tau \bar{A}) \rightarrow (Ea)A(a)$$

$$(Ea)A(a) \rightarrow A(\tau \bar{A})$$

These new axioms are not of the same nature as the ones mentioned in the previous quotations. They are the axioms on which the mathematical building rests. These axioms are logical and arithmetical in characters and are true axioms, in an absolute sense, since they draw their certitude and evidence from how Hilbert is now setting the problem of the foundation of mathematics: a proof theory that tries to justify the ideal elements with finitary tools.

This circumstance corresponds to a conviction I have long maintained, namely, that a simultaneous construction of arithmetic and formal logic is necessary because of the close connection and inseparability of arithmetical and logical truth³³.

³³[76], pp. 1131-1132 in [25]. In [85] can be found a good account of the importance of the combinatorial aspects related to the role of the arithmetic in particular, and calculation, in general. See also [134] in this respect. We do not try here to trace the dividing line between logic and arithmetic. A discussion of Hilbert's logicism can be found in [29], and will be discussed later.

However, since - at least finitary - mathematical statements have a content (*Inhalt*), intuition cannot be ignored in the foundation of mathematics. In next section we will tackle the analysis of the concept of intuition, now it is enough to say that intuition is the source of certainty and evidence for mathematics and it is capable of making mathematical truths absolute. Intuition is the origin of certainty in the finitary setting. So Hilbert, considering the whole mathematics as a complex of formal propositions, founds the certainty of mathematics in the intuitive relationship between the thinking subject and the symbols, “immediately clear and understandable”.

Following the terminology of Feferman³⁴, we could call Hilbert’s logical-arithmetical axioms *foundational* - while the axioms of a non foundational theory can be called *structural*.

Finally in 1925,

Certain of the formulas correspond to mathematical axioms. The rules whereby the formula are derived from one another correspond to material deduction. Material deduction is thus replaced by a formal procedure governed by rules. The rigorous transition from a naïve to a formal treatment is effected, therefore, both for the axioms (which, though originally viewed naïvely as basic truth, have been long treated in modern axiomatics as mere relations between concepts) and for the logical calculus (which originally was supposed to be merely a different language)³⁵.

Here we see Hilbert’s full awareness of the changed concept of structural axioms. Indeed for Hilbert axioms were “originally viewed naïvely as basic truth”, as Euclid did, then at the beginning of this new axiomatic era as “mere relations between concepts”, like he himself did in the *Grundlagen der Geometrie*³⁶. Finally Hilbert thinks that his proof theory brought to the end this process of formalization of mathematics, so that also the structural axioms need to be viewed as meaningless formulas and they do not have more meaning than other mathematical propositions. These ideas mirror an old idea of Hilbert that seems to be constant over time: “Usually, in the story of a mathematical theory we can easily and clearly distinguish three stages of development: naïve, formal and critical³⁷”.

³⁴[28]

³⁵[78], p. 381 in [55].

³⁶As he now sees his foundational work, retrospectively.

³⁷In [59], p. 383 in [83]. In German: *In der Geschichte einer mathematischen Theorie lassen sich meist 3 Entwicklungsperioden leicht und deutlich unterscheiden: Die naive, die formale und die kritische*. My translation.

It is important to notice that in this second period the axiomatic method is still considered a logical tool, subordinate to a safe foundation of mathematics to be pursued by means of proof theory. The working of this method has to be logical. Hilbert says: “The axiomatic method belongs to logic³⁸”. We could read here a sort of logicism in Hilbert’s ideas, but this would be a mistake³⁹. Indeed the axiomatic method is not the tool capable of giving foundation to mathematics, but it is only used to formalize mathematics, like a preparatory study. This preparatory work puts the mathematician in the position of applying proof theory. We thus have an instrumental conception of logic, not a foundational one.

Let us now analyze the concept of intuition. We anticipate that we will find two different concepts of intuition and that this difference is responsible for two different concepts of axiom.

1.2 Hilbert and intuition

The intuitive character of the axioms is what marks the difference between them and other true propositions⁴⁰.

First of all we need to stress the difference between “intuitive” and “evident”, since the confusion between these two concepts has always been source of ambiguity. By “evident sentence” we mean a sentence that does not need to be analyzed to exhibit its truth. By “intuitive sentence” we mean a sentence whose truth, in a given context and with a given background knowledge, is immediately perceived, so that it is possible to skip some step of reasoning that, in other cases, would be necessary. This distinction pertains to the difference between the level of validity and that of justification. An evident sentence need not to be justified and its validity is immediately given, while an intuitive sentence is believed thanks to a good level of knowledge of the subject who considers it and for this reason it does not need a proof for its justification. In the latter case validity does not follow for free but it must be ascertained in a different way, for example with a proof, when it is possible. Unlike evidence, which is innate within our mind, intuition can be educated thanks to training and mathematical practice. There is an important *caveat*, though. The intuition we are talking

³⁸[80], p. 1158 in [25].

³⁹In what follows we will also discuss the logicism of Hilbert in the first period of his inquires in the foundations of mathematics.

⁴⁰There are not many places where it is possible to find a detailed analysis of the role of intuition in Hilbert’s work. Few exceptions are [90], [106] and [137], for what concerns the second period of his foundational studies. For a more comprehensive study two interesting references are [113] and [23].

about, that we could call a contextual intuition, is not an intuition that depends on a specific faculty of the mind, different from intellect. In other words it is not a Kantian-style intuition, i.e. a faculty whose structure depends on pure forms, that are given once and for all, like space and time, and that governs sensible knowledge. On the contrary the intuition we are considering here can be refined by the same knowledge that it helps to create. Besides, for Kant, intuition is not intellectual, since it acts in perception and makes perception possible.

We also distinguish two modes of intuition; following [136] we call them intuition *of* and intuition *that*, to stress the difference between the conception of intuition as a kind of perception - *à la* Gödel⁴¹ - and the idea that intuition can be a propositional attitude. In neither cases intuition is a form of knowledge. What any kind of intuition lacks to become knowledge is the evident characters that make existent - in same sense - the objects of intuition and true the propositions intuited. This evidence if not immediate can be given by proofs or sufficiently reliable arguments. What is important to stress here is that intuition can become knowledge thanks to a rational process.

We will see how these different concepts can determine the nature of axioms.

1.2.1 First period

In the earlier Hilbert's foundational works the context of a mathematical theory plays a fundamental role in the choice of the axioms. Indeed the "axiomatic investigation of their [i.e. of the signs] conceptual content" is relative to an informal theory and allows the "use of geometrical signs as a means of strict proof". Moreover, since "the use of geometrical signs is determined by the axioms", intuition and mathematical practice are connected.

A precise account of mathematical signs is then outlined. Mathematical signs, including geometrical figures, can be used in a proof as far as their conceptual content is adequate to the context; that is when signs formalize principles that are coherent with the basic concepts of the underling theory. Then they can be used as demonstrative tools, in the ways allowed by the axioms. So, the "conceptual content" is just the meaning of signs in the context of use. This meaning depends on the axioms that concur, as implicit definition, in determining the basic principles of a theory. As we see the formal and the pre-formal⁴² sides of

⁴¹In [136] Parsons shows that this kind of intuition, although is explicitly defended by Gödel in [47], is not the only one that can be found in Gödel's works. Starting from this right remark, it would be interesting to analyze the analogies and differences between the - at least - two different conception on intuition in Gödel's thought in comparison with Hilbert's.

⁴²By pre-formal here we mean "before the axiomatical presentation of an intuitive theory". Indeed such an intuitive theory gathers both formal and informal knowledge.

mathematics mutually influence each other. Axioms determine the meaning of signs and their demonstrative use, but where the axioms come from and how can they match with ideas and use? A link and a correspondence then must be found between these two sides of mathematical knowledge.

Hilbert's solution appeals to intuition as he says in the beginning of the *Grundlagen der Geometrie*: axioms express "certain [...] fundamental facts of our intuition". This fact could sound strange, if one considers Hilbert's positions as expressed in his correspondence with Frege, where he marks clearly his distance from any conception of geometry that sees in the spatial intuition the source of legitimacy of its axioms. Indeed the *incipit* of the *Grundlagen der Geometrie* and its reference to intuition is partly the result of an immature reflection on the sources of knowledge in geometry⁴³, but it also springs from a notion of intuition that is not only the empirical intuition of space, as in the Euclidean formulation. Although recognizing the intuition of space as the starting point of any geometrical reflection, Hilbert maintains that it is not the ultimate source of meaning and truth of geometrical propositions. A different notion of intuition leads Hilbert to argue that the analysis of the foundations of geometry consists of "a rigorous axiomatic investigation of their [of the geometrical signs] conceptual content"⁴⁴. As a matter of fact Hilbert is explicit in recognizing that the axioms of geometry have different degrees of intuitiveness.

A general remark on the character of our axioms I-V might be pertinent here. The axioms I-III [incidence, order, congruence] state very simple, one could even say, original facts; their validity in nature can easily be demonstrated through experiment. Against this, however, the validity of IV and V [parallels and continuity in the form of the Archimedean Axiom] is not so immediately clear. The experimental confirmation of these demands a greater number of experiments.⁴⁵

In order to clear this intricate connection, maybe it could be useful to see in details how the axiomatic method works, as described by Hilbert. The process of axiomatization starts from an intuition concerning a domain of facts

⁴³For a detailed study of the origins and the early influences on Hilbert's conception of geometry see [175], [174] and [176]. In [23] Corry argues that the progression of Hilbert's works marks the shift, for what concerns geometrical knowledge, from intuition to experience, thanks to his involvement in the research on general relativity theory. We maintain that there is indeed an evolution but we think that the intuition in the period of the foundation of geometry is not a kantian-style intuition as Corry's article seems to presuppose. We will come back later on this point when discussing the primacy of Euclidean geometry in Hilbert's *Grundlagen*.

⁴⁴[57], p. 1101 in [25].

⁴⁵[62], p. 380 in [52].

(*Tatsachen*), then, while formalizing it, it tries to clear the logical relationships within the concepts of the theory. This process, as Hilbert describes it leads from the subject matter of a theory to a conceptual level⁴⁶.

The method of the axiomatic construction of a theory presents itself as the procedure of the mapping [*Abbildung*] of a domain of knowledge onto a framework of concepts, which is carried out in such a way that to the object of the domain of knowledge there now correspond the concepts, and to statements about the objects there correspond the logical relations between the concepts⁴⁷.

The work of the axiomatic method is not exhausted by the formalization of an informal theory, because one of its tasks is to analyze the meaning of signs, by means of formal methods. Indeed, by deepening the foundations of a domain of knowledge one elucidates, at once, the logical structure of the theory and the intuitions about the subject matter of the theory. Therefore axioms have a double role with respect to signs. On the one hand axioms, through the axiomatic enquiry, are used to give meaning to signs, on the other hand they grant the demonstrative power of signs, linking intuition to mathematical practice in the act of justification of their use. Indeed intuition both precedes axiomatization and guides the work of a mathematician.

[O]ne should always be guided by intuition when laying things down axiomatically, and one always has intuition before oneself as a goal [*Zielpunkt*]. Therefore, it is no defect if the names always recall, and even make easier to recall, the content of the axioms, the more so as one can avoid very easily any involvement of intuition in the logical investigations, at least with some care and practice⁴⁸.

⁴⁶Recall that at the beginning of [81] Hilbert quotes Kants' *Critique of pure reason* and writes "All human knowledge begins with intuitions, thence passes to concepts and ends with ideas". This quotation, though not kantian in spirit, explains how Hilbert wanted to use the axiomatic method in his researches. Indeed for Hilbert the mathematical objects - or relations; recall Frege's criticism, built on the distinction between first and second order - defined by the axioms of the *Grundlagen der Geometrie* are not strictly speaking geometrical objects but conceptual entities that can be interpreted as geometrical objects. The intended interpretation is of course that of geometry, but this does not narrow the range of possible interpretations that can be give to formulas that constitute the formal system. We then can see three distinct levels of things: 1) empirical entities 2) formal objects 3) elementary ideas of Geometry. This distinction also mirrors the evolutive steps of a theory: naïve, formal and critical.

⁴⁷[75], p. 3. Translation in [51].

⁴⁸[71], pp. 87-88. Translation in [51].

The kind of intuition that allows to give meaning to mathematical propositions, is not evidence, but it is a contextual intuition that develops in parallel with the demonstrative techniques. It is the same intuition according to which mathematicians isolate and choose the axioms of a theory⁴⁹. It is the intuition that one develops when working within a theory. The axiomatic method then consists in formalizing, by means of signs - figures, symbols or diagrams - a *modus operandi* acquired by habit. Indeed, in 1901, Hilbert, in discussing the primacy of his work with respect to Kline's program, maintains that the concepts of Euclidean geometry are more familiar, not because of our outer intuition of the world, but thanks to our elementary study of the subject at school.

On the basis of Riemann and Helmholtz Lie set up a system of axioms which differs fundamentally from those systems that are developed according to the Euclidean model. Lie's axioms contain function-theoretic parts since he requires motion to be expressed by differentiable functions. [...] The question arises whether the function-theoretic components are only necessary because of the desire to apply this (group-theoretic) method, or whether they are foreign to the subject matter itself and are thus superfluous. It turns out that in fact they are. Thereby we once again draw closer to the old Euclid, insofar as we don't need to impose the additional infinitesimal properties on the concept of motion which Lie still thought necessary. Instead, the elementary postulates which are already contained in the Euclidean concept of congruence suffice, a concept with which we are all familiar, due to the theorems about the congruence of triangles known from school⁵⁰.

Following the terminology fixed before it is an intuition *that*: a propositional attitude towards mathematics, that can be formalized and gains certainty, once a consistency proof is given for the formal system that embodies its syntactic counterpart: the signs. It is not an innate intuition, but it is sufficiently reliable to be used as an heuristic criterion and that can be formalized, once it is shown to be correct. Obviously this criterion is not always safe:

⁴⁹When working on the foundation of geometry Hilbert explains his goal in the following way: "we can outline our task as constituting a *logical analysis of our intuition [Anschauungsvermögens]*" ([63], p. 2), i.e. an analysis of the most fundamental principles of geometry, conducted with formal means. Among these principles there are of course also our spatial intuitions, but "the question of whether spatial intuition has an *a priori* or empirical character is not hereby elucidated" ([63], p. 2). As a matter of fact, in these years, there is no philosophical analysis of the faculty of intuition. Nevertheless the quotation above (from [71]) shows that the intuition involved is not just a faculty of sensation.

⁵⁰From a lecture before the Royal Academy of Science in Göttingen, 1901. [113], p. 61.

[...] we do not habitually follow the chain of reasoning back to the axioms in arithmetical, any more than in geometrical discussions. On the contrary we apply, especially in first attacking a problem, a rapid, unconscious, not absolutely sure combination, trusting to a certain arithmetical feeling for the behavior of the arithmetical symbols, which we could dispense with as little in arithmetic as with the geometrical imagination in geometry⁵¹.

All these remarks show that at the beginning of Hilbert's reflections there is no coincidence between the notion of intuition and the notion of evidence. Indeed, Hilbert's explicit purpose, while writing the *Grundlagen der Geometrie*, was to give a safe basis to geometry different from space intuition, unlike the euclidean axiomatic setting. Hilbert wanted to justify also non-euclidean geometries, so, after refusing evidence as a criterion for truth, he looked for a sufficiently general and comprehensive principle to give foundation to geometry, i.e. the axiomatic method⁵². Nevertheless signs need meaning, in order to avoid a meaningless discourse. This is the "conceptual content" mentioned by Hilbert, where the intuition that gives meaning to signs is not the pure intuition of space - in case of geometry - but it is the intuition of the basic concepts of the theory that are formalized by means of axioms. This intuition is contextual to the formal system, it is the intuition that allows us to determine, thanks to the implicit definitions of the axioms, what are points, lines and space, as far as they are geometrical entities, i.e. part of a geometrical formal theory.

We can find an antecedent of this kind of intuition in Klein's words:

Mechanical experiences, such as we have in the manipulation of solid bodies, contribute to forming our ordinary metric intuition, while optical experiences with light-rays and shadows are responsible for the developement of a 'projective' intuition⁵³.

However a different conception of the axiomatic method and of a formalistic treatment of mathematics⁵⁴ will lead Klein to a different approach to geometry. Indeed Klein's geometrical enquires and the Erlangen's Programme will always presuppose an uncritical treatment of the intuitive data on the nature of space, contrary to the basic principle that aims Hilbert's axiomatic method. Indeed, while Klein will try to analyze and classify the different kind of spaces, Hilbert

⁵¹[57], p. 1101 in [25].

⁵²Even if, at the time, Hilbert lacked the logical tools.

⁵³In [95], p. 593.

⁵⁴On this subject see [177].

will deal with intuitions prior to the concept of space. We will come back later to this point, while recalling the different stages that Hilbert saw in the development of a science.

Once we cleared the notion of intuition that Hilbert had in this first period is now time to come back to the problem of axioms and see how they can match with this contextual intuition. In the preface to the *Grundlagen der Geometrie*, Hilbert is explicit in pointing out the requirements that a system of axioms must meet to be considered a good presentation of a theory.

The following investigation is a new attempt to choose for geometry a *simple* and *complete* [*vollständiges*] set of *independent* axioms and to deduce from these the most important geometrical theorems in such a manner as to bring out as clearly as possible the significance [*Bedeutung*] of the different groups of axioms and the scope of the conclusions to be derived from the individual axioms. ⁵⁵.

Here we see clearly that the meaning of the axioms is related to the technical tools they provide, as they are used in proving geometric theorems. This meaning is therefore intrinsic to the context of the theory.

Hilbert requires that a formal system should be simple, complete and independent. As is common in mathematics the more the ideas are simple, the more they are deep and fundamental⁵⁶. Indeed the problem of simplicity is linked with the idea of deepening the foundation of a theory⁵⁷. Moreover, the demand for independence is for Hilbert, as we saw, a necessary condition for a good application of the axiomatic method. Indeed, for Hilbert the independence of a system of axioms is an index of the depth of the principles expressed by the axioms⁵⁸. Of course a system should also be consistent, but this is not a major

⁵⁵[81], p. 1. Also in [64].

⁵⁶We will not discuss here the problem of simplicity, although it is partially linked to that of purity of the methods we will address later. In the *Mathematische Notizbücher* ([56]) Hilbert writes: “The 24th problem in my Paris lecture was to be: Criteria of simplicity, or proof of the greatest simplicity of certain proofs. Develop a theory of the method of proof in mathematics in general. Under a given set of conditions there can be but one simplest proof. Quite generally, if there are two proofs for a theorem, you must keep going until you have derived each from the other, or until it becomes quite evident what variant conditions (and aids) have been used in the two proofs. Given two routes, it is not right to take either of these two or to look for a third; it is necessary to investigate the area lying between the two routes.

⁵⁷In [167] is argued how this notion of simplicity is also connected to the development of Hilbert’s proof theory. But we do not know if this notion of simplicity can be seen as a link that motivates the passage from the first to the second period.

⁵⁸Notice however that the system of axioms proposed by Hilbert was not entirely independent. A truly independent system of axioms for geometry, but not categorical, will be proposed in 1904 by Oscar Veblen in [178].

concern of this period, even if it should be noted that a consistency proof carries also the burden of making possible mathematical knowledge, since it has to be knowledge of something true and existent. Finally it is now time to analyze the notion of completeness, and its formal counterpart: the Axiom of Completeness that is said, by Hilbert, to be the cornerstone of his foundation of geometry.

1.2.2 Completeness

In 1899, after a series of lectures on geometry held at the University of Göttingen⁵⁹, Hilbert published the “Foundations of Geometry” (*Grundlagen der Geometrie*)⁶⁰.

The system of axioms that Hilbert sets up in the *Grundlagen der Geometrie* is divided into five groups. In order: connection, order, parallels, congruence and continuity. We have two axioms of continuity: Archimedes’s axiom and the Axiom of Completeness.

We now want to analyze the latter and try to understand what led Hilbert to formulate this axiom and why it occupies such an important role in the whole construction of the foundation of geometry.

To the preceding five groups of axioms, we may add the following one, which, although not of a purely geometrical nature, merits particular attention from a theoretical point of view⁶¹.

Moreover Hilbert argues that the Axiom of Completeness “*forms the cornerstone of the entire system of axioms*”⁶².

In the first German edition of 1899 there is no trace of the Axiom of Completeness. It appears from the second, in 1903, to the sixth, in 1923, in the following form:

⁵⁹See [175] and [52], for a precise exposition of the origins of the *Grundlagen der Geometrie* and of the development of Hilbert’s reflections on geometry in this early period.

⁶⁰When referring and quoting it we will use [64] to indicate the first German edition and [65] for the first French edition. Otherwise [81] refers to the first English edition, translated from the second German edition ([68]), while [84] indicates the second English edition, translated from the tenth German edition ([82]). However, when quoting from [84], we will point to the German edition where the quote first appeared. Moreover, when quoting [81], we will indicate if the quote can be found also in [64].

⁶¹[81], p. 15.

⁶²[84], p.28; original emphasis. From the seventh German edition onward. This addition in the 1930 edition of the *Grundlagen der Geometrie* points in the direction of the importance of the Axiom of Completeness also in the second period of Hilbert’s foundational inquiries. As a matter of fact, we will argue that, even if Hilbert’s methods and conceptual background change through time, the role of the Axiom of Completeness, for Geometry, is analogous to that of proof theory, for mathematics.

V.2 (Axiom der Vollständigkeit) Die Elemente (Punkte, Geraden, Ebenen) der Geometrie bilden ein System von Dingen, welches bei Aufrechterhaltung sämtlicher genannten Axiome keiner Erweiterung mehr fähig ist, d.h.: zu dem System der Punkte, Geraden, Ebenen ist es nicht möglich, ein anderes System von Dingen hinzuzufügen, so dass in dem durch Zusammensetzung entstehenden System sämtliche aufgeführten Axiome I-IV, V 1 erfüllt sind⁶³.

The axiom, however, appeared in print for the first time in the French edition, in 1900, in the following form.

Au système de points, droites et plans, il est impossible d'adjoindre d'autres êtres de manière que le système ainsi généralisé forme une nouvelle géométrie où les axiomes des cinq groupes I-V soient tous vérifiés; en d'autres termes: les éléments de la Géométrie forment un système d'êtres qui, si l'on conserve tous les axiomes, n'est susceptible d'aucune extension⁶⁴.

There is also an axiom of completeness for the axiomatization of real numbers in *Über den Zahlbegriff*, published in 1900.

IV.2 (Axiom of Completeness) It is not possible to add to the system of numbers another system of things so that the axioms I, II, III, and IV 1 are also all satisfied in the combined system; in short, the numbers form a system of things which is incapable of being extended while continuing to satisfy all the axioms⁶⁵.

Furthermore, from the seventh edition onward the Completeness Axiom is replaced by a Linear Completeness Axiom, which in the context of the other axioms implies the Axiom of Completeness in the apparently more general form.

V.2 (Axiom of Line Completeness) It is not possible to extend the system of points on a line with its order and congruence relations in such a way that the relations holding among the original elements as

⁶³[68], p. 16.

⁶⁴[65], p. 25.

⁶⁵[66], p. 1094 in [25]. In German: *IV.2 (Axiom der Vollständigkeit) Es ist nicht möglich dem Systeme der Zahlen ein anderes System von Dingen hinzuzufügen, so dass auch in dem durch Zusammensetzung entstehenden Systeme bei Erhaltung der Beziehungen zwischen den Zahlen die Axiome I, II, III, IV.1 sämtlich erfüllt sind; oder kurz: die Zahlen bilden ein System von Dingen, welches bei Aufrechterhaltung sämtlicher Beziehungen und sämtlicher aufgeführten Axiome keiner Erweiterung mehr fähig ist.*

well as the fundamental properties of the line order and congruence following from Axioms I-III and from V.1 are preserved⁶⁶.

The literal translation of the Axiom of Completeness is the following.

V.2 (Axiom of Completeness) The elements (points, straight lines, planes) of geometry form a system of things that, compatibly with the other axioms, can not be extended; i.e. it is not possible to add to the system of points, straight lines, planes another system of things in such a way that in the resulting system all the axioms I-IV, V.1 are satisfied.

In order to set about analyzing the content of this axiom, it is important to understand the terms involved: Axiome, Dingen, Geometrie. We tried before to give an idea of Hilbert's conception of axioms, finding a notion of intuition at its base, but encountering Frege's problem: how it is possible to match intuition and axioms? Moreover we hinted to a weak form of realism in Hilbert's conception of the geometrical objects and some logical difficulties, again raised by Frege, in understanding what the implicit definitions define. Finally we would like to understand what Hilbert means by geometry. However the main outcome of the analysis of the Axiom of Completeness is the recognition that it is not possible to keep separate Hilbert's notion of geometry from the role that this axiom plays in the process of its axiomatization. This fact will also explain at once how the Axiom of Completeness is to be considered as a solution to Frege's problem.

In what follows, first of all, we will try to understand the notion of geometry underlying the axiom system of the *Grundlagen der Geometrie*, explaining the feature of completeness with respect to the axioms of geometry. Then, we will try to understand the relevance of analytic geometry in Hilbert's work and the role of the Axiom of Completeness in the foundation of geometry.

Completeness of the axioms

In the lectures on projective geometry in 1891, Hilbert divides geometry in three parts:

The divisions of geometry.

1. Intuitive geometry.

⁶⁶In [84], p.26.

2. Axioms of geometry.
(investigates which axioms are used in the established facts in intuitive geometry and confronts these systematically with geometries in which some of these axioms are dropped)
3. Analytical geometry.
(in which from the outset a number is ascribed to the points in a line and thus reduces geometry to analysis)⁶⁷.

There is here an important distinction: the one between geometry and geometries. It is also possible to find this distinction in the *Grundlagen der Geometrie*, but for orthographic reasons it can be found only in the French version of 1900, where in the statement of the Axiom of Completeness we can find the distinction between *Géométrie* and *géométrie*. The presence of new additions and comments indicates that Hilbert followed closely the editing of this translation⁶⁸. From now on, with Geometry we mean the intuitive theory that is the object of formalization in the *Grundlagen der Geometrie*. In this section we want to explain the formal characters of Hilbert's work in the foundation of Geometry and why he does not accept uncritically the notion of space.

Prior to the problem of the matching between Geometry and the axiom system presented in the *Grundlagen*, one may ask why Hilbert chose to formalize euclidean geometry, and then he added the Axiom of Completeness, in order to develop analytic geometry. As we will see, his aim was not to give a foundation to analytic geometry. Then, why Hilbert decided to complement the system of the *Festschrift* with the *Axiom der Vollständigkeit*? These questions brake down into two: 1) why euclidean geometry? 2) why analytical tools? Now, we want to clear the first part, deferring to the next section the answer to the second. To this aim the following quotation is helpful.

We arrive now to the construction of geometry, in which axiomatics was fully implemented for the first time. In the construction of arithmetic, our real point of departure was in its intuitive (*anschaulischen*) foundation, namely the concept of natural number (*Anzahlbegriff*) which was also the starting point of the genetic method. After all, the number system was not given to us as a network of concepts (*Fachwerk von Begriffen*) defined by 18 axioms. It was intuition that led us in establishing the latter. As we have started from the concept

⁶⁷[58], p. 3.

⁶⁸In the volume [52] there is a careful account of the editorial vicissitudes of the French translation.

of natural number and its genetic extensions, the task is and naturally remains to attain a system of numbers which is as clear and as easily applicable as possible. This task will evidently be better achieved by means of a clearly formulated system of axioms, than by any other kind of definition. Thus it is the task of every science to establish on the axioms, in the first place, a network of concepts, for which formulation we let intuition and experience naturally serve as our guides. The ideal is, then, that in this network all the phenomena of the domain in question will find a natural place and that, at the same time, every proposition derivable from the axioms will find some application⁶⁹.

As we tried to show before, the contextual intuition that guides Hilbert's foundation of geometry is a mixture of experience and pre-formal knowledge. Then the reference to experience and the possibility to apply the theorems of geometry is not surprising. Indeed, as it well explained in [23], at that time Hilbert thought that euclidean geometry was the 'right' geometry to be applied to the outer world. However, the relevance of euclidean geometry for describing reality does not imply that the form of intuition that Hilbert acknowledges is a kantian intuition. Quite the contrary, we argued that this contextual intuition is an intellectual intuition *that* - a propositional attitude - contrary to Kant's sensible intuition, that acts in perception.

However, contrary to other geometrical investigations of that time - as for example Klein's representation of geometries as groups of transformations over manifolds - Hilbert's work did not have the goal to analyze the nature of space, but to make an axiomatic inquire of our geometrical intuitions. These intuitions are prior to the concept of space and hence they cannot presuppose anything about it.

In [69], Hilbert too contributed to the clarification of the nature of the space, assuming continuity since the beginning. However, since a foundation and not just a classification was sought in the *Grundlagen der Geometrie*, Hilbert sees his work as a contribution to the *kritische* stage of the development of Geometry. Then Hilbert's task is to analyze critically the continuity assumption hidden in the intuition of space. Thus, following the basic principle of the axiomatic method of deepening the foundations, Hilbert tries to elucidates the more fundamental principles of Geometry.

Here is outlined one of the most difficult tasks of Hilbert's axiomatization of Geometry: to find a system of axioms able to formalize all the means, also

⁶⁹[71], p. 35-36.

analytical, used in geometrical proofs. Linked to these problems, there are considerations on the purity of method, but we will face them later. Here it is sufficient to say that Hilbert is not concerned with problems of uniformity of methods of proofs⁷⁰.

In the same lectures on projective geometry we can find the following sentence, which still suffers from a conception that shortly thereafter would be radically changed.

Geometry is the theory about the properties of space⁷¹.

However, in Hilbert's lectures for the summer semester, in 1894, entitled *Die Grundlagen der Geometrie* there is no longer an explicit definition of geometry, but rather of geometrical facts. It is also worth noting that in the 1899 *Grundlagen der Geometrie* we do not find a definition of space.

Among the phenomena, or facts of experience that we take into account observing nature, there is a particular group, namely the group of those facts which determine the external form of things. Geometry concerns itself with these facts⁷².

Here there is a subtle, but basic, shift in addressing the problem of a foundations for Geometry. Hilbert is not trying to define what Geometry is by means of the axioms, on the contrary he just tries to find a simple, independent and consistent system of axioms that allows a formalization of all geometrical facts. The completeness of the axioms to which Hilbert refers at the beginning of the *Grundlagen der Geometrie* has therefore to be understood in the sense of maximizing the class of known⁷³ geometrical facts that can be proved thanks to the proposed system of axioms.

In 1894, Hilbert was explicit in describing the goals he wanted to achieve by means of his foundational studies.

Our colleague's problem is this: what are the *necessary* and *sufficient*⁷⁴ conditions, independent of each other, which one must posit

⁷⁰This is a concern typical of a classical conception of the axiomatic method that dates back to Aristotele: "[...] we cannot in demonstrating pass from one genus to another. We cannot, for instance, prove geometrical truths by arithmetic" (Posterior Analytics: 75a29-75b12). For an historical survey of this subject see [24].

⁷¹[58], p. 5.

⁷²[60], p. 7.

⁷³We will see later that Hilbert thinks that the axiomatic method does not allow to discover new true propositions.

⁷⁴My emphasis.

for a system of things, so that every property of these things corresponds to a geometrical fact and vice versa, so that by means of such a system of things a complete description and ordering of all geometrical facts is possible⁷⁵.

Hilbert's statement of intent is clear: find necessary and sufficient conditions to describe every geometrical fact. Then the problem of defining Geometry disappears, since it is implicitly and extensionally defined by geometrical facts. This is precisely the purpose of an analysis conducted with the axiomatic method. As a matter of fact, in 1902, Hilbert says:

I understand under the axiomatic exploration of a mathematical truth [or theorem] an investigation which does not aim at finding new or more general theorems being connected with this truth, but to determine the position of this theorem within the system of known truths in such a way that it can be clearly said which conditions are necessary and sufficient for giving a foundation of this truth⁷⁶.

Thanks to this precise statement, we can make some general consideration on the axiomatic method. First of all, this method is primarily designed to formalize an already developed field of knowledge. Therefore it is a method that can be applied when a science has already reached a sufficient level of maturity, such that it can be divided from other branches of knowledge. Then it is possible to develop an intuition internal to the theory capable of identifying the class of facts that have to be axiomatized, together with the basic principles that allow their proofs. Moreover, it should be noted that Hilbert says explicitly that the goal of the axiomatic method is a clear understanding of geometrical proofs, thanks to the analysis of the *meaning* of the axioms⁷⁷, and not the discovery of new theorems.

Besides, Hilbert does not consider the axiomatic method primarily as a source of mathematical rigor⁷⁸, but rather, from a methodological perspective, as a tool which allows us to answer in an objective way the question: why some proofs are possible and some others are not.

One of Hilbert's greatest achievements in the field of the foundational studies has been to recognize not only the distinction of levels between theory and

⁷⁵[60], p. 8.

⁷⁶[67], p. 50.

⁷⁷Recall the quote from the introduction of the *Grundlagen der Geometrie* (p. 1), where Hilbert declares that the aim of the book is "to bring out as clearly as possible the significance [*Bedeutung*] of the different groups of axioms".

⁷⁸See [134] in this respect.

metatheory, but also to understand that the metatheory was analyzable with mathematical tools. However, Hilbert considered meta-mathematical investigation as a deepening of knowledge about mathematics, and not as a genuine source of new results; contrary to his subsequent work and what the development of twentieth-century logic would have show⁷⁹.

In 1908, Hilbert still expresses opinions similar to those of 1902.

In the case of modern mathematical investigations, ... I remember the investigations into the foundations of geometry, of arithmetic, and of set theory—they are concerned not so much with proving a particular fact or establishing the correctness of a particular proposition, but rather much more with carrying through the proof of a proposition with restriction to particular means or with demonstrating the impossibility of such a proof⁸⁰.

If the main point in axiomatizing Geometry is the axiomatization of all geometrical facts, what distinguishes them from other facts, whether empirical or mathematical? Hilbert answers this question clearly: the axioms of the *Grundlagen*, but he is not clear on what motivates his choice; and it is on this terrain that Frege's problem regains strength.

Axiom der Vollständigkeit

Hilbert's aim is to find necessary and sufficient conditions to prove all relevant geometrical facts. So that it is possible to define Geometry as the field of knowledge whose true propositions are the theorems that can be proved by means of the axioms presented in the *Grundlagen der Geometrie*. However, the axioms that we can find in the *Festschrift* are not sufficient for this purpose. This observation brings us to the answer to the question: why Hilbert supplemented his system of axioms with the *Axiom der Vollständigkeit*? and why analytical tools?

For what concerns the second question, as we just saw, Hilbert's critical investigation of our geometrical intuitions was also meant to take care of the continuity principles that are deeply linked with our intuition of space. This partially explains Hilbert's attention to analytical geometry. Judson Webb, in

⁷⁹Following this line of reasoning it is perhaps reasonable to find an explanation for Hilbert's mild reaction to Gödel's incompleteness theorems. However, the quotes above are from the first period of Hilbert's interest on foundational issues i.e. before the twenties; while Gödel's theorems were proved in 1930.

⁸⁰[72], p. 72. Translation in [134], p. 100.

[188], suggests that Hilbert's goal was to free Geometry from analytical considerations, in order to restore its dignity and autonomy. However, more than historical considerations, there is also another methodological reason that led Hilbert to deal with analytic geometry.

As a matter of fact, logic and analysis always play an important role in Hilbert's foundational work. As we have already noticed, in 1922 Hilbert expresses this view in these terms:

This circumstance corresponds to a conviction I have long maintained, namely, that a simultaneous construction of arithmetic and formal logic is necessary because of the close connection and inseparability of arithmetical and logical truth⁸¹.

The foundational view proposed here by Hilbert is radically different from the standard one that tries to ground all mathematical knowledge on a single concept. This is what the logicist tradition - like Frege⁸² and Russell - tried to do with logic; or how a set theoretical, functional or categorical foundation of mathematics is interpreted in modern times. Rather Hilbert was convinced that the tools offered by logic and arithmetic were essential for a proper development of any branch of mathematics. In other words, Hilbert does not seem to have any ontological or epistemological commitments in using numbers and logic; rather it is a methodological concern⁸³.

In all exact sciences we gain accurate results only if we introduce the concept of number⁸⁴.

However, according to Hilbert these tools must be investigated in a critical manner.

But if science is not to fall into a bare formalism, in a later stage of its development it has to come back and reflect on itself, and at least verify the basis upon which it has come to introduce the concept of number⁸⁵.

⁸¹[76], pp. 1131-1132 in [25].

⁸²However we will see in the next chapters that this aspect does not exhaust Frege's foundational work.

⁸³This is why it is not easy to attribute any philosophical position to Hilbert, although the problems he addresses have obvious philosophical implications.

⁸⁴[60]. In German: *In allen exakten Wissenschaften gewinnt man erst dann präzise Resultate wenn die Zahl eingeführt ist.*, in [52], p. 194.

⁸⁵[61]. In German: *Aber wenn die Wissenschaft nicht einem unfruchtbaren Formalismus anheimfallen soll, so wird sie auf einem späteren Stadium der Entwicklung sich wieder auf sich selbst besinnen müssen und mindestens die Grundlagen prüfen, auf denen sie zur Einführung der Zahl gekommen ist*, in [52], p. 194.

In order to introduce the concept of number in Geometry, Hilbert defines a calculus of segments and then he uses the axiomatic method to show which algebraic properties of the calculus follow from the validity of geometrical propositions.

Here the axiomatic method is used with the aim of understanding the demonstrative role of the axioms of Geometry. The idea is to generate a coordinate system internal to Geometry, showing that some fundamental theorems imply certain properties of numbers that are used as coordinates. In this way, many properties of the system of real numbers are not imposed from outside, as in the standard presentation of analytic geometry, but arise from geometrical considerations.

For example, the validity of Pappus's theorem (called Pascal's theorem by Hilbert) is used to show that the multiplication that is possible to define on the coordinate system must necessarily be commutative. Thanks to axioms I-V1 Hilbert shows that the coordinate system thus defined forms an Archimedean field. However, since this Archimedean field can be countable, it is clear to Hilbert that the geometry that satisfies all axioms I-V1 can not be immediately identified with analytic geometry.

Indeed, the domain of the latter is uncountable, because it makes use of all real numbers. So, Hilbert's major concern is to define axiomatically a bijection between the points of a straight line and the real numbers. The solution of this problem is precisely the mathematical content of the Axiom of Completeness

If in a geometry only the validity of the Archimedean Axiom is assumed, then it is possible to extend the set of points, lines, and planes by "*irrational*" elements so that in the resulting geometry on every line a point corresponds, without exception, to every set of three real numbers that satisfy the equation. By suitable interpretations it is possible to infer at the same time that *all* Axioms I-V are valid in the extended geometry. Thus extended geometry (by the adjunction of irrational elements) is none other than the ordinary space Cartesian geometry in which the completeness axiom V.2 also holds⁸⁶

In this quotation it is possible to see how the Axiom of Completeness is used to fill that gap between Hilbertian plane geometry and analytic geometry. The way to achieve this is by adding irrational elements to the coordinate system presented in the *Grundlagen der Geometrie*. As a matter of fact, the axiomatization of the real numbers is simultaneous with the introduction of the Axiom

⁸⁶[81], pp. 35-36.

of Completeness for geometry⁸⁷.

The irrational elements are also called ideal elements, by Hilbert. However, he immediately makes it clear that the ideal character of these elements is only relative to the specific presentation of the system⁸⁸.

That to every real number there corresponds a point of the straight line does not follow from our axioms. We can achieve this, however, by the introduction of ideal (irrational) points (Cantor's Axiom). It can be shown that these ideal points satisfy all the axioms I-V [...]. Their use is purely a matter of method: *first with their help is it possible to develop analytic geometry to its fullest extent.*⁸⁹

The reference to irrational elements echoes the problem of the purity of methods, which is explicitly mentioned by Hilbert. However Hilbert's solution is not to restrict the demonstrative tools, allowing just those conforming to the essential properties of the objects of the theory. Indeed, the same idea of an extra-logical property of mathematical objects is contrary to the conception of axiomatic method, as Hilbert made clear also in correspondence with Frege.

In fact, the geometric investigation carried out here seeks in general to cast light on the question of which axioms, assumptions or auxiliary means are necessary in the proof of a given elementary geometrical truth, and it is left up to discretionary judgement [*Ermessen*] in each individual case which method of proof is to be preferred, depending on the standpoint adopted⁹⁰.

Since its aim is to show the possibility or the impossibility of a proof, the axiomatic method is the highest expression of the search for the purity of methods. Indeed, in an interlineated addition to the 1898/1899 lessons Hilbert writes: "Thus, solution of a problem impossible or impossible with certain means. With this is connected the demand for the purity of methods⁹¹". Hilbert considers the application of the axiomatic method as a precondition for any consideration on the purity of methods. Indeed, thanks to that it is possible to clear necessary conditions for the proof of a mathematical theorem. So, the choice of the

⁸⁷Remember that the Axiom of Completeness first appears in [66] and then in the first French edition of *Grundlagen der Geometrie*

⁸⁸In [74], p. 149, Hilbert says, "The terminology of ideal elements thus properly speaking only has its justification from the point of view of the system we start out from. In the new system we do not at all distinguish between actual and ideal elements".

⁸⁹[63], pp. 166-167.

⁹⁰[81], pp. 82.

⁹¹See [62], p. 284 in [52].

demonstrative methods becomes a subjective question, since it does not depend on the nature of the problem.

This basic principle, according to which one ought to elucidate the possibility of proofs, is very closely connected with the demand for the ‘purity of method’ of proof methods stressed by many modern mathematicians⁹². At root, this demand is nothing other than a subjective interpretation of the basic principle followed here.⁹³.

Once that it is clear that problem of the purity of methods is nothing else than finding the right conditions for the proof of a theorem, and once it is clear the importance of the notion of number, we can easily understand why Hilbert wanted to develop analytic geometry “in his fullest extent”. But can we consider the *Grundlagen der Geometrie* as a foundation for analytic geometry? We believe not, because Hilbert used analytic geometry to the full in the application of the axiomatic method to Geometry; for example in the proof that it is possible to develop a non-Desarguean geometry. Indeed Hilbert’s goal was not a foundation of analytic geometry in the contemporary standard sense, short of running into an obvious circularity in his reasoning.

In this context we can also explain how the axiomatic method can be used to improve our mathematical knowledge, since this method neither is used for the discovery of new theorems - remember that Hilbert says: “I understand under the axiomatic exploration of a mathematical truth [or theorem] an investigation which does not aim at finding new or more general theorems”⁹⁴ - nor, following [134], in the search for rigor in mathematics.

This basic principle [to enquire the main possibility of a proof] seems to me to contain a general and natural prescription. In fact, whenever in our mathematical work we encounter a problem or conjecture a theorem, our drive for knowledge [*Erkenntnistrieb*] is only then satisfied when we have succeeded in giving the complete solution of the problem and the rigorous proof of the theorem, or when we recognize

⁹²Remember that Hilbert’s proofs were not easily accepted by the mathematical community of the late nineteenth century. Therefore, instead of restricting the methods of proof, Hilbert put forward an analysis of proofs that does not rest on external considerations on the nature of mathematical entities, but that aims to show if a demonstrative tool is necessary in a particular proof. Moreover, given the link between methods of proof and axioms, the justification of the means used in a proof is brought back to the justification of the axioms and to the inference rules.

⁹³[81], p. 82.

⁹⁴[67], p. 50.

clearly the grounds for the impossibility of success and thereby the necessity of the failure⁹⁵.

We can clearly see in Hilbert's thought a dichotomy between the subjective side of the choice of the demonstrative tools and the objective side of the logical relations between mathematical propositions. However, the objectivity of mathematics is presupposed in Hilbert's foundation of Geometry. Indeed the correctness of the analytical methods employed in his foundational work is not argued. This would require a consistency proof for analysis; but this is a different task that does not pertain to the foundation of Geometry. The emphasis given to the objectivity of the logical relations between mathematical propositions - obtained by means of rigorous proofs - is just a matter of justification of the methods of proof, hence of the axioms. We need to stress here the difference between giving a foundation and giving a justification, and it is here that we can see the novelty of Hilbert's foundations with respect to the standard ones⁹⁶. As a matter of fact, if we try to interpret Hilbert's foundational efforts as an attempt to ascertain the truth of the geometrical propositions and to clear the meaning of the concepts of point or line, we have to face the problem that, on one side, the truth of geometry rests on the correctness of the methods used - mainly on the truth of analysis - and, on the other hand, that there is no clear definition of the basic geometrical objects, nor of space or geometry. Moreover, if we try to see the *Grundlagen der Geometrie* as a reduction of Geometry to analysis, we ran into an apparent circularity of the argumentation, because analytical methods are used in order to show the necessity of the axioms that should give a foundation for analytic geometry. This seems to support the autonomy of Hilbert's foundation of mathematics⁹⁷. But this point of view is incorrect: a reduction-style-foundation is not found, because there is no foundation in this sense. Hilbert does not try to reduce Geometry to analysis, nor he tries to find an ontological classification of geometrical entities. On the contrary he tries to justify the possibility to give a formal treatment of an intuitive theory; and in this sense we have to consider Hilbert's work as foundational. Even if Hilbert avoids any extra-logical commitments about objects and methods of proof, however, the way he constructs the formal theory for Geometry is not autonomous from extra-mathematical considerations; we can say philosophical. Indeed Hilbert justifies the formalization of a theory appealing to intuition, logical reasoning and the concept of number. These concepts seem to be for Hilbert the starting points for any mathematical knowledge and construction. Appealing to these

⁹⁵[81], p. 82.

⁹⁶We will elaborate on this in the next chapter

⁹⁷See for example [33].

notions he is able to say that the axioms presented in the *Grundlagen der Geometrie* formalize precisely analytic geometry, in its intuitive character - and in the extension of its theorems. This choice is indeed philosophical, because it implies a precise definition of mathematics: the science of calculation, carried out by logical means. This conception is quite astonishing if we think of the development of mathematics at that time, pushing towards always more abstract methods. However it explains the role of arithmetic in Hilbert's conception of mathematics, throughout all his work; where arithmetic is here to be understood in the widest sense, including also transfinite cardinal arithmetic.

Indeed Hilbert's notion of logic is very wide; even too much to ascribe to him a sort of logicism in the normal sense that this word has gained in the philosophy of mathematics tradition⁹⁸. Indeed we could call it an arithmetical-logicism. Then, notice that, even if the use of the Axiom of Completeness can be seen as a good solution for the foundation of Geometry, we are still left with the justification of the coherence of this solution with the context of the axiomatic method. In other words: if Hilbert does not accept any extra-logical tool, in a foundation of a theory, how it is possible that talking about objects, domains and their extensions we still remains in the realm of logic?

We cannot find an explicit answer to this concern in Hilbert's writings, but Husserl, in his notes after his conference *Das Imaginäre in der Mathematik* held in Göttingen in 1901, reports a comment of Hilbert during the discussion of the possibility to give a complete⁹⁹ foundation of number theory. Indeed Husserl writes Hilbert's objection in this way: "Had I [Husserl] been justified to say that every sentence entailing only the positive integers is either true or false on behalf of the axioms for the positive integers?¹⁰⁰" In order to report Hilbert's argument against this possibility, we report part of Husserl's note.

When we say that a sentence is decided on the ground of the axioms of a domain, what can we use besides the axioms? *Alles Logische. Was ist das?* All propositions that are free of special features of a domain of knowledge, which independently by all "special axioms", applies to all matter of knowledge. But here we have a wide range of possibilities. The domain of algorithmic logic, the domain of numbers, the domain of combinatorics, the domain of the general numerical series - and the theory of ordinal numbers. And fi-

⁹⁸See [29] in this respect.

⁹⁹Here complete is intended in one of the different senses that it is possible to understand from Husserl's lecture. See [158] and [112], in this respect.

¹⁰⁰Majer, in [113], sees in this quotation Hilbert's awareness of the problem of the relation between the formal and the informal side of mathematics.

nally, even the most general theory of set isn't it purely logical? [...] The algorithmic logic is sufficient to derive the theorems of incidence from Pascal's Theorem (without the axiom of continuity); the logic of numbers is at stake when we use Archimedes, and in order to use the Axiom of Completeness we have to appeal to set theory.

We can see here Hilbert's conviction that, even in number theory, we need to use all the power of 'logic' - to be intended here in Hilbert's very broad sense - in order to be able to decide all the sentences "entailing only the positive integers". Coming back to the problem of the coherence of the Axiom of Completeness with the context of the axiomatic method, then we can say that, granting that set theory is a part of logic, Hilbert's solution for the completeness of Geometry is perfectly acceptable and 'logical'.

The completeness sought by Hilbert is a completeness of the domain of discourse: the possibility to fix and define what Geometry is¹⁰¹, knowing since the beginning what are the more important geometrical theorems. Indeed, in 1905, in [71], Hilbert defines what is a complete axiom system in the following way: "it actually have all facts under consideration (*alle vorgelegten Thatsachen*) as logical consequences". This may be seen, in a terminology introduced by Detlefsen, a qualificatively completeness of a theory T :

For every sentence σ in a language \mathcal{L} , if σ is \mathcal{D} , then $T \vdash \sigma$.

However if \mathcal{D} stands for the set of geometrical facts under consideration, the problem is then to determine them in a clear way. Hilbert's solution, appealing to the idea that axioms are implicit definition, is to collapse the task of this determination to the implicit definitions given by the axioms. Then the axioms of the *Grundlagen* define not only points, lines and planes but also Geometry in its integrity. We believe that the the Axiom of Completeness brings a normative - more than a descriptive - element in the axiomatization of Geometry.

Recalling that Hilbert's goal was to find necessary and sufficient conditions for proving the more relevant geometrical facts, we can affirm that the axioms of groups I-IV, together with the Axiom of Archimedes, are necessary conditions for the development of analytic geometry, and the Axiom of Completeness plays the role of a sufficient condition to adapt the formal presentation given in the *Grundlagen der Geometrie* to the intuitive idea of a geometrical theory that makes use of the whole class of real numbers. Already in 1872 Cantor, as also Dedekind, felt the need for an axiom to make compatible these two sides of Geometry.

¹⁰¹Of course there is here a mixture of semantical and syntactical completeness, and of categoricity, but none of them in particular.

In order to complete the connection [...] with the geometry of the straight line, one must only add an axiom which simply says that conversely every numerical quantity also has a determined point on the straight line, whose coordinate is equal to that quantity [...] I call this proposition an *axiom* because by its nature it cannot be universally proved. A certain objectivity is then subsequently gained thereby for the quantities although they are quite independent of this¹⁰².

So we can distinguish two different kinds of axioms: the ones that are *necessary* for the development of a theory and the *sufficient* ones used to match intuition and formalization.

In the lectures that precede the first edition of the *Grundlagen der Geometrie* Hilbert proposed that continuity could be formalized in ways similar to Cantor's¹⁰³ and Dedekind's¹⁰⁴, which were able, together with the other axioms, to guarantee the existence of a bijection between the point lying on a straight line and the real numbers. However, Hilbert soon realized that there was need of less continuity for developing Geometry. Thus, following the general principle of the axiomatic method of outlining the necessary conditions, Hilbert chose the Axiom of Archimedes, in order to answer the why questions that motivated his work. In a letter to Frege, on December 29th 1899 (in [36], pp. 38-39), Hilbert wrote: "It was of necessity that I had to set up my axiomatic system: *I wanted to make it possible to understand* those geometrical proposition that I regard as the most important results of geometrical inquiries" (my emphasis).

By the above treatment the requirement of continuity has been decomposed into two essentially different parts, namely into Archimedes' Axiom, whose role is to prepare the requirement of continuity, and the Completeness Axiom which *forms the cornerstone of the entire system of axioms*. The subsequent investigations rest essentially only on Archimedes' Axiom and the completeness axiom is in general not assumed¹⁰⁵.

We are now in the position to understand the importance of the Axiom of Completeness, even if it "is in general not assumed". Its role is to fix uni-

¹⁰²In [19], p. 128.

¹⁰³(Cantor continuity axiom): every descending (with respect to the relation of inclusion) sequence of non empty real intervals has non-empty intersection.

¹⁰⁴(Dedekind continuity axiom): given any partition of the real line in two classes $A \leq B$ (i.e. $\forall a \in A$ and $\forall b \in B$, we have $a \leq b$) there is a real number c such that $a \leq c \leq b$, for every $a \in A$ and $b \in B$.

¹⁰⁵[84], p. 28. From the seventh German edition onward.

vocally the relevant set of sentences that forms the theory of Geometry; as a consequence, thanks to the contextual intuition that helps in fixing the set \mathcal{D} , the Axiom of Completeness finds a practical justification. Due to its sufficient character and its consequent possibility to define Geometry, Hilbert thought to have answered Frege's problem: we do not need to know that a point is to determine if a pocket watch is a point or not, but only if the right geometrical theorems apply to it. However if a definition of point, or line, or even of Geometry is not needed outside the formal system of Geometry, where Hilbert's axioms come from? Hilbert found a way to justify the choice of his axioms thanks to necessary and sufficient conditions, but a starting point is needed, in order to lay down the axiomatic setting for Geometry. Then, we see that Frege's problem is not fully solved. Some sort of intuition or knowledge is needed in order to determine, from the outset, which are the relevant theorems that we want to formalize. This is exactly the contextual intuition we described before: a mixture of experience and spatial intuition that helps Hilbert to avoid any commitments with the position that Geometry is an *a priori* science. Then Hilbert's work answers only the problem of the matching between intuition and formalization, but not the questions about the role of intuition in the axiomatization of Geometry, as it is the case with Frege's objection. Then, we can say that, thanks to the Axiom of Completeness, Hilbert's managed to shift Frege's problem from 'what is a geometrical object?' to 'what is Geometry?', but he failed to solve it.

However, a remarkable aspect of the Axiom of Completeness is that its sufficiency is not essential for the proof of the most important geometrical facts; its role is mainly conceptual. Following this line of reasoning, the Axioms of Completeness can be seen as the first, historically documented, instance of Skolem's paradox; of course Hilbert was not driven by considerations on the nature of logic, but the Axiom of Completeness can be seen as a way of solving the problem of the existence of a theory for analytic geometry that cannot prove that real numbers are uncountable.

Indeed Hilbert seems to argue in favor of an connection with the continuum independent from logic. Writing against the genetic method that tries to define real numbers, starting with natural numbers, Hilbert says:

The totality of real numbers, i.e. the continuum [...] is not the totality of all possible series of decimal fractions, or of all possible laws according to which elements of a fundamental sequence may proceed. It is rather a system of things whose mutual relations are governed by the axioms set up and for which all propositions, and only those, are true which can be derived from the axioms by a finite

number of logical processes¹⁰⁶.

In other words, this matching of intuition and formalization, which tries to harmonize the intuitions behind the system of real numbers and the real line, is the intuitive content of the fifth group of axioms of the *Grundlagen der Geometrie*, in the same line as Hilbert's work on the concept of number: “[w]e therefore recognize the agreement of our number-system with the usual system of real number¹⁰⁷”.

To summarize, Hilbert's analysis of the notion of continuity led him to formalize the Axiom of Completeness as a sufficient condition for analytic geometry, in the form of a *maximality* principle about propositions, once this theory is given in an intuitive way.

There are some presuppositions that need to be made explicit in Hilbert's ideas. First of all, the scope of the axiomatization needs to be known right from the beginning. Moreover, Hilbert chose to include analytical tools in the formalization of Geometry. This choice seems surprising if we consider that at that time the development of geometry led to the introduction of very abstract topological concepts, not only from classical geometry, but also from considering calculation as the most important tool in Geometry. The answer to this problem is to be found in Hilbert's conviction that “In all exact sciences we gain accurate results only if we introduce the concept of number¹⁰⁸”.

All this shows how important logic and arithmetic are for Hilbert. So, together with the fact that formalization needs to take care of demonstrative methods used in a certain field of knowledge, it makes clear the conceptual background from which Hilbert's ideas developed and then shaped his proof theory.

1.2.3 Second period

In order to stress and explain the difference between the first and the second periods, and the corresponding two different kinds of intuition, there is a first

¹⁰⁶In [57], p. 1105 in [25]. Again in this quotation we could hear an echo of realism: the existence of the continuum does not depend on the axiomatization, but it is something stable that we get to know thanks to our mathematical work in a way that has a phenomenological flavor. This interpretation of Hilbert's philosophy of mathematics could also hint to an interesting parallel between Hilbert and Plato. What we try to understand in mathematics are *ideas* - recall the kantian *exergo* at the beginning of the *Grundlagen der Geometrie*. Indeed Plato was not platonist at all, since for him mathematical objects are abstractions from reality, but still a pale and incorrect image of the eternal ideas, of which only it is possible to predicate truth.

¹⁰⁷[55], p. 1095.

¹⁰⁸[60].

question that need to be answered: why Hilbert's solution of the first period does not have an analog - or an extension - in the second one? The easy answer is that, once it is the whole of mathematics that needs a foundation - and not a single theory - it is not possible to find a set of axioms that act as necessary and sufficient conditions for being a mathematical theorem, because mathematics is essentially incomplete. However, although this answer is true, it does not answer our question.

There is no clear answer in Hilbert's writings - and so we do not know if it has been a concern for him - but it is possible to point to two different reasons that would have surely been problematic in any attempt to do so.

The first one is related to the importance that the use of ideal elements has in mathematics, for Hilbert.

We come upon quite another, wholly different interpretation, or fundamental characterization, of the notion of infinity when we consider the method - so extremely important and fertile - of *ideal elements*. The method of ideal element has an application already in the elementary geometry of the plane.¹⁰⁹

Indeed this is one of the main reason to introduce the Axiom of Completeness: in order to introduce the irrational elements as the ideal elements with respect to a countable model of the other axioms. However, granting the freedom - and the open character, as we will see later - of the development of mathematics and the usefulness of the possibility to introduce ideal elements, how it is possible to formulate a completeness axiom for the entire mathematics? This question is not addressed by Hilbert, but it would be an insurmountable obstacle in any attempt to generalize the method that Hilbert used in the first period of his foundational reflection.

Secondly there is the problem of the formalization of logic. As a matter of fact, in the Twenties, Hilbert thought that Russell's formalization of logic was not adequate for the propose of a foundation, and since one of Hilbert's aim was to contrast the irrational pushes that disturbed the mathematical community, a careful formalization of the logical tools was needed for their use. In other words it was not possible a careless use of the logical background that was used to formulate an axiom as the Axiom of Completeness.

Then some new ideas were needed. Indeed, in the Twenties, when engaged in the foundations of mathematics for the second time, Hilbert's new conception of axiom mirrors a deeper enquiry about the concept of intuition, in the direction

¹⁰⁹[78], p. 372 in [55].

of a Kantian-style notion. Thanks to that Hilbert thought to have solved the problem of a safe foundation for mathematics.

We start with two quotations which sound very Kantian.

Instead, as a precondition for the applications of logical inferences and for the activation of logical operations, something must already be given in representation [*in der Vorstellung*]: certain extra-logical discrete objects, which exist intuitively as immediate experience before thought. [...] Because I take this stand point, the objects of number theory are for me [...] the sign themselves, whose shape can be generally and certainly recognized by us – independently of space and time, of the special conditions of the production of the sign, and of insignificant differences in the finished product¹¹⁰.

Kant taught [...] that mathematics treats a subject matter which is given independently of logic. Mathematics therefore can never be grounded solely on logic. [...] As a further precondition for using logical deduction and carrying out logical operations, something must be give in conception, viz. certain extralogical concrete objects which are intuited as directly experienced prior to all thinking¹¹¹.

In these quotations Hilbert's Kantism is clearly outlined. As Kant did before, Hilbert tries to give a foundation to the certainty of mathematical truths, not by means of logic, but reflecting on the very possibility of any mathematical knowledge. For Kant the *a priori* conditions for geometrical and arithmetical knowledge were the pure intuitions of space and time. Hilbert, going further in the same direction, gives a foundation to the certainty of mathematical truths by means of the sensible pure intuition of mathematical signs. He thinks that the intuition of mathematical symbols - sensible intuition, since symbols are written on some physical support, and pure, since it does not depend on the shape the signs assume - is necessary to make knowledge within a formal framework possible. So, since every piece of mathematics is formalizable, symbols are pre-conditions of any form of mathematical knowledge.

Hilbert's purposes are clearly kantian. It remains to be seen how much Hilbert's ideas towards the realization of those purposes are also kantian. The affinity of the two thinkers looks in fact merely verbal and, maybe, for Hilbert functional to philosophers' approval. Indeed at that time the forms of neo-Kantism were quite spread and often quite far from Kant's original ideas. Indeed

¹¹⁰[76], p. 1121 in [25].

¹¹¹[78], p. 376 in [55].

we can describe Hilbert's work as a *critical deduction of mathematical knowledge*: one of the tasks of the *Neue Fries'sche Schule*¹¹² founded by the neo-kantian philosopher Leonard Nelson, who was a colleague of Hilbert in Göttingen, during the Twenties.

The work *Naturerkennen und Logik* (1930) is the opportunity for a deep reflection on Hilbert's part on the philosophical meaning of this new conception. First of all Hilbert explicitly says that the older conception of the axiomatic method, offered at the beginning of the century, is not sufficient:

How do matters stand with this axiomatics, which is today on everybody's lips? Now, the basic idea rests on the fact that generally even in comprehensive fields of knowledge a few propositions – called axioms – suffice for the purely logical construction of the entire edifice of the theory. But their significance is not fully explained by this remark¹¹³.

It follows an analysis of the sources of human knowledge. Hilbert claims that there are not just intellect and sensation - in kantian terms - but there is a third way: “besides logic and experience we have a certain *a priori* knowledge of reality”¹¹⁴. The latter is possible, for Hilbert, thanks to the intuition of mathematical symbols - of “their properties, differences, sequences and contiguities¹¹⁵” - that formalize this knowledge. This intuition is pure and sensible, and deeply linked to the finitary method.

William Tait, in [163], showed with clear arguments that Kant's intuition is intuition *of*, since it is active in the process of perception. As far as this kind of intuition is concerned, also Hilbert's conception of intuition, in the second period, is of the same kind, since it is sensible. Nevertheless there is an important difference here between the two thinkers in what concerns the aspects of evidence linked to this mode of intuition. Indeed for Hilbert intuition is evident as far as it is a kind of knowledge, the one that is able to ground mathematical reasoning: “also [...] mathematical knowledge in the end rests on a kind of intuitive insight [*anshaulicher Einsicht*]”¹¹⁶. On the contrary intuition, for Kant, is not a kind of knowledge, since in the intuitive process lack the concepts under which the objects, given in the intuition, fall.

This conception of intuition and the way we handle mathematical symbols

¹¹²See [139] on this subject.

¹¹³[80], p. 1158 in [25].

¹¹⁴[80], p. 1161 in [25].

¹¹⁵[78], p. 376 in [55].

¹¹⁶[80], p. 1161 in [25].

determine the foundational axioms to be assumed for proof theory¹¹⁷. Indeed if we want to avoid an infinite regression, we must justify these axioms extra-mathematically. Then Hilbert’s idea is to appeal to the similarity between formalization and the way we are used to think mathematically. On this ground manipulation and calculation become two sides of the same idea. Moreover, Hilbert wanted to give a foundation to all mathematics and so he maintained that the foundational axioms formalize the “fundamental elements of mathematical discourse¹¹⁸”, that are, for Hilbert, pre-conditions of any knowledge within a formalized discourse.

The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds¹¹⁹.

Hilbert then goes a step further: he claims that our intuitions of symbols have not only an *a priori* character, but they also manifest typical features of evidence.

The subject matter of mathematics is [...] the concrete symbols themselves whose structure is immediately clear recognizable¹²⁰.

This is the main difference between the first and the second period, and also the main difference between Kant’s and Hilbert’s intuition. In the first period

¹¹⁷For an interesting study on how intuition is also used by Hilbert to justify the correctness of material deduction, i.e. the manipulation of symbols that takes place in intuitive arithmetic, see [106]. However there is no analysis of this use of intuition in Hilbert’s work, nor textual evidence. We agree with Legris, but we attain to Hilbert’s work.

¹¹⁸[77].

¹¹⁹[79], p. 475 in [55]. Hilbert’s interest for an application of Kant’s transcendental method to mathematics and the possibility to give an extra-logical foundation to mathematics can be dated back to the first period of his foundational work. In 1905 Hilbert defines the following principle and he calls it “axiom of reasoning” or “philosopher’s *a priori*”: “I have the ability to think things, and to designate them by simple signs ($a, b, \dots X, Y, \dots$) in such a completely characteristic way that I can always recognize them again without doubt. My thinking operates with these designated things in certain ways, according to certain laws, and I am able to recognize these laws through self-observation, and to describe them perfectly” (in [71], p. 219). In this early period Hilbert is interested in what we could call the transcendental aspect of our reasoning about mathematics, namely the *deepening the foundations* of the axiomatic method. On the contrary in the second period Hilbert’s aim is to show how the axioms for his proof theory make a “protocol of the rules according to which our thinking actually proceeds”; that, in kantian terms, amount in a deduction of the axioms from *a priori* principles. See [140] for an interesting discussion of the philosophical background related to the search for an *a priori* foundation of mathematics among philosophers and mathematicians in Göttingen in the late Twenties.

¹²⁰[78], p. 376 in [55].

intuition and evidence are kept apart, in the second one they coincide, thanks to a kantian-style - for Hilbert - notion of intuition, that is like Leibniz's intuitive knowledge, as described in [107]: clear, distinct and adequate.

Hilbert thinks that this kind of knowledge makes basic arithmetic safe and it makes possible to extend knowledge to the transfinite domain, assuming consistency of the formal system that incorporates mathematics. Logical tools are just, as Kant said, harness of the reason. Hence, in this second period symbolic logic is a tool used for a complete deployment and correct use of intuition. Moreover intuition, since it coincides with evidence, is able to give knowledge from and certainty to finitary arithmetic and logic. There is though an objection that easily arises: if Hilbert is right and he managed to give an extra-mathematical foundation of mathematical knowledge, thanks to this kantian-style intuition, why do we need a consistency proof for arithmetic? The solution is to be found in the reasons that motivate the invention of the proof theory. Indeed Hilbert's goal was to secure infinitary mathematics - part of which is also general arithmetic - by means of a finitary consistency proof, that is to justify the use of mathematical symbols in a meaningless context, as a source of knowledge. Then a consistency proof for arithmetic does not secure contentual mathematics, for which intuition is already the source of its soundness, but it gives a safe foundations to all mathematics, since it allows a consistent use of the same symbols, that are used in contentual mathematics, in a more abstract and also meaningless way. Where intuition is not available, then manipulation of symbols replaces intuitive arithmetic, once we know that this manipulation cannot generate any contradiction¹²¹.

It is important to stress that the intuition underlying Hilbert's foundational studies, at the time of the discovery of the proof theory, even if it is an intuition *of*, does not witness an evolution towards a stronger realism in Hilbert's thought. Indeed the intuition *of* in this later period is not a philosophical shelter from mathematical problems. It is not intuitions of the numbers, whatever they are. It is the intuition that witnesses the accordance between the formalization of arithmetical-logical concepts, by means of the symbols, and the concepts themselves¹²². This accordance is given, on the one hand by the perception of the

¹²¹Recall also that Hilbert objected to Husserl that it was possible to give a complete axiomatization of arithmetic in arithmetical terms. As a matter of fact Hilbert acknowledge that there is a gap between contentual and formal mathematics already in the field of number theory.

¹²²This is the reason why Nelson's critic does not effect Hilbert's proposal. What is here at stake is not a "metaphysics of chalk", as Nelson said in [132], but a more general accordance with our rules of thought and the way we formalize the basic arithmetical and logical concepts. On this epistemological aspects of Hilbert's critical deduction of mathematical knowledge, see [140].

signs and on the other hand by the awareness of the fact that the way we use signs mirrors how we are used to reason about arithmetical and logical concepts. In other words Hilbert's foundational effort is not ontological, but epistemological and transcendental in character. The matching between formalization and intuition is found, in this second period, at the level of contentual mathematics, and then extended to abstract mathematics by a consistency proof, for the logical-arithmetical tools given by the axioms of the proof theory.

To sum up, at the base of two different concepts of axiom there are two different conceptions of our intuitive relation with formal mathematics. Initially axioms define basic concepts of intuitive theories. So the content of symbols depended on the axioms not only since they define formalized concepts, but also because axioms determine the use of symbols, and so their meaning, in mathematical practice. Then axioms have both a definitory and an operational function and their choice depends on a contextual intuition that is used to isolate the basic principles of a theory. It is intuition *that*¹²³ lacking an evident character and, thanks to a consistency proof, it gives content to mathematical knowledge.

Later, at the time of Hilbert's program, evidence and intuition are identified and this coincidence is made apparent in the perception of mathematical symbols. The finitary point of view, together with Hilbert's proof theory, is based on this intuition. Intuition of mathematical symbols determines the *a priori* principles of mathematical reasoning, in its formalized framework, - through self-observation - and hence the choice of the logical-arithmetical axioms. In this second period the manipulation of the signs mirrors our combinatorial abilities and intuition allows to tie together the subjective and the objective sides of mathematical knowledge. The intuition described in this period is then intuition *of* and thanks to a consistency proof mathematical knowledge can be extended from the finitary to the transfinite domain.

1.3 Hilbert's position(s) now

It is now time to leave aside the historical analysis, hoping to have given a sufficiently clear picture of Hilbert's conception of axioms. Our problem, at the beginning of this chapter was: which argument can we give to accept a new axiom in set theory? Of course we did not get even closer to a partial answer, but we tried to set the stage for at least understanding properly the problem, clearing the modern concept of axiom or at least its origin. Nevertheless, did we learn something from what we found in Hilbert's foundational studies? Maybe.

¹²³Hence, not kantian.

The problem is that we do not know, at this stage, if it is possible to export Hilbert's work in a more contemporary setting. In this last section we will try to do exactly this. We will argue that there are more than superficial affinities between Hilbert's work and how the problem of the justification of the axioms is set in the contemporary philosophy of mathematics and among set theorists. Once this link will be argued and secured it will be possible to draw some theoretical conclusion from our historical analysis and propose a partial answer to our original question.

1.3.1 Hilbert's program and Gödel's position

One of the outcome of the previous section is that there are - at least - two different conception of axiom in Hilbert's work. Then we will need to check if one of the two or both can be considered valuable as a methodology in the search for criteria that can justify modern axioms in set theory. We start from the one expounded in the second period. As we hinted before, the discovery of the phenomenon of incompleteness is the main problem for it.

Therefore we start by considering the impact of Gödel's theorems, in the light of the distinction between the first and the second period of Hilbert's enquiry. As we tried to show, what was at stake was not only the conception of axioms, but also the fiddly muddle of relations among axioms, theorem and proofs, that, according to Hilbert, have to be defined within a formal system.

If by "Hilbert's program" we refer to the attempt to give a consistency proof, by means of finitistic methods, of the formal system of infinitary mathematics, then we can say that substantially, accepting Church-Turing's thesis, Gödel's second incompleteness theorem marks the end of this program.

This wasn't Gödel's point of view in 1931, at the time of the discovering of the incompleteness phenomenon. Even if, in [44] and in [45], he holds a substantial identification between Primitive Recursive Arithmetic and Hilbert's finitism, in [48], he says that "Due to the lack of a precise definition of either concrete or abstract evidence there exists today, no rigorous proof for the insufficiency (even for the consistency proof of number theory) of finitary mathematics". Nevertheless at the level of the basic ideas of Hilbert's proof theory, Gödel himself is aware of the fact that the kind of intuition that Hilbert used to give a foundation to mathematical knowledge was not sufficient:

Since finitary mathematics is defined as the mathematics of *concrete intuition*, this seems to imply that *abstract concepts* are needed for the proof of consistency of number theory . . . By abstract concepts, in this context, are meant concepts which do not have as their content

properties or relations of *concrete objects* (such as combinations of symbols), but rather of *thought structures* or *thought contents* (e.g. proofs, meaningful propositions, and so on), where in the proofs of propositions about these mental objects insights are needed which are not derived from a reflection upon the combinatorial (space-time) properties of symbols¹²⁴.

In this quote we can see that, even if Gödel tried to clear the concept of finitism in the sense of what it should be, he thought that the finitary point of view given by Hilbert, by means of the intuitive relationship between the subject and mathematical symbols, was not adequate to give a concrete definition of finitism. In other words even if Gödel did not abandon the idea that it could exist a meaningful and useful sense that could be given to the expression “finitary methods”, he thought that Hilbert’s philosophical explanations were not acceptable.

Then we can say that not only Gödel’s technical results destroyed Hilbert’s program but also the general philosophical background of Hilbert’s intuition was not accepted by the author that mostly shaped the problem of the justification of the axioms in the last century. The importance given to Gödel’s reflection is not a way to appeal to a principle of authority, but the recognition of the importance of Gödel’s setting of the problem of the justification of the axioms in modern times: the so called *Gödel’s program*¹²⁵ in set theory.

This program aims to extend ZFC, the first order formalization of set theory, with new axioms, in order to decide problems proved to be independent from ZFC. The goal of this program is in a sense the same as Hilbert’s program, i.e. to remove any *ignorabimus* from mathematics. In Gödel’s words:

It is well known that in whichever way you make [the concept of demonstrability] precise by means of a formalism, the contemplation of this very formalism gives rise to new axioms which are exactly as evident as those with which you started, and that this process can be iterated into the transfinite. So there cannot exist any formalism which would embrace all these steps; but this does not exclude that all these steps [...] could be described and collected together in

¹²⁴[48]. We don’t try here to clear what material intuition is for Gödel, neither what finitism should be for him. What we are interested in is Gödel’s position towards Hilbert’s philosophical ideas behind the finitary proposal.

¹²⁵As far as I know this terminology dates back to Feferman who explicitly call it in this way in [27], but it is now a standard terminology, not also among philosophers of mathematics - for example [54] - but also among working set theorists; see [190].

some non constructive way¹²⁶.

This conceptual background is the starting point of every axiomatic investigation of set theory, after the discovery of the phenomenon of incompleteness and the discovery of forcing as a general mean to prove independence in set theory. Of course one can choose different strategies as changing the underlying logic of ZFC or refuting the axiomatic setting, but these are totally different approach that deserve a completely different study. We say explicitly here that the first order axiomatic setting will also be the horizon of our inquiry.

1.3.2 Hilbert's position and Gödel's program

We now turn our attention to the first period. In Hilbert's earlier conception we find a notion of formal system sufficiently dynamic so that we can claim that Hilbert's foundational program could partially survive.

In 1900 Hilbert was explicit in considering the axiomatic method as an always evolving process towards a better understanding of the basic concepts of a theory. These concepts are defined by the axioms, but it is possible that our experience¹²⁷ can lead us to widen the definitions and consequently complete the axiomatic system. These ideas partly survive in the second period and, in 1922, Hilbert says:

Thus the concept "provable" is to be understood relative to the underlying axioms-system. This relativism is natural and necessary; it causes no harm, since the axiom system is constantly being extended, and the formal structure, in keeping with our constructive tendency, is always becoming more complete¹²⁸.

Even if the process of extending a formal system, made necessary by Gödel's theorem, never stops, the idea that the axiomatic method is a tool to deepen the understanding of the basic concepts of a theory survived after Gödel's theorems. Indeed the development of always stronger¹²⁹ formal systems, capable of showing the consistency of the weaker ones, is exactly what gave rise to Gödel's

¹²⁶[46], p. 151 in [49].

¹²⁷Here experience is to be intended in the widest sense possible. It is not a sensorial experience - akin to the platonic one - but it is rather the progress of a theory, that by adding new theorems deepens the understanding of the basic concepts of a theory.

¹²⁸[76], p. 1127 in [25].

¹²⁹We refer here to the order given by the consistency strength of different large cardinals hypothesis. These cardinals form a linear order thanks to which we can compare the strength of different theories, as far as the theories are equiconsistent with a sentence asserting the existence of a large cardinal. For a detailed presentation of the subject see [89].

program. Of course there are profound differences between the two authors on how this process of extension should work. Just to hint at a major difference on the role that intuition plays in this context, consider that for Gödel the new axioms are “exactly as evident as those with which you started”, contrary to the separation between intuition and evidence that we found as a distinctive character of the first period of Hilbert’s reflections. Moreover, in the above quotation what becomes more complete is the formal structure, contrary to Gödel’s phenomenological idea that the extension of the axioms of set theory helps in completing our concept of set. Even though, as we saw, there are some similar aspects also in Hilbert’s position.

However, even from different starting points, both Hilbert and Gödel ask for a deepening of the basic concepts of a formal theory. Indeed Gödel, in 1964, thought that

a complete solution of these problems [e.g. the continuum hypothesis] can be obtained only by a more profound analysis (than mathematics is accustomed to give) of the meaning of the terms occurring in them (such as “set”, “one-to-one correspondence”, etc.) and of the axioms underlying their use¹³⁰.

We can see here a clear similarity between the two authors: to Gödel’s “a more profound analysis of the meaning of the terms”, corresponds Hilbert’s “deepening the foundations”, that amounts in an “axiomatic investigation of [...] conceptual content [of the signs]”. As a by-product of this method we obtain new axioms able to complete a theory, with respect to the contextual intuition we developed from it. Of course the kinds of intuitions are different, and to different intuitions corresponds different things that are intuited; moreover the completeness for Hilbert is with respect to an intuitive theory, while for Gödel corresponds to a degree of understanding of the concept. Nevertheless we would like to suggest the presence of a methodological affinity of the two thinkers, strong enough to enable us to try to apply what we learned from Hilbert to the modern problem of the justification of the axioms. Indeed, the only criterion given by Hilbert echoes closely a gödelian one, usually known as the one of *success*.

There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole discipline, and furnishing such powerful methods for solving given problems (and even solving them, as far as possible, in a constructivistic way) that quite irrespective

¹³⁰[47], p. 257 in [49].

of their intrinsic necessity they would have to be assumed at least in the same sense as any established physical theory¹³¹.

In 1925, in [78], Hilbert advances a very similar idea, linking successful and justification.

[I]f justifying a procedure means anything more than proving its consistency, it can only mean determining whether the procedure is successful in fulfilling its purpose. Indeed, success is necessary; here, too, it is the highest tribunal, to which everyone submits¹³².

Interestingly enough, in *Die logischen Grundlagen der Mathematik* Hilbert gives an outline of how the development of mathematics works in the context of an extension of formal system, hinting at a constant dialectics between the demonstrative moment and meta-theoretical analyses that lead to the adoption of new axioms. Indeed, leaving aside the purely formal aspects that characterize proofs at the metamathematical level, we can find in Hilbert's conception of the axiomatic method the roots of the useful relationship, for the mathematical discourse, between theory and practice, that parallels what normally happens in the development of modern set theory.

Thus the development of mathematical science as a whole takes place in two ways that constantly alternate: on the one hand we drive new provable formulae from the axioms by formal inferences; on the other, we adjoin new axioms and prove their consistency by contentual inference¹³³.

This last quotation is just a suggestion, since no clear description of this dialectical process is outlined. What is important to notice, is the presence, even in the mathematical world of Hilbert, of dynamic elements in the development of mathematics.

In what follows we will argue in favor of a more inductive development of a set theoretical foundation of mathematics, that takes into account the impossibility - given by Gödel's results - of a consistency proof and that can rather be assimilated to Russell's view as described in the 1907 lecture *The Regressive Method of Discovering the Premises of Mathematics*.

¹³¹[47], p. 265 in [49]

¹³²[78], p. 370 in [55].

¹³³[77], p. 1138 in [25].

But when we push analysis farther, and get to more ultimate premises, the obviousness becomes less, and the analogy with the procedure of other sciences becomes more visible. The various sciences are distinguished in their subject matter, but as regards method, they seem to differ only in the proportions between the three parts of which every science consists, namely (1) the registration of ‘facts’, which are what I have called empirical premises; (2) the inductive discovery of hypotheses, or logical premises, to fit the facts; (3) the deduction of new propositions from facts and hypotheses¹³⁴.

Even if the two above quotations differ radically in the method of justification of the axioms, they show an interesting similarity in the useful interaction between the theoretical and the practical level.

1.3.3 Idea of completeness and contemporary axiomatics

We are now in the position to draw the moral of Hilbert’s position and see if we really learned something from it. The central concept we found in the first period - that we argued to have similarities with the modern reflections on the problem of the justification of the axioms - is that of completeness. Having explained what Hilbert means by completeness and what he was aiming for in placing it at the center of his axiomatic presentation of Geometry, we would like on the one hand to extract from it some useful ideas that can be used as criteria for justify new axioms in set theory and on the other hand to check if similar ideas has been used somewhere else in mathematics, after Hilbert. We would like to say here that we do not want to explain how the notion of completeness became what we now call semantic completeness, syntactic completeness and categoricity¹³⁵. On the contrary, we would like to see if the idea of a maximal axiom that tries to match intuition and formalization has been used in other contexts.

We now want to list three property that the *Axiom der Vollständigkeit* manifest and that we will later try to apply in the justification of new axioms of set theory. The first aspect we would like to hold from our analysis is the *Sufficiency* of the Axiom of Completeness with respect to the other axioms of Geometry. Notice that this property is relational, both to the a set of independent axioms - that act like necessary conditions - and to a set of propositions that need to be formalized.

¹³⁴[148], p. 282.

¹³⁵See [5] and [6] in this respect

Hilbert's solution is satisfactory as far as the Axiom of Completeness, translated into a modern terminology - using second order logic - implies the categoricity of the model. However, since it is possible to develop arithmetic in the system of the *Grundlagen der Geometrie*, by the first Gödel's incompleteness theorem, this system is deductively incomplete, with respect to first order logic. However, Hilbert did managed to build a complete system of axioms, i.e. capable to prove all relevant theorems of Euclidean geometry and to formalize all methods of proof used in it. Indeed, and this lead us to the second property, a sufficient axiom should have *Completeness* with respect to an intuitive theory; i.e. be able to 1) prove all facts in a given set of sentences and 2) close the relevant domain with all the allowed method of definition of new objects.

In the analysis of the foundations of Geometry, Hilbert faced the problem of finding a link between the subjective perception of mathematical reality and the objective character of mathematical truth. However, this link was not independently justified, because he never even tried to address the problem of explaining the concept of Geometry. Indeed, one of the main outcome of our analysis of the Axiom of Completeness is the fact that Hilbert used it to define what Geometry is. So, the third the aspect we want to stress here - and keep for later use - is the *Prescriptiveness* of this axiom, contrary to the descriptive character that is often linked to a traditional conception of axioms.

This novelty can be seen as a consequence of the development of logic in the late Nineteenth Century. Indeed, there is a substantial link between the problem of matching intuition and formalization and the problem of a mathematical treatment of logic. As a matter of fact whenever there is a need to formalize concepts that have intuitive roots, we have to reflect on whether reasoning on these concepts is really possible; and at the border between subjectivity of judgements and objectivity of truths there is logic. In this respect the Axiom of Completeness is used to delimit the scope of axiomatization and it witnesses an extra-logical relation with the subject matter of Geometry, together with the choice of what counts as a geometrical proposition.

A different treatment deserves the criteria of *Success*, that partly is subsumed in the unifying power of an axiom, once the background intuitive theory and the theorems that need to be proved are identified from the outset. Moreover, Hilbert acknowledges that the axiomatic method is not used to discover new truths. Nevertheless we will discuss the role of success among the criteria for new axioms in set theory later, in connection to the problem of what counts as a natural axiom.

Other examples of completeness

In the end of this chapter we would like to show that there exist analogs of the Axiom of Completeness in other branches of mathematics and therefore Hilbert's work can be seen as an instance of a more general procedure aiming to establish some necessary conditions for the development of a theory, together with some sufficient formal condition capable of unifying and defining an intuitive theory.

Indeed, another example, besides the case of geometry, is the formalization of the concept of computability. In this case the need for a principle capable for completing the theory is really important, since what is formalized is an intuitive concept. In this context, the analogue of the Axiom of Completeness is Church-Turing thesis. It says that the class of functions defined by the λ -calculus - equivalently of general recursive functions and of functions computable by a Turing machine¹³⁶ - is the class of all the functions that are intuitively computable. Then, since all these functions are intuitively computable, Church-Turing thesis is a sufficient condition that characterizes the class of computable functions. There seems to be an important difference between the Axiom of Completeness and Church-Turing thesis, since the former is an axiom, but the latter is a thesis. However the difference is only apparent, because as far as their use in proofs is concerned both serve as a justification of the use of the other axioms. Indeed Hilbert says explicitly that the Axiom of Completeness is not used in his geometrical investigations; exactly as the Church-Turing thesis is not used in proving theorem of recursion theory, but just invoked to justify that all and only those functions are intuitively computable, and so giving a conceptual unity to the intuitive theory. Again we can see that Church-Turing thesis bridges the gap between the formalization of a concept and our intuitive idea.

Another example of this kind is the formalization of the concept of natural number by means of the Peano-Dedekind axioms¹³⁷. In this case the presentation is completed by the axiom of induction as a second order principle: given a non empty set M , an element $0 \in M$ and an injective function $S : M \rightarrow M$

$$\forall P \subseteq M (0 \in P \wedge \forall x (x \in P \rightarrow S(x) \in P) \rightarrow P = M).$$

This axiom says that every subset of M satisfying the axioms and closed under the successor function, must necessarily be the structure of natural numbers. In

¹³⁶Indeed all these classes are provably the same.

¹³⁷Besides the scheme for induction we have:

1. $\forall x (x \neq 0 \rightarrow \exists y (S(y) = x))$,
2. $\neg \exists x (S(x) = 0)$,
3. $\forall x, y (x \neq y \rightarrow S(x) \neq S(y))$.

other words is not possible to extend the system of natural numbers with new objects and to get a new system of things that satisfies the Dedekind-Peano axioms, minus induction.

As in the case of the Axiom of Completeness we are here dealing with a method which, by using second order principles, fixes uniquely a structure, intended to formalize an intuitive concept. As for Geometry, by means of these axioms we give a definition of natural number. It is interesting to note that in both situations the result is achieved through the identification of a property which formalizes the demonstrative power of a concept: continuity in the first case, induction in the second.

Therefore, it is interesting to ask whether this kind of principles can be found also in set theory. As we already argued, we are not asking here for a completeness axiom for the whole of set theory, because of the phenomenon of the open character of the mathematical work, together with the fact that set theory is a foundational theory for mathematics. More intuitively, if we try to formulate an axiom that makes set theory complete with respect to the intuitive idea of set, one collides with some conceptual difficulties. These are due to the fact that the very concept of set is a mental operation of reducing to unity a plurality of things. Therefore the “set of” operation cannot be limited to a fixed domain, without asking if this latter is itself a set. The history of the axiomatization of the concept of set is in fact a continuing attempt to impose the least restrictive limitations, in order to avoid an inconsistent system; as Russell’s paradox showed for Frege’s system.

However, even facing these inherent difficulties, the need for an axiom similar to Hilbert’s Axiom of Completeness was historically felt quite early in the development of set theory.

In 1921 Fraenkel expressed this idea as follows:

Zermelo’s axiom system does not ensure any character of “categorical” uniqueness. For this reason there should be an “Axiom of Narrowness” similar, but opposite, to Hilbert’s Axiom of Completeness, in order to impose the domain to be the smallest possible, compatibly with the other axioms. In this way we can eliminate those classes, existing in Zermelo’s system, that are unnecessary for a mathematical purpose¹³⁸.

¹³⁸[31]. In German: *Das Zermelosche Axiomensystem sichert dem Bereich keinen “kategorische” Eindeutigkeitscharakter. Dazu ist ein weiteres, dem Hilbertschen Vollständigkeitsaxiom umgekehrt analoges “Beschränktheitsaxiom” erforderlich das dem Bereich den kleinsten mit den Axiomen verträglichen Umfang auferlegt. Hierdurch werden verschiedene, für mathematische Zwecke unnötige Klassen von Mengen ansageschieden , die im Zermeloschen System*

Moreover if we try to apply Hilbert's foundational ideas as outlined in these pages - i.e. to apply the axiomatic method in order to find necessary and sufficient conditions - we can say that set theory provide fine tools to analyze the main possibility of proof of a theorem¹³⁹ and a unifying language where it is possible to pose any mathematical problem.

Hence our next step consists in understanding the foundational role of set theory and see if this aspect of the theory influences the criteria for extending its axioms.

Platz haben. My translation. Notice that this interesting quotation goes directly against Maddy's naturalism, because it asks in the context of the mathematical research for a maxim opposed to the criterion of MAXIMIZE, as exposed in [109]. Then Maddy should explain in which sense set-theoretic practice of Woodin's research is to be considered more valuable than Fraenkel's, for what concern the possible extensions of the Zermelo-Fraenkel axiom system.

¹³⁹This is also the aim of what is now called Reverse Mathematics, although its main focus are systems that lie in between RCA_0 and second order arithmetic. For this reason in Reverse Mathematics the axiomatic method is applied to theorems about countable structures. So, even if its analysis is finer, its scope is much smaller than that of set theory. See [119], [154] and [159] for a presentation of aims and methods of Reverse Mathematics.

Chapter 2

Set theoretical foundations of mathematics

Dealing with these foundations has, surprisingly enough, tuned out to be not only a job that had to be undertaken for reasons of intellectual sincerity or philosophical meticulousness but something that was infinitely rewarding, exciting and fruitful.

A. Fraenkel, Y Bar-Hillel and A. Levy *Foundations of set theory*

A wonderful aspect of mathematical work is the possibility to create useful interactions between apparently different areas. This aspect, that we may call the *unity of mathematics*, is a distinctive aspect of modern mathematics. The tools and the ideas that come to light thanks to this global point of view are so powerful that allow to overcome the Aristotelian *caveat* about the different *genus*, for example, between geometry and arithmetic. Moreover, the birth of modern mathematical logic and the need to keep together a very vast and disperse development of mathematics were among the reasons that allowed and pushed toward the foundational programs of the beginning of the last century. Nevertheless history frustrated these foundational efforts. Not only contradiction were discovered, but also a deep and unsolved tension between syntax and semantics: two very new branches of mathematical enquire. We can say that all foundational programs did not succeed in the sense they were conceived.

Nevertheless foundational inquires are still open and there are mathematical problems that have a foundational flavor. This situation calls for an explanation of what a foundation is and how it is possible to propose one, nowadays. We think that among the many reasons that push for a foundation of mathematics, there is a goal that is common to every foundation, that is to shape the mathematical field. By this we mean that any kind of foundation, if it does not defines,

at least distinguishes between mathematical and non-mathematical work and, in some way, characterizes mathematical practice, as being of a certain kind and obeying some specific rules. It is in this sense that we can find concerns for the unity of mathematics also in the foundational context and we believe that this is a common aspect of all different foundations of mathematics.

In this chapter we propose to look at set theory not only as a foundation of mathematics in a traditional sense, but as a foundation for mathematical practice. For this purpose we distinguish between a standard, ontological, set theoretical foundation that aims to find a set-theoretical surrogate to every mathematical object, and a practical one that tries to explain mathematical phenomena, giving necessary and sufficient conditions for the proof of mathematical propositions. We will present some examples of this use of set-theoretical methods, in the context of mainstream mathematics, in terms of independence proofs, equiconsistency results. We will also discuss some recent results that show how it is possible to complete the structures $H(\aleph_1)$ and $H(\aleph_2)$. Moreover, in the central part of this chapter we will claim that a practical foundation of mathematics can be considered relevant not only for the practice of doing mathematics, but also for a philosophical perspective, showing its import in the context of the philosophy of mathematical practice. This latter task will be done considering the explanatory role of the set-theoretical axioms and discussing Kitcher account on the matter of scientific explanation. In the end we will propose a more general distinction between two different kinds of foundation: a practical one and a theoretical one, drawing some examples from the history of the foundations of mathematics.

2.1 Set theoretical foundation as unity

In order to explain this concept of unity we can see how it is realized in the context of the most common foundation of mathematics: set theory. This character of set theory has always been stressed by many people. We offer just one quotation for many, by Penelope Maddy:

For all that, set theoretic foundations still play a strong unifying role: vague structures are made more precise, old theorems are given new proofs and unified with other theorems that previously seemed quite distinct, similar hypotheses are traced at the basis of disparate mathematical fields, existence questions are given explicit meaning, unprovable conjectures can be identified, new hypotheses can settle old open questions, and so on. That set theory plays this role is

central to modern mathematics, that it is able to play this role is perhaps the most remarkable outcome of the search for foundations¹.

However, instead of describing the almost ‘standard’ set-theoretical foundation following which every mathematical entity is intended to be a set, we propose to look at set theory as a means to give a foundation to mathematical practice². Indeed, the universality character of set theoretical language - i.e. the possibility to formalize any piece of mathematics inside set theory and to find a set-theoretic surrogate for any mathematical object - has not *a priori* any ontological meaning. Set theory is not to be intended here as only ZFC, as it is often the case when set theory is called upon arguing for a standard foundation, but as a general method that makes use of set-theoretical principles to analyze mathematical practice. As part of this method we include also reverse mathematics and all the useful set-theoretical assumptions, sometimes called axioms, that extend ZFC³. Clearly the term “theory” here is an abuse of language from a logical point of view, because, neither we think of a consistent set of sentences, nor of an intuitive theory with its intended interpretation. What we have in mind is a general method that is widely and sometimes tacitly used in mathematical practice.

2.1.1 Hilbertian origins

There are two aspects of this set-theoretical perspective that in our perspective model the corresponding foundation of mathematics. They explain in which sense set theory meets the requirements of unity of mathematics. These ideas can be traced back to Hilbert’s foundational works. They follow the evolution of Hilbert’s thought on foundational issues, belonging to two different periods, also chronologically distant. Hence they do not characterize his point of view on this subject.

The first one pertains to Hilbert’s period of foundations of Geometry. In his mind the question of why a theorem is true was equivalent to the problem of

¹[109], pp. 34–35.

²On this ground we will be inspired by Resnik’s idea that mathematics is a science of patterns and that a set theoretical foundation can be seen as a macro-pattern: “There is another phenomenon which has greatly changed mathematics and which could be called a reduction. This is the set theorizing of mathematics. I have in mind the use of the language of set theory as the background language of working mathematics and the attendant objectification (or, in my terms, positionalization) of mathematical structures.” [145], p.540. However, contrary to the realist structural position of [146] and [145] we will try to show how to make sense of the notion of pattern in an epistemological way and not only ontological. Our aim indeed will be to support a set theoretical foundation of mathematics, deprived of its standard realist proposal.

³Among them a special status is hold by large cardinals, but we will discuss it later.

elucidating the main possibility of a proof.

I understand under the axiomatical exploration of a mathematical truth [or theorem] an investigation which does not aim at finding new or more general theorems being connected with this truth, but to determine the position of this theorem within the system of known truths in such a way that it can be clearly said which conditions are *necessary and sufficient*⁴ for giving a foundation of this truth⁵.

Of course, despite Hilbert's ideas, history then showed that metamathematics can give rise to new truly mathematical results and it is a powerful method not only to determine general properties of the axiomatic setting of a formal theory. What is important to stress here is that this attitude is an attempt to give an answer to possible 'why questions' that can rise in the mathematical discourse. Indeed this is exactly what Hilbert was hoping to do in his foundation of geometry. In a letter to Frege, dated December 29th, 1899 (in [36], pp. 38-39) Hilbert wrote: "I wanted to make possible to *understand* and answer such questions as *why* the sum of the angles in a triangle is equal to two right angles and how this fact is connected with the parallel axiom⁶".

The second idea that we would like to recover from Hilbert is the conviction that a good axiomatization of mathematics should be a catalog of the principles that we use in our mathematical practice.

The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds⁷.

In Hilbert's program, this belief was related to the expectation that few arithmetical and logical axioms were able to characterize every piece of mathematics. Since this has been shown to be impossible, we accept this suggestion to be compatible with an open ended list. Indeed this idea of a catalog of principles could *a priori* involves also incompatible principles. We are not looking for a categoricity theorem that permits to define what a theory is about, but a theory that can explain our mathematical work, showing its uniformity of methods and arguments, in order to account for its unity.

We think that these two ideas are also able to account for the explanation of a mathematical fact, outlining the main conditions of its proof and pointing at

⁴My italics.

⁵In [67], page 50.

⁶Italics mine.

⁷[79], p. 475 in [55].

the reasons for accepting his truth⁸. Since we are dealing with a demonstrative context, what is often essential for overcoming the difficulty of an argument is a combinatorial aspect of the proof, that reveals the key ingredient for the solution of a problem. This is the reason why many set-theoretical principles have a combinatorial character, but this does not prevent us from listing them in the catalog, as long as they contribute to account for the unity of mathematics - i.e. they are not *ad hoc* and they have many and different applications. What is relevant in showing that some principles are necessary and/or sufficient is their role in the argumentative structure of a theorem. Sometimes these principles go hand in hand with a more general understanding of a whole field.

The general idea behind this conception of set theory is that it is a method that can be applied to all other branches of mathematics. Indeed this is how it was conceived, at least by Zermelo, in the Thirties.

Our axiom system is non-categorical after all, which, in this case, is not a disadvantage, but an advantage. For the enormous significance and unlimited applicability of set theory rests precisely on this fact⁹.

A good conceptual reason for arguing in favor of set theory as an open-ended foundation lays in the fact that - like mathematics, as we will argue in the next section - the subject matter of set theory is not sufficiently clear to immediately characterize a model and isolate its axioms. This is one of the main concern in contemporary research in set theory, but what is important to stress is the distinction between a foundational role of set theory and the research that tries to single out one true model for ZFC, out of the many we can conceive. This is not an easy task, because if we have a good intuition, for example in the case of the real line, of what are the pathological aspects we would like to avoid, like the Banach-Tarski paradox and the consequent non-measurability of some sets of reals, this is much more difficult as soon as we proceed in the hierarchy of the transfinite. Far from being a weakness, this hazy boundary is what allows set theory to account for the unity of mathematics. However, this aspect of vagueness, common to both set theory and mathematics, has been criticized and has always been subject of discussions, in the foundational context, where, exactly, to draw the dividing line between mathematics and non mathematics. There are people like Feferman and Weaver that would like to put the crossbar much lower than the level of ZFC¹⁰. However if we are trying to explain the unity of mathematics, and therefore we are working at a foundational level, we

⁸Later we will argue more on this point.

⁹In [192], page 427.

¹⁰See, for example, [187] and [28].

cannot drop so easily set theory. Indeed either we discredit modern research in set theory as being mathematics, or we have to propose a sufficiently wide framework, where it is possible to place it. In a slogan: there are more things in mathematical research and mathematical practice than are dreamt of in ZFC. What we propose here is to consider the methods offered by set theory as a framework for mathematics, part of which is of course set theory.

2.1.2 Practical reasons

We now plan to show why set theory can offer a foundation for the practice of doing mathematics. Before we start we need a definition that is fundamental in what follows.

Definition 2.1.1. *We say that a theory T , that extends ZFC, has consistency strength stronger than a theory S if in first order Peano arithmetic it is possible to prove $\text{Con}(T) \rightarrow \text{Con}(S)$, where $\text{Con}(T)$ is the sentence expressing the consistency of T . Moreover, for a sentence A written in the language of set theory, we refer to $\text{Con}(A)$ as an abbreviation for $\text{Con}(\text{ZFC} + A)$.*

There are three reasons that support the idea of a set-theoretical foundation of mathematical practice.

1. *Independence proofs.* This is the main subject of modern research in set theory. Since the invention of forcing¹¹, in the Sixties, many problems were shown to be independent from ZFC, like for example the Continuum Hypothesis (CH) and Souslin's Hypothesis. This kind of proofs is used, as Hilbert did, to prove that a set of axioms is not sufficient for a mathematical result.
2. *Combinatorial principles.* The discovery of the independence of a proposition does not conclude its mathematical analysis. Indeed the examination of an independent problem often brings together the identification of a technical *impasse* and the corresponding combinatorial principles that are sufficient for its solution. For a safe use of these principles, the method of forcing is used to show that they are consistent relative to some theory like, for example ZFC. But sometimes ZFC is not sufficient for this task. It is here that large cardinals come into play.
3. *Large cardinals.* These are hypotheses on the existence of cardinals large enough¹² to prove $\text{Con}(\text{ZFC})$. They are used to determine the power

¹¹See [102] for a very good introduction to this subject.

¹²If κ is large enough, so that V_κ is a model of ZFC, then we say that κ is a large cardinal. However this is not a definition of what is a large cardinal.

of sentences stronger than ZFC, in terms of their consistency strength. Indeed many natural sentences stronger than ZFC can be proved to be equiconsistent - in the context of ZFC - with the existence of suitable large cardinals. Then large cardinals can be viewed, modulo equiconsistency, as necessary and sufficient conditions for the proof of sentences stronger than ZFC.

There is an important reason for using large cardinals as the backbone for the analysis of the propositions that transcend the deductive power of ZFC.

Empirical fact: the order induced by the consistency strength of large cardinal hypothesis is, except few cases, linear and well founded.

The use of large cardinals in set theory is twofold: on the one hand they serve to compare different principles, using equiconsistency results and the linear order given by their consistency strength; on the other hand they supply the means to give relative consistency proofs, being the key ingredient of theorems of the form: given an independent statement A , written in the language of set theory, if the hypothesis I stating the existence of a large cardinal κ holds, then there is a model where A holds; in other words I implies $Con(A)$.

There exists an epistemological tension between logical deduction and consistency strength and this aspect is responsible for the richness of the analysis that set theory can offer of necessary and sufficient conditions. Indeed the epistemological value of the search for necessary and sufficient conditions for the proof of a theorem consists in the discovery of its place in the logical structure of a theory. This process can work in two directions: starting from an axiomatic system and asking which of its axioms are needed for the proof of a theorem, or starting with a proposition and looking for the axioms that are needed for its (non-trivial) proof, without specifying the axiomatic context. In the first case this analysis is informative on the content of a theorem, like for example Hilbert's work on Desargue's theorem - where the aim is to clear its spatial content. But in the second case, when the goal is a context-free analysis that looks for the principles that are needed for the proof of a proposition - for example a proposition independent from ZFC - the discovery of necessary and sufficient conditions consists just in finding logically equivalent formulation of the proposition. In this latter case the progress in our knowledge maybe given by a combinatorial character of an equivalent formulation, or its relevance in a different field, but it is not informative for what concern the possibility of its proof, nor for its content - i.e, we cannot give an answer to the question

“why this proposition is a theorem of set theory?”. On the contrary, a result of equiconsistency is well more informative on the epistemological status of a proposition. Indeed, such a proof outlines the fact that we have to believe not only in the truth of a sentence, but also in the existence of a particular class of models of ZFC: the ones whose existence is guaranteed by the equiconsistency proof. Moreover, logical equivalence and equiconsistency cannot be assimilated, without collapsing truth and existence. While the former is a syntactical notion, the latter is semantical and expresses the fact that we need to believe in something, possibly, stronger than ZFC in order to believe the truth of a particular sentence. This is the reason why the use of large cardinals in the consistency proofs fills the gap between logical deduction and consistency strength, because not only we can have a theorem of the form $I \rightarrow A$, but also $I \rightarrow Con(A)$, i.e. we can show that the sentence expressing the existence of a large cardinal, logically implies not only another sentence, but also the fact that there is a model where this is true.

Hence, large cardinals provide a precise answer to “why” questions. In this sense large cardinals can be seen as more fundamental principles that give more and more powerful means, not only to prove new propositions, but also to analyze our beliefs in their truth¹³. After the discovery of the difference between, truth, provability and existence, we have to accept the original Sin of Gödel’s theorem, but what large cardinals - together with the method of forcing - offer is a way to analyze and stratify the degree of incompleteness that we find in our mathematical practice.

2.1.3 Example of sufficient conditions

We would like here to present some examples that show how set theory is used to analyze mathematical problems. It is important here to stress not only the fine and deep explanation that is given by a set theoretical investigation, but also the fact that the problems discussed come from some of the characteristic fields of classical mathematics: group theory and functional analysis. This aspect is important since it acknowledges the importance of set theoretical method not only in logical or pathological context, but in classical domains of mathematical practice.

¹³If we accept this point of view we also have to accept the consequence that the more fundamental principle is a contradiction. As a matter of fact many large cardinals hypothesis can be seen as stating the existence of a non trivial elementary embedding of two universes of set theory; the more similar are these classes, the higher is the consistency strength of the corresponding large cardinal. Pushing this process at the limit we get a statement that postulate the existence of a non trivial elementary embedding of the universal class V in itself. This statement has been shown to be inconsistent, with ZFC, by Kenneth Kunen.

We shall describe the solutions given to Whitehead's problem and a recent result by Farah on operators algebras of an Hilbert space.

Definition 2.1.2. (*Whitehead's problem (WP)*) *Is every Whitehead group (i.e. an abelian group A such that, whenever B is an abelian group and $f : B \rightarrow A$ is a surjective group homomorphism, whose kernel is isomorphic to the group of integers \mathbb{Z} , then there exists a group homomorphism $g : A \rightarrow B$ with $fg = id_A$) a free group (i.e. a group A that has a subset X , called the set of generators, such that every element of A can be written uniquely¹⁴ as a finite combination of elements in X and their inverses)?*

In the Seventies Shelah proved the following theorems.

Theorem 2.1.3. (Shelah [151]) *If $V=L$, then the answer to WP is yes.*

Theorem 2.1.4. (Shelah [150]) *If Martin's Axiom (MA)¹⁵ and the negation of the Continuum Hypothesis ($\neg CH$) both hold, then the answer to WP is no.*

We then have another proof of the fact that $V=L$ and $\neg CH$ are incompatible. Moreover, since $Con(ZFC + V=L) \iff Con(ZFC) \iff Con(ZFC + MA + \neg CH)$ we have sufficient conditions for both answers to WP, without exceeding the consistency strength of ZFC; that is, without an overshooting that would confuse the problem.

Another example is the following result in the context of functional analysis.

Definition 2.1.5. *The Calkin algebra $\mathcal{C}(H)$ is the quotient of $B(H)$, the ring of bounded linear operators on a separable infinite-dimensional Hilbert space H , by the ideal $K(H)$ of compact operators.*

It is natural question to ask if every automorphism is inner; i.e. it is induced by the operation of conjugation. The answer to this question is again sensible to the background set theoretical hypothesis.

Theorem 2.1.6. (Philips and Weaver [142]) *If CH holds there is an automorphism of $\mathcal{C}(H)$ that is not inner.*

Theorem 2.1.7. (Farah [26]) *If the Open Coloring Axiom (OCA)¹⁶ holds all automorphism of $\mathcal{C}(H)$ are inner.*

¹⁴Modulo equivalence of the form $ab = axx^{-1}b$.

¹⁵We will not give the definition of MA here. What is important to know is that it is one of the weakest Forcing Axioms. We refer, for the interested reader, to [88].

¹⁶As in the case of MA, we refer to [88] for the definition of OCA.

It is interesting to note that, in this case, the first version of Farah's theorem used the Proper Forcing Axiom (PFA), whose consistency strength is much higher than $Con(ZFC)$. Then, the analysis of why PFA was used in the proof led to the discovery that just OCA, that is a combinatorial consequence of PFA, was needed. Then since $Con(ZFC + OCA) \iff Con(ZFC) \iff Con(ZFC + CH)$ we have found again sufficient conditions for the solution of a natural mathematical problem; the best possible solution with respect to consistency strength.

For the last example consider Fubini-Tonelli's theorem, that is provable in ZFC.

Theorem 2.1.8. (*Fubini-Tonelli*) *Given $f : [0, 1]^2 \rightarrow [0, 1]$, if the function f is Lebesgue measurable then the integral not only exists and*

$$\int_1^0 \int_1^0 f(x, y) dx dy = \int f(x, y) dm_2 = \int_1^0 \int_1^0 f(x, y) dy dx,$$

where m_2 stands for the Lebesgue measure on \mathbb{R}^2 .

Then if we ask whether the theorem still holds, dropping the request on the measurability of the function, we obtain a problem that is independent from ZFC.

Theorem 2.1.9. (*Sierpinski [157]*) *If CH holds then there is a function $f : [0, 1]^2 \rightarrow [0, 1]$ for which the iterated integrals $\int_1^0 \int_1^0 f(x, y) dx dy$ and $\int_1^0 \int_1^0 f(x, y) dy dx$ exist but are not equal.*

Theorem 2.1.10. (*Laczkovich, Friedman and Freiling [103], [38], [37]*) *It is consistent with ZFC that for every function $f : [0, 1]^2 \rightarrow [0, 1]$ the existence of the iterated integrals $\int_1^0 \int_1^0 f(x, y) dx dy$ and $\int_1^0 \int_1^0 f(x, y) dy dx$ implies that they are equal.*

In this case our solution seems to be different from the one to Whitehead Problem, because on one side we know that if the generalization of the Fubini-Tonelli's theorem holds than CH must fail, while on the other hand we know nothing about a model where the Fubini-Tonelli's theorem does not hold, because $Con(CH) = Con(ZFC)$ and CH is independent from ZFC; we just know that they exist. We can say that one theorem is more informative than the other. This example explains why we look for theorems of the form: if some principle P holds, then some sentence S holds. Indeed, contrary to what happens in an independence proof where no principle is shown to be sufficient for the proof of S , we get information of the context in which S can be true. It is not just a proof of the existence of a model, but we have more information about this model.

2.1.4 Examples of equivalence and equiconsistent results

Of course there are also examples of necessary and sufficient conditions for sufficiently natural mathematical problems. They indeed show equivalences between different principles that can be epistemically informative for their combinatorial content, or just useful for finding new and unexpected link between different areas of mathematics.

Theorem 2.1.11. (*Freiling [37]*) *The following are equivalent, over ZFC:*

1. *Continuum Hypothesis: $2^{\aleph_0} = \aleph_1$,*
2. *Axiom of Symmetry: for any function f that associates countable sets of real numbers to real numbers, i.e, $f : \mathbb{R} \rightarrow [\mathbb{R}]^{\aleph_0}$, there are $x_0, x_1 \in \mathbb{R}$ such that $x_0 \notin f(x_1)$ and $x_1 \notin f(x_0)$.*

Nevertheless the strength of the set theoretic method can be mostly appreciated in combination with large cardinals and so when necessary and sufficient conditions are such, up to equiconsistency. The best example is Solovay's model for the following very natural property for sets of reals, that started the study of descriptive set theory.

Definition 2.1.12. (*LM, BP and PSP*) *Given $X \subseteq \mathbb{R}$, we say that X is Lebesgue measurable (LM) if it belongs to the σ -algebra generated by the Lebesgue measure on \mathbb{R} . We say that X had the Property of Baire (BP), if there is an open set U such that $U \Delta X$ (the symmetric difference) is a meager set (i.e. small). We say that X has the Property of the perfect set (PSP), if it is either countable or has a nonempty perfect subset: a closed set with no isolated point.*

Theorem 2.1.13. (*Solovay [160]*) *The following are equivalent, over ZFC:*

- *Con(ZF+ all sets of reals are LM and have BP and PS),*
- *Con(There exists an inaccessible¹⁷ cardinal κ).*

The epistemological meaning of this theorem is that it explains what we need to believe in term of consistency to accept that all subset of the reals behave very nicely with respect to some natural properties.

¹⁷This is the weakest notion of large cardinal. We say that κ is inaccessible if it is regular and such that for every $\lambda < \kappa$, we have $2^\lambda < \kappa$.

2.1.5 Necessary and sufficient conditions

We would like here to present a new point of view on the application of the forcing method to the general phenomenon of independence¹⁸. They are part of a more general program that helps in making more precise the methodology suggested in Gödel's program and that is now called *Woodin's program*¹⁹. This program aims at finding a satisfactory description of the universe of set theory step by step; that is, giving a sufficiently complete description of initial segments of the class $V = \{x : x = x\}$.

We need a definition in order to make precise the "step by step" methodology of this program.

Definition 2.1.14. (*A cumulative hierarchy*) *We can build V stage by stage in the following way: for every $\lambda \in \text{Card}$ the structure $H(\lambda)$ consists of the sets of cardinality hereditarily less than λ .*

$$H(\lambda) = \{x : |x| < \lambda \text{ and } \forall y (y \in \text{tc}(x) \Rightarrow |y| < \lambda)\}.$$

Then Woodin's way to phrase his program is the following.

One attempts to understand in turn the structures $H(\aleph_0)$, $H(\aleph_1)$ and then $H(\aleph_2)$. A little more precisely, one seeks to find the relevant axioms for these structures. Since the Continuum Hypothesis concerns the structure of $H(\aleph_2)$ ²⁰, any reasonably complete collection of axioms for $H(\aleph_2)$ will resolve the Continuum Hypothesis²¹.

Notice that one of the main motivation is the solution of CH. Now we need a definition in order to make precise the sense in which the program attempts to complete the initial segment of the universe of set theory, ruling out the trivial incompleteness phenomena given by Gödel's sentences and the sentences expressing the consistency of a theory.

Definition 2.1.15. *ψ is called a solution of a structure M , that models enough of ZFC, iff for every sentence $\phi \in \text{Th}(M)$,*

$$ZFC + \psi \vdash \ulcorner M \models \phi \urcorner \text{ or } ZFC + \psi \vdash \ulcorner M \models \neg\phi \urcorner.$$

¹⁸Some of the results we quote are old, but what is new is their presentation, the context in which they are placed and consequently the meaning they assume in this new context; see [184].

¹⁹see [190] for a presentation of this program.

²⁰Indeed $\mathbb{R} \subseteq H(\aleph_2)$, since $\mathcal{P}(\omega) \subseteq H(\aleph_1)$ and so every subset of \mathbb{R} belongs to $H(\aleph_2)$.

²¹[190], p. 569.

We now presents some initial result of Woodin's program and some other by Viale, that show how the forcing axioms fit in this program. The importance of these results are to be found in the possibility of using forcing not only to give sufficient conditions, but also necessary. Indeed the slogan that motivates them is the following.

Key idea: The method of forcing is a tool that allow to prove theorems over certain natural²² theories T which extend ZFC.

The first result of this kind is a reformulation of Cohen's forcing theorem, that shows how - the Σ_1 -theory with real parameters of - any transitive model of ZFC overlaps with the Σ_1 theory of $H(\aleph_1)$.

Theorem 2.1.16. (Cohen, Levy, Schoenfield) [21] Assume T extends ZFC. Then for every Σ_0 -formula $\varphi(x, p)$ and every parameter p such that $T \vdash p \subset \omega$ the following are equivalent:

- $T \vdash \ulcorner H(\aleph_1) \models \exists x \varphi(x, p) \urcorner$
- $T \vdash$ There is a partial order P such that $\Vdash_P \exists x \varphi(x, p)$.

Assuming large cardinals the above theorem can be extended to all formulas, with parameters in $H(\aleph_1)$. The next theorem says that it is possible to find a solution for the theory of $L(\mathbb{R})$ (i.e. the class of all set that are constructible with real parameters). Observe that $H(\aleph_1) \subseteq L(\mathbb{R})$ and $H(\aleph_1)^{L(\mathbb{R})} = H(\aleph_1)$. Thus the following result is really an extension of the previous one.

Theorem 2.1.17. (Woodin [104]) Assume T extends ZFC + There are class many Woodin cardinals²³. Then for every formula $\varphi(p)$ and every parameter p such that $T \vdash p \subseteq \omega$ the following are equivalent:

- $T \vdash \ulcorner L(\mathbb{R}) \models \varphi(p) \urcorner$
- $T \vdash$ There is a partial order P such that $\Vdash_P \varphi^{L(\mathbb{R})}(p)$.

So, modulo the method of forcing, large cardinals are a solution for the structure $H(\aleph_1)$; i.e. they decide the theory of $H(\aleph_1)$ with parameters in $H(\aleph_1)$. Indeed the fact of interpreting every result that we present in this section as a

²²Of course there is a an important philosophical problem behind the concept of natural, but we will come back on this later. Luckily the main concept of naturalness is sufficiently natural to be easily understood, but nevertheless it deserves a philosophical analysis.

²³This notion is agin too technical to be defined here; see [88]. What is important to know is just that Woodin cardinals are large cardinals.

good solution for the corresponding theory depends heavily on the assumption that the forcing is the only effective way to obtain independence result for set theory. However this is a theoretical innocuous assumption, that goes hand in hand with the general pragmatism of the methodology that is used in every mathematical research.

Another point that is important to rise here is that Woodin's solution for $H(\aleph_1)$ not only decides all the relevant statement - i.e. the statement true in $H(\aleph_1)$ with parameters in the structure - but it is also complete with respect to the means given by the method of forcing, once we relativize them to $H(\aleph_1)$. As a matter of fact, if we reason from the point of view of this structure it is true that for every countable model M of ZFC and every forcing notion \mathbb{P} if we let $\mathcal{D}_M^{\mathbb{P}}$ the family of all dense subset of \mathbb{P} belonging to M , we have, by Choen's theorem, that there exists a $\mathcal{D}_M^{\mathbb{P}}$ -generic filter; i.e. a filter $G \subseteq \mathbb{P}$ that has non-empty intersection with every $D \in \mathcal{D}_M^{\mathbb{P}}$. This is true because $M \in H(\aleph_1)$ implies that M is countable. Hence it should be noted that this solution is complete also with the respect to the method of 'proof' we have at disposal in this setting; i.e. the structure is 'closed' under forcing.

The results presented so far are about the structure $H(\aleph_1)$ But what happen if we try to go one step farther and look for a solution of $H(\aleph_2)$?

The next step is to find an axiom that can be a solution for the theory of $H(\aleph_2)$ with parameters in $H(\aleph_2)$ ²⁴. Of course we would like to have an analog solution, that is compatible with the one for $H(\aleph_1)$ and able to extend it. The first attempt would be to look for even stronger large cardinal axiom, since this method has been already so successful. The first result in this direction is by Woodin; later refined by Koellner.

Theorem 2.1.18. *(Koellner, Woodin [96]) Given a model M of ZFC, we have that CH is a solution for M , with respect to the class of all Σ_1^2 sentence φ (i.e. an existential statement of third order arithmetic the same complexity of CH ²⁵) written in the language of set theory. Moreover, if a Σ_1^2 statement ψ is another such a solution for M , then $ZFC + \psi \vdash \ulcorner M \models CH \urcorner$. Under the assumption of the existence of a class of measurable Woodin cardinals, this fact cannot be changed under forcing²⁶.*

We are in the same situation we had with Cohen's absoluteness Lemma: a

²⁴There are also attempts to find solution of $H(\aleph_2)$ with parameters in $H(\aleph_1)$; like Woodin's (*). However, unless we consider it only as a partial result, we do not think that this approach is in the spirit of Woodin's program.

²⁵Indeed CH asserts that there is a well-ordering of the reals, every initial segment of which is countable.

²⁶See [96].

partial result in the good direction, that need to be extended. Unfortunately, this cannot be achieved because of the following result by Aspero, Larson and Moore.

Theorem 2.1.19. *(Aspero, Larson, Moore [4]) There exist sentences ψ_1 and ψ_2 which are Π_2 over the structure $(H(\aleph_2), \in, \omega_1)$ such that*

- ψ_1 can be forced by a proper forcing not adding ω -sequences of ordinals (i.e. ψ_1 is consistent with CH);
- if there exists a strongly inaccessible limit of measurable cardinals, then ψ_2 can be forced by a proper forcing which does not add ω -sequences of ordinals (i.e. ψ_2 is consistent with CH);
- the conjunction of ψ_1 and ψ_2 implies that $2^{\aleph_0} \neq \aleph_1$.

This theorem says that there are two Π_2 statements (over the structure $H(\aleph_2)$) that are mutually compatible with CH, but whose conjunction is incompatible with CH. Hence, any attempt to save CH together with large cardinals and find a good solution of $H(\aleph_2)$ is doomed to fail, if it aims to include all Π_2 -statements provably consistent by means of forcing. Worse than that it has been showed by Koellner and Woodin in [96] that if it is possible to give a solution for the class of the Σ_3^2 -sentences that implies CH, then it is possible to find another solution for the class of the Σ_3^2 -sentences that implies the negation of CH.

Thus on the one hand we have to direct our attention towards principles that negate CH, and on the other hand we have to go beyond large cardinals and look for different principles that can settle all possible problem that can be phrased in the structure $(H(\aleph_2), \in, \omega_1)$.

Another problem, less technical but more conceptual, is related to the inexpressibility of the universe of set theory. It is well known that it is not possible to describe the universe of set theory with a single first order formula - neither of second order, by Zermelo quasi-categoricity theorem. Thus if we are looking for a completion of the universe of set theory step by step, why, if we managed to find a solution for an initial segment, we should expect that it is possible to extend this same method to other larger segment, in order to exhaust, sooner or later, all the problems that we are facing? Wouldn't be this a way to describe in a unique way the endless and indefinite universe of set theory? Indeed this idea of extending, from one step to another, the results we have already obtained seems to be at odd with the idea of the undefinability of the universe and the consequent possibility to reflect general property to initial part of its

hierarchy²⁷.

Then we want to find axioms that:

1. decide the largest possible fragment of $H(\aleph_2)$,
2. negate CH²⁸ (because of Theorem 2.1.19),
3. are not large cardinal axioms²⁹,
4. extend Woodin's result on $H(\aleph_1)$ (thus they should be at least compatible with the existence of class many Woodin cardinals),
5. extend the Cohen-Levy-Schoenfield result: $(H(\aleph_2), \in, \omega_1) \prec_1 (V^{\mathbb{P}}, \in, \omega_1)$, for \mathbb{P} a forcing notion,
6. saturate the structure $H(\aleph_2)$, with respect to the means given by the method of forcing; i.e. for every $M \in H(\aleph_2)$ and every forcing notion \mathbb{P} , letting $\mathcal{D}_M^{\mathbb{P}}$ as before, there exist a $\mathcal{D}_M^{\mathbb{P}}$ -generic filter $G \subseteq \mathbb{P}$.

This would be an optimal solution, but we still have to work a bit, in order to achieve it. As a matter of fact, allowing parameters in $H(\aleph_2)$ is a strong requirement and indeed incompatible with the possibility of considering any forcing extension whatsoever. As an example consider the statement that there is a bijection between ω and ω_1 . This statement is of course wrong, but it is possible to force it over $H(\aleph_2)$, allowing ω_1 as a parameter.

Luckily enough we have a good candidate for a slightly weaker versions of all these requirements: Forcing Axioms³⁰. Indeed it was noticed empirically that all the proposition about the structure $H(\aleph_2)$ that were proved to be independent from ZFC, either could be proved thanks to Forcing Axioms, or a counterexample to the proposition could be found in ZFC. Moreover many of them - MM, PFA and FA(ccc* σ -closed) - imply the negation of CH, the strongest among them are consistent relative to a supercompact cardinal - a cardinal much larger than a Woodin cardinal - and they have, as consequences, some restricted form of

²⁷We will come back on this point later, when discussing the possible extension of the Forcing Axioms.

²⁸Notice that this could be seen as an argument against CH, even if this would be subject to the - never ending - success of Woodin's program.

²⁹Notice that this is already required together with asking that they imply \neg CH, because large cardinals do not decide CH.

³⁰We will not define these axioms here, together with many other technical notions, because we will do it later in the second part of this work. What matters now is the general form of the theorem that mirrors the analogues that was presented before. We advice the interested reader to come back on this part later, for a better understating.

the Cohen-Levy-Schoenfield theorem - see [8] in this respect. In particular the axiom MM - one of the strongest among the Forcing Axioms - makes also more understandable the last requirement we proposed. Indeed MM implies that the class of the forcing notions \mathbb{P} , for which every family \mathcal{D} of dense $D \subseteq \mathbb{P}$, with $|\mathcal{D}| \leq \aleph_1$, has a generic filter G intersecting every $D \in \mathcal{D}$, coincides with the class of stationary set preserving forcing.

In conclusion, since we need to limit the class of partial orders we can use, because of possibility to force that ω_1 is countable with a forcing that destroys the stationarity of a subset of ω_1 , let us focus on the stationary preserving forcing³¹ and the Forcing Axioms, in order to decide all the problems that can be stated in $H(\aleph_2)$.

Theorem 2.1.20. *(Viale [184]) Assume T extends $ZFC + MM^{+++}$. There are class many Woodin cardinals. Then for every Π_2 -formula $\varphi(x)$ in the free variable x and every parameter p such that $T \vdash p \in H(\aleph_2)$ the following are equivalent:*

- $T \vdash \ulcorner H(\aleph_2) \models \varphi(p) \urcorner$
- $T \vdash$ *There is a stationary set preserving partial order P such that $\Vdash_P \varphi^{H(\aleph_2)}(p)$ and P preserves BMM.*

Moreover it is possible to extend this result to a full solution of $H(\aleph_2)$.

Theorem 2.1.21. *(Viale [182]) Assume T extends $ZFC + MM^{+++32}$. There are class many super huge cardinals. Then for every formula $\varphi(x)$ in the free variable x and every parameter p such that $T \vdash p \in H(\aleph_2)$ the following are equivalent:*

- $T \vdash \ulcorner H(\aleph_2) \models \varphi(p) \urcorner$
- $T \vdash$ *There is a stationary set preserving partial order P such that $\Vdash_P \varphi^{H(\aleph_2)}(p)$ and P preserves MM^{+++} .*

Hence, thanks to the Forcing Axioms, we have a good description of the structure $H(\aleph_2)$ and, in keeping with the general idea that lies behind and foundation of mathematics - as we saw in the example of the Whitehead Problem and of the Calkin Algebra - they are capable of unifying different branches of mathematics. Indeed both MA and OCA are consequences of MM^{+++} .

It is now time to come back to the question that motivated our investigations on the foundational role of set theory: does it concurs in shaping the criteria

³¹See [184] for a detailed argumentation of why this class is the best one at our disposal.

³²See [182] for the definition of this principle.

for accepting new axioms for it? The outcome of our survey is that, granting that unity is the aim of every foundation of mathematics, *Unification* is indeed something we are willing to ask for our axiom candidates. This criterion has been proposed, yet, by Penelope Maddy in his book *Naturalism in mathematics*, but only at an intuitive. What we plan to do now is to propose a careful philosophical reflection about this notion.

2.2 Axioms as explanations

Later on we will come back to sum up all the criteria we collected so far and, as the reader may suspect, we will try to test them in the case of the Forcing Axioms. But now we want to analyze more philosophically the claim that set theoretical axioms should be able to unify mathematics, as far as set theory is viewed as a foundational theory. In doing so we will also elaborate on our claim that set theory should be considered as a foundation for mathematical practice. Indeed, so far we showed its relevance and usefulness for what concern the practice of doing mathematics, but our claim is stronger and refers to the philosophy of mathematical practice: a recent tradition in the philosophy of mathematics, as a quick look at contemporary bibliography clearly shows³³. Our main thesis of this section is that many of the principles that are used in contemporary set theory, many of which are called axioms, manifest specific characteristics that can be assimilated to, at least, one important account of mathematical explanation - one of the more studied and developed area of the philosophy of mathematical practice. Hence our derived claim is that some set theoretical axioms can be seen as explanation of the mathematical phenomena.

The aim of our analysis will be to link unification and explanation in a foundational context. Many authors have proposed a philosophical inquire on the notion of explanation in science in term of unification, but few of them have proposed it as a way to understand the notion of mathematical explanation i.e. the role of mathematics in the scientific explanation. Among them Kitcher's account, presented in term of the unification power of scientific theories, can be compared with the unifying features that some axioms in set theory have. We will not present here his position in details, but we refer to the primary bibliography ([92], [94] and [93]) and to Molinini's Phd thesis ([126]), for a good presentation of the subject, able to clear many of the obscure passages that can be found in Kitcher's work.

In trying to find some explanatory aspects in the concept of axioms in set theory, we connect our arguments with a long tradition in the philosophy of

³³See among the others [116].

mathematics and in the foundational studies, that has its roots in the empiricist positions of John Stuart Mill.

From these considerations it would appear that Deductive or Demonstrative Sciences are all, without exception, Inductive Sciences; that their evidence is that of experience; but that they are also, in virtue of the peculiar character of one indispensable portion of the general formulae according to which their inductions are made, Hypothetical Sciences³⁴.

Indeed Mill takes a step forward in the direction of a connection between inductivism and explanation.

A hypothesis is any supposition which we make (either without actual evidence, or on evidence avowedly insufficient) in order to endeavor to deduce from it conclusions in accordance with facts which are known to be real; under the idea that if the conclusions to which the hypothesis leads are known truths, the hypothesis itself either must be, or at least is likely to be, true. If the hypothesis relates to the cause or mode of production of a phenomenon, it will serve, if admitted, to explain the facts as are found capable of being deduced from it. And this explanation is the purpose of many, if not most, hypotheses³⁵.

To summarize Mill's position, inductive reasoning brings justification for the scientific hypotheses - both in mathematics and empirical science - and so it allows to explain the phenomena, in terms of the hypotheses that make possible their deduction.

The form of this argument, in the context of mathematical axioms, is what Paolo Mancosu has called the h-inductivism in [115]³⁶: the position that sees in the success of an hypothesis, and its ability to give a systematization of a discipline, the main justification for its acceptance. On the same par, we can find also Russell and Gödel. What is important to stress now, is the relevance of this tradition to the subject of explanation.

The case of Russell is exemplary in showing the dialectic correspondence between the matter of explanation and that of justification. Indeed in the 1907 lecture *The Regressive Method of Discovering the Premises of Mathematics* Russell

³⁴Mill

³⁵Mill

³⁶For what concern the presentation of this historical overview we follow [115].

gave a clear description of how the search for axioms proceeds in the foundation of mathematics.

In mathematics, except in the earliest parts, the propositions from which a given proposition is deduced generally *give the reason why* we believe the given proposition. But in dealing with the principles of mathematics, this relation is reversed. Our propositions are too simple to be easy, and thus their consequences are generally easier than they are. Hence we tend to believe the premises because we can see that their consequences are true, instead of believing the consequences because we know the premises to be true. But inferring the premises from consequences is the essence of induction; thus the method of investigating the principles of mathematics is really an inductive method, and is substantially the same as the method of discovering general laws in any other science³⁷.

On the other hand Gödel has never said explicitly, in print, that axioms can be seen as explanation of their conclusion, even if Charles Parsons, in the introduction of *Russell's mathematical logic* contained in Gödel's collected works acknowledges that this interpretation is "difficult to refute"³⁸. However, in [121] is reported an interesting opinion of Gödel.

The limited effect of the failure of Hilbert's program upon the dependability of the impressive cluster of mathematical theories which he tried to place on a common 'foundation' can be clarified by reference to certain relevant views of Gödel which he informally conveyed to me, some years ago, during a discussion we had at Princeton, N.J. According to Gödel, an axiomatization of classical mathematics on a logical basis or in terms of set theory is not literally a foundation of the relevant mathematics, i.e., a procedure aiming at establishing the truth of the relevant mathematical statements and at clarifying the meaning of the mathematical concepts involved in these theories. In Gödel's view, the role of these alleged 'foundations' is rather comparable to the function discharged, in physical theory, by explanatory hypotheses. Thus, in the physical theory of electromagnetic phenomena, we can explain why the sky looks blue to us under normal circumstances, and we are even able to produce the same phenomenon in the laboratory. Both the explanation of the

³⁷[148], pp. 273–274. Emphasis mine.

³⁸See [20] for a critical exposition of this Gödelian thesis.

physical phenomenon under consideration and its production under laboratory conditions are due to the logical fact that the statements describing the blue of the sky or that of an artificially produced area in the laboratory are theorems provable within an axiomatic system the postulates of which are concerned with hypothetical laws governing electro-magnetic phenomena, the composition of the atmosphere, etc. It would not occur to a physicist that these electro-magnetic assumptions which enjoy the role of postulates in an axiomatized, or axiomatizable physical theory, are more dependably known to be true than the pre-scientific phenomena (like the blue of the sky) which are being explained by being shown to be provable theorems in the aforementioned physical theory. Thus, the actual function of postulates or axioms occurring in a physical theory is to explain the phenomena described by the theorems of this system rather than to provide a genuine ‘foundation’ for such theorems. Professor Gödel suggests that so-called logical or set-theoretical ‘foundations’ for number-theory, or any other well established mathematical theory, is explanatory, rather than really foundational, exactly as in physics³⁹.

This quotation gives strength to our theses in three respects: (1) Hilbert’s axiomatic setting for the problem of the foundation of mathematics is not jeopardized by Gödel’s theorems, (2) axioms can be seen as explanation, (3) set theory is a good framework for an explanatory foundation of mathematics. Gödel is motivated by strong philosophical positions that push towards a strong analogy between mathematics and natural sciences, that, as we will show in analyzing Kitcher’s position, suffers of many conceptual problems.

However, we want to be clear that we do not endorse the thesis that axioms can be seen as explanation to argue in favor of a realist conception of mathematical object. As a matter of fact we argue for a foundation of mathematics free from any ontological commitment. Instead, in what follows we will try to give arguments in favor of the thesis that, in the context of the axiomatic setting, explanation and justification are two sides of the same coin: unification. Once this point is achieved, then, there will be no reasons for justifying the axioms in terms of an existing mathematical reality.

2.2.1 Applying Kitcher’s account?

What interests us here is the role of mathematical explanation, if there is any, inside mathematics. Indeed mathematical explanation can mean both the use of

³⁹[121], pp. 86-87.

mathematics in explaining physical phenomena and the use of explanatory considerations in the context of pure mathematics. From now on by mathematical explanation we will mean the latter: mathematical explanation of mathematical phenomena.

For the sake of precision, there is no precise account of mathematical explanation in any of the writings of Kitcher, but, instead, of scientific explanation of physical phenomena. Nevertheless the possibility to export this model from physics to - pure - mathematics is proposed by Kitcher himself⁴⁰, in the light of his holistic point of view on scientific knowledge.

[G]iven my own views on the nature of mathematics, mathematical knowledge is similar to other parts of scientific knowledge, and there is no basis for a methodological division between mathematics and natural sciences⁴¹.

We will see in a moment how weak is this thesis, but what we want to save from Kitcher's way to set the problem is the global point of view on the problem of mathematical explanation - i.e. to consider how a general theory can explain some of the phenomena it is able to formalize or deduce - as opposed to a local point of view that tries to look for the explanatory characters of a proof. Indeed, this latter task is much more complicated and there are hints that it is not possible to give a detailed and objective account of why a proof counts for more explanatory than others. A seminal, but isolated, case of such a work can be found in Paseau's study of the different proofs of the compactness theorem ([138]). The main thesis of Paseau, with whom we agree, is that the explanatory virtue of a proof always depends on the context and so it is hard, if not impossible, to give an objective account of it. As we will see, one of the points we will make in this section is that also a global explanation depends on the background theory in a substantial way, contrary to a long tradition that dates back to Aristotle's times and that numbers, among his members, also Bolzano⁴². Leaving aside this difficulties, we want to stress how a global account

⁴⁰And sustained by his readers, as it is done in [164] - where Tappenden says (pp.158-159) "However, mindful of the fact that some explanations in physics and mathematics do seem to be governed by the same principles, I'll count it as an advantage of an account that it supports a uniform treatment of some mathematical and some physical explanations. A promising candidate to support a uniform treatment of some pure mathematical cases and some non-mathematical ones is the treatment of explanation as unification as proposed in the seventies by Michael Friedman and Philip Kitcher" - and in [118].

⁴¹[94], p. 423

⁴²See [114], for a detailed historical presentation of Bolzano's theory of mathematical explanation.

of explanation can in principle fit with a global mathematical point of view, as it is the case when dealing with a foundation of mathematics.

Once we are assured that the general setting of the two problems is compatible, let us see more in details Kitcher's account and his affinities with our proposed set-theoretical practical foundation for mathematics. The more important conceptual similarity is, of course, the individuation of unity as the main virtue of both a practical foundation and a global account of mathematical explanation. Indeed, when discussing Hempel's model of explanation, Kitcher describes his position in these terms.

This unofficial view, that regards explanation as unification, is, I think, more promising than the official view. My aim in this paper is to develop the view and present its virtues⁴³.

This is done at least at two levels: making clear the aim of an explanation and clearing the nature of an explanation. For what concern the former Kitcher is explicit in saying that an explanation is an answer to a why question. Exactly how, echoing Hilbert's foundation of geometry, our practical foundation aims to do.

I shall restrict my attention to explanation-seeking why-questions, and I shall attempt to determine the conditions under which an argument whose conclusion is S can be used to answer the question "Why is it the case that S "?⁴⁴

Notice that this declaration of intention is not as narrow as Kitcher seems to argue. Indeed, when we get to mathematical explanation, the question "Why is it the case that S "? can be interpreted in different ways according to the context in which the question is asked. For example we can take S to be Fermat's Last Theorem (FLT), but if we ask "why is the case that FLT" in a context where it is possible to understand and state the question, like that of number theory, we cannot even formulate a possible answer because we do not have an elementary proof of it. Contrariwise, this is a reasonable question in the context of a theory sufficiently strong to incorporate scheme theory and algebraic geometry. This example is meant to show that in the context of pure mathematics the methods of proof do have a role in the determination of an answer to a why-question.

⁴³[92], p. 508

⁴⁴[92], p. 510. Notice that this quotation manifests Kitcher's global point of view. Indeed, his aim is not to show which are the properties that an argument should have to be considered as explanatory, but which are the general - global - condition under which an argument should be considered as explanatory.

Moreover, what we suggest is that, in the particular case of mathematical explanation, an answer to a why-question can hide many different problems, like for examples considerations on the purity of methods, that bring together some controversial and less objective positions towards the nature of mathematical discourse.

For what concerns the second aspect Kitcher is explicit in saying that an argument is a derivation⁴⁵ and - this is the main thesis contained in [92] and [94] - that, given a set of sentences K , there is an *explanatory store* $E(K)$, that consists in the best *systematization* - read formalization in the context of the axiomatic method - of K . What makes $E(K)$ the best systematization is the possibility to associate to it a set A of arguments - called a *basis* - that instantiate general arguments patterns and that, better then other systematizations, unify K , in the following sense:

So the criterion of unification I shall try to articulate will be based on the idea that $E(K)$ is a set of derivations that makes the best tradeoff between minimizing the number of patterns of derivation employed and maximizing the number of conclusions generated⁴⁶.

It is not clear, in Kitcher's work, how this minimizing-maximizing effect should act in the process of choosing an argument instead of another⁴⁷. We maintain that Kitcher is appealing here to an intuitive principle of success that our arguments in $E(K)$ should surely fulfill. However, every attempt to clarify the notion of explanation that makes use of cardinality arguments or strictly syntactical ones - call them quantitative - fall short of the objection of the 'bad' - or casual - generalizations that unify but do not explain⁴⁸. We agree with [118] and [164] that a substantial account of explanation would need not just quantitative methods but also qualitative ones, that as Jamie Tappenden tries to argue in [164], need to take into account semantical consideration on the naturalness or the fruitfulness of a unifying principle. We will try to address some of the difficulties related to the qualitative methods with an analysis of the concept of naturalness, in the next chapter.

⁴⁵In Kitcher words: "a sequence of statements whose status (as a premise or as following from a previous members in accordance with some specified rule) is clearly specified". In [94], p. 431.

⁴⁶[94], p. 432.

⁴⁷The problem, with this counting arguments, is that it can work just in the case of finitely many consequences. And not with a mathematical theory with countably many consequences.

⁴⁸This is an objection discussed in length by Kitcher and attributed to Goodman. However, we will not enter in this debate.

What is important to stress here is that, for Kitcher, what is fundamental in the analysis of explanatory unification is the notion of argument pattern. Kitcher offers a description of these arguments as detailed as vague. What is important to keep from Kitcher's idea of argument pattern is that it is a general structure of an argument - not only a logical structure - sufficiently general to be applied in many different contexts and in many different forms: it is what permits to recognize that different proofs are essentially the same. We will not give a detailed presentation of the notion of argument pattern as it can be found in [92] mostly because of the lack in Kitcher's work of a clear analysis of the notion of similarity.

This suggest that our conditions on unifying power should be modified, so that, instead of merely counting the number of different patterns in a basis⁴⁹, we pay attention to similarities among them. All the patterns in a basis may contain a common core pattern, that is, each of them may contain some pattern as a subpattern⁵⁰.

Although Kitcher tries to clear what is an argument pattern, it is better to keep this notion as intuitive, although vague, as possible⁵¹. Then at this level of generality we could ask: which are the similarities between some axioms of set theory - especially those exceeding ZF, that we discussed in the last section - with respect to the argument patterns, as far as both are responsible for the unity of the theory? To answer this question we can recall Zermelo's idea about the "unlimited applicability of set theory". In order to make the argument more concrete recall the problem whether all the automorphisms of the Calkin algebra are inner. The way in which the proof works is by finding enough similarities between this algebra of operators and the structure $\mathcal{P}(\omega)/fin$. Then, it is possible to use the axiom OCA to perform the same argument on both sides. Indeed, this is a general methodology when facing a mathematical problem and looking for its solution. Indeed this is one the main advantage in relying on set theory as a foundations of mathematics: the concept of set and the methods used in set theory are so general and abstract that can be applied - possibly - to any field of mathematical inquiry. Hence, the use of the axioms for set theory

⁴⁹Just to recall it, a basis is the set of arguments that instantiate, in the more unifying. way all the relevant argument patterns of a given systematization.

⁵⁰[92], p. 521.

⁵¹In passing let us just notice how this vagueness on the concept of argument pattern can be related to a more qualitative analysis for the notion of explanation. Indeed any account, that aims to ascribe the explanatory power of an axiomatic setting that minimizes the *relevant* arguments patterns, seemed to be promising. But without a corresponding analysis of the concept of relevant this account would be useless.

permits to show the similarity, in the arguments, of many different mathematical reasonings. When attacking a problem, the first attempt of a mathematician is to bring the difficulties to a more clean and comprehensible level, where a solution is easier to find. This operation, of cleaning a problem from the irrelevant aspects, amounts in recognizing similar patterns or making more evident the core of the problem that often has - as the methodology of set theory shows, in abstracting from any content - a combinatorial aspect. Here by combinatorial aspects of a proof we do not only mean the part of an argument that is performed by pure calculation without a broader overview of the structure of a proof, but also the steps of an argument that manifest a necessity character similar, for strength, to calculation, and that act like the fundamental ingredient of a theorem. In a set theoretical context, the combinatorial aspect of a proof are often found when abstracting from the particular properties of the subject matter of a theorem and when outlining the general set theoretical properties that make possible to perform an argument. Indeed the combinatorial character of some axioms, or some principles that flow from them, is capable of showing in a pure form what is needed for the proof of a sentence: they show how to overcome the main difficulty one finds in a problem, acting as the key ingredient for its solution. Indeed this method works also in the opposite direction, from solutions to axioms and sometimes brings together the discovery of new, tacitly used, principles as it is the case of the Axiom of Choice. Even one of the more influential proponent of the more outstanding alternative foundation of mathematics: category theory, acknowledges the ability of ZFC to reduce many arguments to few.

The rich multiplicity of mathematical objects and the proofs of theorems about them can be set out formally with absolute precision on a remarkably parsimonious base⁵².

Moreover, whenever these principles are proved to be independent from ZFC, even if we do not have logical necessity, we have a deductively dependency of a proposition on the principle used in its proof that shows the insufficiency of other methods to solve a particular problem. Hence, as for the argument patterns, the axioms of set theory can be seen as the reasons for an argument to work. Then we can say that the axioms that extend ZFC can be considered as argument pattern.

The underlying notion of argument pattern is of course stretched to its limit and it could be argued that, in the context of the axioms of set theory, it is hardly recognizable. But our claim is not that argument pattern are set-theoretical axioms, because we acknowledge that there are of course different methods of

⁵²[108], p. 358.

proof and argument patterns in different areas of mathematics that have nothing to do with a set theoretical methodology. What we argue is that, when it comes to the foundation of mathematics, some axioms of set theory explains why a given proposition is a mathematical theorem, providing its proof; not why it is a theorem of a particular theory, say geometry or analysis. Moreover, we are not claiming that any set-theoretical axiom, singularly, can be seen as an instantiations of argument patterns, but that set theory as a whole can be seen as an explanation of why it is possible to prove a theorem - showing the core argument that allows a proof to work - once it has been cleared that the sense of explanation we use is related to a form of unification.

2.2.2 Kitcher's problem for mathematical explanation

We are aware of the fact that the arguing for the relevance of the role of the axioms in the context of argument patterns is a subtle and far from easy task⁵³ - even if Kitcher seems to support this view, as we will see in moment in the only passage were he discusses the possibility to use his account for analyzing the notion of mathematical explanation - but there is a preliminary problem that need to be cleared. Even if we give for granted that axioms act as - or instantiate - a form of argument patterns⁵⁴: what is the epistemological argument in favor

⁵³Notice that, at the ontological level, also Resnik does not exclude the fundamental relevance that axioms can have with respect to patterns: "But I must elaborate a bit on this answer, since one might remark that I am saying that, in effect, the premisses to which we appeal in proving the theorem "implicitly define" the pattern or class of patterns to which the theorem pertains. Now I have no problem *per se* with calling such premisses (or a more condensed set of axioms from which they might be derived) an implicit definition, so long as this is not taken to imply that the premisses are known a priori in some absolute sense. Of course the axioms constituting the clauses of an implicit definition are trivial consequences of this definition. Thus it is a matter of definition that they characterize the pattern they help specify." [146], p. 237-238. In general, the aims and the context of the notion of pattern, in Resnik's work, is very far from Kitcher's and our's, but when we get to the more fundamental elements of doing mathematics, even in a platonic context as Resnik's, the notions of axiom and of pattern tend to collide. What is really different here is the background idea of mathematics: a science of existing structures for Resnik, while the domain of rigorous arguments for us.

⁵⁴Notice that Kitcher rejected the arguments proposed by Friedman in [39] - who also proposed to identify explanation and unification - saying that the major difference between his account and Friedman's consisted in what was to be assumed as the basic notion in the process of explanation: for Friedmann were the physical laws, while for Kitcher were the argument patterns. In Kitcher words: "Finally, I think that it is not hard to see why Friedman's theory goes wrong. Although he rightly insists on the connection between explanation and unification, Friedman is incorrect in counting phenomena according to the number of independent laws. [...] What is much more striking than the relation between these numbers is the fact that Newton's laws of motion are used again and again and that they are always supplemented by laws of the same types, to wit, laws specifying force distributions, mass distributions, initial ve-

of the coincidence between unification and explanation? In other words, even if it seems at first sight a convincing matching, what are the arguments in favor of this identification? The main argument that Kitcher advances, to hold the epistemological link between the act of unifying a theory and that of explaining why some of phenomena described by the theory hold, is that the limit goal of all science is to unveil the causal structure of the world.

The growth of science is driven in part by the desire for explanation, and to explain is to fit the phenomena into a unified picture insofar as we can. What emerges in the limit of this process is nothing less than the causal structure of the world⁵⁵.

This position of course has profound consequences, one of which is the dependency of the concept of causality from that of explanation. Kitcher is well aware of this fact.

Indeed, I have been emphasizing the idea (favored by Mill, Hempel, and many other empiricists) that causal notions are derived from explanatory notions. Thus I am committed to

(2) If F is causally relevant to P , then F is explanatory relevant to P ⁵⁶.

Without entering in the discussion of the robustness of this philosophical position, we want to outline the main difficulty that this position suffers in the context of an analysis of explanation internal to mathematics: mathematics is not a causal world. There is a general agreement on this point and even a truly platonistic-minded thinker, like Gödel, has always advanced only an analogy between the physical world and the mathematical realm.

As we said before, Kitcher has never fully addressed the matter of the applicability of his model to the mathematical explanation. However, In the 1989

locity distributions, etc. Hence the unification achieved by Newtonian theory seems to consist not in the replacement of a large number of independent laws by a smaller number, but in the repeated use of a small number of types of law which relate a large class of apparently diverse phenomena to a few fundamental magnitudes and properties. Each explanation embodies a similar pattern: from the laws governing the fundamental magnitudes and properties together with laws that specify those magnitudes and properties for a class of systems, we derive the laws that apply to systems of that class", in [91], p.212. However Kitcher criticism is directed to some technical points raised by M. Friedman's proposal, hence nothing prevents, in principle, to argue in favor of the possibly that axioms - or laws - can capture some essential feature of an argument pattern.

⁵⁵[94], p. 500.

⁵⁶[94], p. 495.

paper *Explanatory unification and the causal structure of the world*, there is a, although short, attempt to discuss the problem.

For even in areas of investigation where causal concepts do not apply - such as mathematics - we can make sense of the view that there are patterns of derivation that can be applied again and again to generate a variety of conclusions. Moreover, the unification criterion seems to fit very well with the examples in which explanatory asymmetries occur in mathematics. Derivation of theorems in real analysis that starts from premisses about the properties of the real numbers instantiate patterns of derivation that can be used to yield theorems that are unobtainable if we employ patterns that appeal to geometrical properties. Similarly the standard set of axioms for group theory covers both the finite and the infinite groups, so that we can provide derivations of the major theorems that have a common pattern, while the alternative set of axioms for the theory of finite groups would give rise to a less unified treatment in which different patterns would be implied in the finite and the infinite case. Lastly, what Lagrange seems to have aimed for is the incorporation of the scattered methods for solving equations within a general pattern, and this was achieved first in his pioneering memoir and later, with greater generally, in the work of Galois.

The fact that the unification approach provides an account of explanation, and explanatory symmetries, in mathematics stands to its credit.

As it is clear, Kitcker argues that the bare possibility of applying the same pattern again and again is responsible for the unificatory virtue of a systematization without demanding a causal connection. This thesis is compatible with the claim that “If F is causally relevant to P , then F is explanatory relevant to P ”, but it is not with the idea that the unification process shows, in the limit, the causal relations between phenomena, because this would imply an even stronger thesis:

(2*) F is causally relevant to P if and only if F is explanatory relevant to P .

But this is of course false when dealing with mathematical explanation in pure mathematics. Then if we want to argue that axioms act as explanations we should look for a framework with different motivations than Kitcker’s, in order to justify the link between unification and explanation, in a non causal context.

As a matter of fact there is a theoretical, and not just methodological, difference between argument patterns in science and axioms in pure mathematics. What differs in the two contexts is the nature of the why-questions for which the explanatory unification looks for an answers. Indeed, if not partial, an answer to a why-question is such that it is not possible to ask any further why-question. When we get to pure mathematics and we discuss explanation in an axiomatic context we need to distinguish between the explanation of “why it is the case that S ”, for a sentence S and “why it is the case that A ”, for an axiom A . Kitcher chooses to explain “why it is the case that S ” appealing to the best possible unification of all sentences of the theory to which S belongs. This strategy can work as long as we remain in the context of physical phenomena, where we do not need to ask why questions on the physical laws. Indeed the example on Newton’s theory of gravitation, proposed in *Explanatory Unification*, is acceptable, since nobody would ever ask why it is the case for Newton’s law of Universal Gravitation. The reason is that reality serves as a bedrock in the search for the causes of a phenomenon. As a matter of fact, when Kitcher deals properly with mathematical explanation he proposes to explain group theory by means of its axioms, but what explain the axioms - as far as they are mathematical propositions - if we cannot make reference to a physical world where groups exist i.e. if there is no causal connection between the groups and the theorems of group theory?

Before trying to find a solution to the problem of “why it is the case that A ”, for an axiom A , let us look at the question of “why is the case that S ”, for a mathematical sentence S , without appealing to the causal structure of the world i.e. without appealing to the bad rock of reality that can stop the rise of new why-questions. If we are not dealing with a self-evident propositions, nor we are referring to some metaphysical property of mathematical objects⁵⁷, an answer to a why question, in terms of argument patterns, can be considered satisfactory only when we are not anymore in the position to ask reasons that could explain why some proposition hold. Only in this case it is possible to give

⁵⁷As for example Steiner seems to do in his *Mathematical explanation*. It seems that for him the question “why it is the case that S ” needs to be answered referring to some essential properties of the mathematical objects to which S refers. For example we can take S to be Fermat’s Last Theorem (FLT) and ask: why is the case that “FLT”? Steiner answer is that it is the case that FLT if there is a property of the natural numbers such that for every $n \in \mathbb{N}$ there is no positive integers a, b and c such that $a^n + b^n = c^n$. This type of answer recall closely a tarskian definition of truth: “for every $n \in \mathbb{N}$ there is no positive integers a, b and c such that $a^n + b^n = c^n$ ” iff for every $n \in \mathbb{N}$ there is no positive integers a, b and c such that $a^n + b^n = c^n$. However such a move, on one side, hides a strong realist position toward the existence of mathematical objects that needs to be argued and, on the other side, is tautological and hence non explanatory.

objective reasons for a proposition S . Then we cannot make reference to any extra-mathematical - informal - property, but we have to ground our answer on something as objective and indubitable as the logical structure of mathematics. In other words, the explanation needs to be internal to the mathematical discourse. So, if we accept that arguments are derivations - as it is the case in Kitcher's account - and the fact that axioms act as argument patterns, then, following Hilbert's suggestion, the best answer to "why is the case that S " amounts in showing the necessary and sufficient conditions for the proof of S .

However, such an answer seems to be clearly unsatisfactory when we restrict it to a single sentence S , because it is often the case that logical equivalences are not explanatory at all⁵⁸. But this objection misses an important aspect of doing mathematics, because the why-questions for which we are normally seeking answers are not "why is the case that S " independently from the mathematical context, but, once the background theory T is made explicit: i.e. "why is the case that $S \in T$ ". Indeed, granting that the act of explanation is a global matter, given by unification, necessary and sufficient conditions need to be given to explain the claim that S is a theorem of T . Hence when we restrict to countable languages our thesis is that a set of axioms $\mathcal{A} = \{A_i : i \in \omega\}$ can act as a unifying explanation of a theory T only if it is possible to show that, for any of its sentence S

$$S \in T \iff \exists n \in \omega \exists A_{i_0}, \dots, A_{i_n} \in \mathcal{A} \text{ such that } A_{i_0} \wedge \dots \wedge A_{i_n} \vdash S.$$

Nevertheless, the possibility to ask a general, context-independent, why-question has value and deserves to be considered. Then the question "why it is the case that S " becomes a question about the mathematical pedigree of S . If such a context-free question can ever find an answer, this will be in a sufficiently broad framework where it is possible to ask why we can consider S as a mathematical theorem: exactly the context given by a foundational theory as it is the case for set theory. For this reason we can say that set theoretical axioms as large cardinals and MM^{+++} can be seen as explanations of a mathematical proposition S , for what concern the question "why it is the case that S " i.e. why S is a mathematical theorem.

To sum up our argument so far, we tried to check the relevance of Kitcher's theory of explanation to the framework of the set theoretic foundation of mathematics we proposed, inspired by the common goal of unification. The similarities has been found in considering why-questions - even though Kitcher does not analyze the peculiarity of the mathematical why-questions - and in arguing for

⁵⁸This is not always the case as some equivalence theorems show. FIND AND EXAMPLE.

unification on the ground of the possibility to perform similar argument again and again - on the foundational side this aspect is made evident by the combinatorial form of many set theoretical principles. So we argued in favor of an analogy between axioms and argument patterns - also acknowledged by Kitcher. However a major conceptual difficult has been found in Kitcher account, when applied to mathematics: the non causal story of mathematics. This led us to propose a different link between unification and explanation, in the context of Kitcher's proposal applied to pure mathematics: the possibility to show that mathematical axioms can act as necessary and sufficient conditions for a whole mathematical theory, and not for just a single sentence.

Then we are finally in the position to come back to our initial problem: to give a philosophical justification of the claim that set theory can be seen as a foundation for mathematics as long as it is capable of unifying mathematical practice. The answer then is to be found in its possibility to explaining mathematical phenomena, given necessary and sufficient condition, at least when it is possible to make clear reference to a theory T , that we can easily describe and recognize; as it is the case for an initial segment of the cumulative hierarchy, in terms of an $H(\theta)$, as it has been seen for Woodin's and Viale's results.

However we are left with the problem "why it is the case that A ", for an axiom A , that is, the search for justifications of the axioms. And remember that our goal, at the beginning of this section was to give a philosophical analysis of the criteria of unification in the search for sound and rational reasons to accept new axioms in set theory. The outcome of this inquire is that explanation and justification are tied together by a sort of completeness theorem that links axioms and propositions, in the attempt to unify a theory. One side of the if-and-only-if-condition, from right to left, shows how it is possible to explain that a given proposition S belongs to some theory T - this is done by showing that S is a consequence of the set of axioms A . Then, starting from the axioms, we have an implicit definition of T , as it was indeed the case for Hilbert's Axiom of Completeness. On the other hand, the implication from left to right presupposes an intuitive description of T and then asks for the axioms that can prove the whole of its theorems and, thus, unify the theory - in the sense described by Kitcher, as argument patterns. If this second implication hold than it is possible to match the intuitive theory and its axioms, and so we are able to justify the axioms in term of their unification power. This is the place where extra-mathematical, qualitative aspects come into play in the justification of the axioms: where it is possible to find consideration of naturalness in the search for new axioms⁵⁹.

⁵⁹Tappenden, in [164] (p. 176), has rightly pointed in the direction of the rise of naturalness judgment in the context of qualitative argument in any account of mathematical explanation:

These judgments are indeed related to an informal description of a theory T and so presuppose its intuitive description. To come back to the argument we proposed at the end of our historical examples of the axiom-as-explanation position, we think that justification and explanation are two side of the same coin: a complete unification. Indeed unification allows the proof of completeness theorem of the form we have just described, where a link is established between syntax and semantics. The correctness direction, from left to right, amounts to the justification of the axioms of a theory T , while the completeness direction, from right to left, amounts to the explanation of why the sentences can be seen as proposition of T . Hence we maintain that the answer to the question “why it is the case that A ”, for an axiom A , consists in its justification - whenever it is possible to give an intuitive description of a theory T , for which A acts as an axiom - thanks to a completeness theorem of form we have just outlined.

The goal of the next chapters will be to assess the naturalness of the Forcing Axioms making clear which is the intuitive description of set theory with which these axioms match. Then, finally, the unification criterion will be philosophically explained and the Forcing Axioms will be justified also in term of this criterion.

In the end, we want to be clear that our thesis is neither that explanation always comes, in mathematics, through axioms, nor that the explanatory unification of set-theoretical principles is always granted by their acting as argument patterns. Indeed, as we argued, the epistemological import of an equiconsistency proof gives different reasons for the explanatory role of large cardinal axioms. However, the possibility to apply Kitcher’s model to some axioms of set theory is intended to show that explanation is part of the role of these axioms, but neither to make a general theory of the nature of the axioms in mathematics, nor to make a theory of mathematical explanation. On this latter aspect, we acknowledge that explanation does not always come in the context of an axiomatic setting, nor all the axioms are capable of explanation. As a matter of fact, for what concerns explanation we favor a more pluralist conception, capable of taking into account all the different nuances that can have a mathematical explanation, internal to the mathematical work.

Moreover, although Woodin’s and Viale’s results point in the direction of a complete axiomatization of set theory and we maintained that an intuitive description of a theory is needed in order to give a complete axiomatization of a theory T , we do not want to argue neither that an analysis of the concept

“The self-conscious focus on producing general schemata is not an unconditional goal. Once again it is important not only that the properties unify but that they are otherwise the ‘natural’ or ‘right’ ones.”

of set, nor an intuition to this concept - *à la* Gödel - is needed in order to give a foundation of mathematics. Quite the contrary, as we hinted before, the vagueness of this concept is the main reason to argue for the set-theoretical foundation of mathematics we proposed. As a matter of fact, we think that a foundation of set theory - with the word ‘foundation’ intended in the sense explained by Gödel: “a procedure aiming at establishing the truth of the relevant mathematical statements and at clarifying the meaning of the mathematical concepts involved in these theories⁶⁰” - is really a different task and we will try to make this distinction clearer in the last part of this section.

2.2.3 Towards a more general distinction

To summarize, we hope to have been able to show that set theory, in the extended sense considered, is a good tool in the analysis of necessary and sufficient conditions for the proof of all mathematical problems and in this sense it is to be intended as a foundations for mathematics. The instruments it provides go much beyond the possibilities that are given by the use of solely logical tools - that encouraged the vast application of the axiomatic method in the last century. As a matter of fact, thanks to equiconsistency results, it is possible to find equivalence results that are not only logical, but epistemological in character; and this analysis is a good form of explanation, in terms of the main possibility of proof - able to unveil deep combinatorial aspects. Moreover it is important to stress the difference between the set-theoretical foundation we described and the standard view that sees a big ontological import in the possibility to reduce every piece of mathematics to set theory. As a matter of fact the foundation of mathematics we argued for is ontologically and theoretically neutral: it does not even take a stand about the attempt to single out the true universe of set theory, in the context of the multiple alternatives offered by the method of forcing. This line of research is an interesting and fruitful subject, but it has important theoretical implications that cannot be compatible with a foundation that aims to explain practice and then follows the free and unforeseeable development of mathematics.

Indeed there is an important debate on the main possibility to find such a complete description of V , where platonic-minded mathematicians, like for example Woodin, are opposed to researchers that hold a multiverse point of view, like for example David Joel Hamkins⁶¹. There are also positions in between like Magidor’s who maintains that “some set theories are more equal than others”⁶²

⁶⁰[121], p. 86.

⁶¹See, for example [53].

⁶²This is the title of the draft of a talk that Magidor, gave in Harvard in 2012.

or, similarly, like Sy Friedman and Sharon Shelah, who argue in favor of the possibility to find rational arguments for choosing one model instead of another⁶³. Of course the main existence of these different positions is sufficient to show that this is subject matter for philosophy; and so cannot have a role in the context of a foundation for mathematics in the sense in which it is proposed here. As a matter of fact, we think that a multiverse view on the nature of the set theoretic universe just confuses the foundational role of set theory with its nature of mathematical theory in itself for which the search for a good description of the intended model is a fundamental and natural demand. On the contrary, even if this would be found it would not disqualify all the theorems that does not hold in that model. As a matter of fact the distinction between foundational aspects and infra-theoretical ones is meant to legitimate both analyses.

In conclusion, we want to stress the importance of not confusing the foundational role of set theory with its nature of mathematical theory in itself, for which the search for a good description of the intended model is a fundamental and natural demand.

2.3 Two different foundational ideas

We believe that the difference between the two set theoretical foundations of mathematics we discussed before is the appearance of a more general phenomenon: two distinct attitudes in the foundation of mathematics. Of course we do not pretend to give an exhaustive classification, but at least to indicate that there are two areas that deal with foundational problems with distinctive perspectives: philosophy and mathematics. These two attitudes are of course well interlaced in the foundational works of the last century, but they are, in principle, autonomous. As a matter of fact, these two dispositions act in response to different needs. The choice of the terms to indicate them could be *theoretical* and *practical*. We could have called them philosophical and mathematical but this choice is somehow misleading, because on the one hand there is no sharp distinction between the two subjects at a foundational level and, on the other hand, we do not want to suggest an opposition between philosophy and mathematics, but, on the contrary, a distinction that can produce useful interactions. The antinomy that I would like to propose with this categorization is the one existing between essence and method.

We will use the expressions “theoretical foundation” and “practical foundation” to indicate the corresponding attitude in the foundational enterprise. We will quote same examples of these approaches, but we would like to be clear

⁶³On this topic see the work of Friedman [41] and Shelah’s *Logical dreams* [152].

that we are not proposing a classification of philosophers and mathematicians in two separate categories. On the contrary, we just delineate a distinction for what concerns goals, approaches and, sometimes, admittedly, true predilections. Indeed, it will always be difficult to draw a clear line to distinguish the two kind of foundations in the work of an author, since the reflection on mathematics is always a difficult and broad enterprise. As we outlined in the first part of this chapter the main concern of a foundation is unity. For this reason a theoretical foundation and a practical foundation are both stimulated by this idea and we will describe how they achieve this purpose.

2.3.1 Theoretical foundation

By a theoretical foundation we mean the attitude that sees in the foundation of mathematics the possibility of a *reduction*. This stance tries to answer a question on what there is in the mathematical world and how we can give a mathematical definition of our mathematical concepts. A reduction of this kind deals mostly with ontological or semantical problems; as, for example, in the case of the reduction of mathematical objects to sets (with all the problems related to the fact of assuming that everything is a set⁶⁴). See for example the very beginning of Kunen’s “Set theory. An introduction to independence proofs”, one of the most used textbook in set theory:

Set theory is the foundation of mathematics. All mathematical concepts are defined in terms of the primitive notions of set and membership. In axiomatic set theory we formulate a few simple axioms about these primitive notions in an attempt to capture the basic “obviously true” set-theoretic principles. From such axioms, all known mathematics may be derived⁶⁵.

Another example of this approach can be found in Russell’s logicist program, for which the reduction is even more conceptual.

In constructing a deductive system such as that contained in the present work . . . we have to analyse existing mathematics, with a view to discovering what premisses are employed, whether these premisses are mutually consistent, and whether they are capable of reduction to more fundamental premisses. . . . [T]he chief reason in favor of any theory on the principles of mathematics must always lie

⁶⁴See [14].

⁶⁵[102], page xi.

in the fact that the theory in question enables us to deduce ordinary mathematics.⁶⁶

In an opposite way, also, any attempt of nominalization of the mathematical discourse can be seen as a form of reduction; a reduction of the truth value of a mathematical sentence to a syntactic game that can be played uniformly within any mathematical theory⁶⁷. As a matter of fact, the answer to the question on what there is can be answered in many and incompatible ways, like, for example, everything or nothing. What is peculiar to this attitude is that it tries to give a comprehensive reduction of the whole of mathematical discourse, or sentences, or truths, to some objects or principles that are able to subsume or vanish any peculiar aspect of a particular mathematical field. Not only this kind of foundation tries to unify but also disappear the differences, explaining that the various things we encounter in our mathematical experience are just diverse manifestations of the same phenomenon. What is common to the foundations that share this goal is an holistic and static view of mathematics, that sees mathematical practice as the field where to test if the reduction proposed is sufficiently comprehensive.

Of course there are problematic aspects of a theoretical foundation. These problems arise in trying to give a general account of mathematics and not only of its unity. First of all there is the matter of fact that mathematics is an always evolving enterprise. This makes very difficult to single out, once and for all, the very characteristic marks of mathematics and moreover to confine its existence within rigid boundaries. The horizon of sense and application of mathematics is always moving and follows freely the heavy burden of its history. Secondly there are problems of reference, or aboutness, as in the case of numbers and sets, as outlined in Benacerraf paper *What number could not be*. Indeed, once a reduction is proposed, there should be arguments in favor of that particular reduction instead of another, maybe, of the same kind. For example, following Benacerraf, once we admit that numbers are sets we should be able to explain which sets are the numbers. Finally, and related to this latter point, there is always a metaphysical obscurity that surrounds any reduction: how this reduction works? what is the relationship between what is reduced and the the tools of reduction? We will not try to give an answer to these questions, because this is not a necessary task for a theoretical foundation, even if, of course, we have to admit that these questions deserve an answer, in the context of a philosophical account of mathematics. We would like here just to outline this view, clear its weaknesses and not try to defend it.

⁶⁶[149], Preface, page v.

⁶⁷See, for example, the work of Hartry Field.

2.3.2 Practical foundation

The second attitude we would like to describe is the practical foundation: it aims to *explain* the unity of mathematics without proposing a reduction and it is epistemological in character. The main question that it tries to answer is: why can we prove a theorem? why a proposition can be seen as a theorem of a theory? The main reason for calling it practical, in contrast with theoretical, is the attention that is devoted to mathematical practice. As a matter of fact the motivation for such a foundation is the observation that doing mathematics consists essentially in trying to prove theorems. Moreover this attitude grants that one of the most important task of a serious reflection on mathematics is to explain the nature and the possibility of mathematical knowledge. In contrast with a theoretical foundation, a practical foundation of mathematics is not confined to a fixed set of axioms or to a given set of primitive principles, as in the case of the *Principia Mathematica*, but it makes use of the axiomatic method, trying to give a detailed description of the mathematical work. In this context the unity of mathematics is suggested as a methodological uniformity. The main goal of a practical foundation is to explain in what consists the procedures that allows to recognize an argument as a proof. To qualify something as a proof has the consequence of characterizing the proposition that is proved as a piece of mathematical work. The roots of this attitude can be found in Hilbert's foundational work on geometry. Remember the letter to Frege, dated December 29th, 1899, when Hilbert says that he wanted to understand *why* "the sum of the angles in a triangle is equal to two right angles"⁶⁸.

In a different way, with respect to the role of the axioms in defining the basic ideas of a theory, an attitude of this kind can be found also in Frege's foundation of arithmetic.

By insisting that the chains of inferences do not have any gaps we succeed in bringing to light every axiom, assumption, hypothesis or whatever else you want to call it on which a proof rests; in this way we obtain a basis for judging the epistemological nature of the theorem.⁶⁹

We can see here what explanation means in the context of a practical foundation. The explanation that is given is internal to the theory for which the foundation is proposed. Indeed the explanation of a theorem is given in terms of its place in the logical structure of a theory, as for Hilbert, or in terms of the

⁶⁸[36], pp. 38-39)

⁶⁹In [35], introduction.

elucidation, step by step, of a proof, in Frege's proposal. Then the reasons that explain are to be found in the axioms that characterize a domain of knowledge or in the tools that we use to get from the premises of an argument to its conclusion. It is important to note that in these cases nothing depends on some metaphysical property of the subject matter nor to the recognition of a sort of similarity in the nature of the things involved in the foundational analysis. With a practical foundation mathematical practice is investigated in details and the quest for the reasons terminates only when we stop asking 'why' questions.

We have to pause here for a moment, because a digression is needed for the role that logic plays with respect to both types of foundations; and meanwhile to illustrate why we can find the two main champions of logicism, Frege and Russell, in different horns of the dichotomy we are proposing. The reason for it is the twofold nature of logic. There is an old and venerable tradition, that can be traced back to Leibniz, that acknowledges logic, on the one hand, as a *characteristica universalis* and, on the other hand, as a *calculus ratiocinator*. The former aspect stresses the fact that logic is a universal language that can express any mathematical concept, while the latter indicates the circumstance that logic can be used to perform formal deductions. This bivalent character of logic can be found also in Frege's work but the passage quoted above shows that the relevant feature of his *Begriffsschrift* is to explain why a conclusion follows from its premises, in terms of a rigorous deduction. While Frege maintained that these two aspects of logic cannot be disentangled, he emphasized that the *characteristica universalis* aspect was the most important⁷⁰. However Russell goes much further in the direction of the *characteristica universalis* and he says not only that logic is the language in which mathematical concepts can be expressed, but that every piece of mathematics can be defined in terms of logic. Then, while Frege has a universal view about logic but he thought that the computational aspect was really necessary for any meaningful notion of logic, on the contrary Russell sees in its work a more fundamental reductive stance.

Single theory vs. mathematics

By looking at the quotations above it could be thought that a practical foundation is context-dependent and it works only when a single theory needs a foundation, and not the whole of mathematics. Indeed Hilbert's work was on geometry, whereas Frege's was on arithmetics and, thanks to their work, it is possible not only to explain what is needed for the proof of a proposition but also why we can recognize it as a proposition of geometry or arithmetic. However,

⁷⁰Contrary to Schröder critiques, according to which Frege's foundational work tended mostly towards a *calculus ratiocinator*. See [141] for this debate.

in this case: when a foundation of a particular theory is proposed, it arises the problem of the adequacy of an axiomatization to the theory that is axiomatized. We will not tackle this problem here in its generality, because this involves issues such as clarifying what a mathematical concept is, how it is possible to formalize it and how we manage to know what we formalize. Even if the attempt to give an answer to these questions is among the central tasks of the philosophy of mathematics, it is not in the scope of this work. Both Frege and Hilbert had their personal solutions to the problem of the adequacy: the former believed in the existence of a realm of concepts, while the latter discarded the problem using implicit definitions. What is important to stress here is that there is an insolvable tension between intuition and formalization, that, in the context of a theory for which we feel to have strong intuitions about its subject matter, can rise deep philosophical questions. In the case of geometry or arithmetic there is a tentative solution that comes directly from mathematics: a categoricity proof like Hilbert's for analytic geometry, or the one that is possible to give for natural numbers, using second order Peano axioms. Leaving aside the discussion on the significance of a categoricity proof, we just acknowledge that there are also situations where there are not even such results, as it is indeed the case for the formalization of set theory proposed by Zermelo and Fraenkel, for which it is not even sufficient using second order ZFC, as it is shown by Zermelo's theorem on the quasi-categoricity of the universes of set theory⁷¹.

Despite all the difficulties that emerge in the case of a practical foundation of a single theory, we would like to argue that this is not the case for a practical foundation of mathematics, as a whole - as the set-theoretical one we discussed in the beginning. In this case we do not have neither the problem of reference, as for a theoretical foundation of mathematics, nor the problem of adequacy, as for a practical foundation of a single mathematical theory. Indeed, it is not necessary to know the subject matter of mathematics before we can propose a practical foundation for it. Or, to put it in a different way, knowing why we can prove a theorem does not entail knowledge of what the theorem is about.

In the case of set theory the same difficulty to develop a reliable intuition of the general concept of set was sufficient to show the independence of a practical foundation from a complete knowledge of the matter for which a foundation is sought. I would like, in conclusion, to discuss another example where, even if we feel to have strong mathematical intuition, we can still mark a conceptual distinction between a practical and a theoretical foundation. Let us consider the case of arithmetic. In general, the fact that we know which principles allow

⁷¹See the article "On boundary numbers and domains of sets. New investigations in the foundations of set theory" in [192].

us to solve a problem in number theory does not depend on our knowledge of what natural numbers are. Indeed, there are many cases in which tools that transcend arithmetic are used to solve a problem in number theory, as in the case of Fermat's Last Theorem, while, on the contrary, just second order Peano axioms are able to fix the structure of the natural numbers. This situation could be seen - and it is often seen - as an historical accident. Indeed it is common opinion among mathematicians that for any relevant number-theoretic statement it can be found a proof in elementary number theory. This belief would involve an extensive coincidence of the set of principles that allow to give an explanation of the "epistemological nature of a theorem" in number theory and the set of axioms that are able, in second order logic, to characterize the structure of natural numbers. This then could be seen a cause of ambiguity between the two different foundations that we are proposing. Nevertheless, even granted this quantitative coincidence, there is a qualitative difference in looking at the axioms as characterizing natural numbers and as tools that characterize the work in number theory. In the former case we are tempted to say that the truth of a proposition in number theory depends on the fact that Peano Arithmetic is the right formalization the "natural numbers", while in the latter that it depends on the knowledge of which principles - or axioms - we are using in its proof. This is the reason why it would be a mistake to confuse the level of explanation - of why we can prove a theorem - and the level of justification - of the proof of a theorem in terms of the nature of the terms involved⁷².

In the case of a practical foundation of the entire mathematics this point is even more evident, because we do not have a clear idea of what the subject matter of mathematics is. Quite the contrary, we have a vague and ambiguous intuition of it, whence spring our feeling that mathematics is completely free in its paths and development. To make clear the borders of mathematics is exactly the purpose of a foundation, hence shaping mathematics. This is the reason why we cannot know what mathematics is before giving it a foundation. It is then clear that being able to explain the facts that we encounter in our mathematical practice does not presuppose a precise knowledge of its objects.

In conclusion, we hope to have given a clear picture of these two different aims in the foundations of mathematics. However we do not want to argue in favor of a separation of a more philosophical attitude from a more mathematical one. Of course a useful interaction between these points of view is not only the best way to find a deep understanding of our mathematical experience, but also a good guide for our mathematical work. The recognition of a conceptual distinction

⁷²This distinction echoes the disagreement on the role of the axioms between Frege and Hilbert and the point that we are trying to make is on the same line of Hilbert.

between two different attitudes in the foundational studies does not involve a separation of them in practice, as working tools in the attempt to account for the mathematical phenomena and to wide our mathematical knowledge.

Chapter 3

Naturalness in mathematics

Is perhaps the right way of tackling the question just this – to write down a long list of actually observed uses, taking note of the frequency of each use, and distilling the whole into a statistical table? But is this the sort of a thing a philosopher wants to do? Is he interested in the random fluctuations of speech, that sea with its endless waves and ripples?

F. Waismann, *Analytic–Synthetic IV*

Our attempt in this chapter is to give a philosophical characterization of the notion of naturalness in mathematics. First of all, we have to acknowledge that this is not an easy task for many historical, methodological and intrinsic reasons. To start with, there is not a wide and well-structured literature on this topic¹ and so every step in this direction will be almost like groping in the dark. On the contrary, there is an important philosophical tradition that is labeled naturalism and that will make our investigation even harder, because, as we will argue later, it is quite far from the position expounded here. We will mainly discuss Penelope Maddy's position, in the attempt to clear our view of a dialectical relationship between mathematics and philosophy.

On the methodological side, we believe that our analysis pertains to the philosophy of mathematical practice. Then we have to face the difficulty that such a fairly new branch of philosophy encounters: the absence of a well established method of inquire. As a matter of fact, the first part of this work will be concerned with explaining our argumentative line. Once we find out that the concept of naturalness deserves an analysis we will try to find a suitable philosophical treatment of it, in the context of this new wave in the philosophy of mathematics, hoping, at the same time, to throw light on the methodology

¹We list here all the relevant works, according to our knowledge, that address directly the problem of naturalness in mathematics: [164], [165], [166], [22], [97] and [8].

of the latter. Moreover, although every work that can be labeled as philosophy of mathematical practice brings within itself an inevitable attention to concrete cases, we will distinguish the relevance we give to mathematical and historical examples from the one that is normally given by naturalism.

Our analysis will start from the statistical evidence that the use of the word naturalness has noteworthy increased in the last seventy years. Some methodological considerations then will be needed in order to justify the use of linguistic instruments in a philosophical work. As a matter of fact, we will argue that a statistical overview does not exhaust our analysis and that the descriptive temptation is a pernicious solution that has to be avoided both at the linguistic and the conceptual level. Then, after we reject a Philosophy First approach to the problems of the philosophy of mathematical practice, we will look for a sufficiently neutral starting point for our study. The - in comparison - vast literature on mathematical explanation will help in this task and, inspired by this possibility, we will outline a general method for dealing with the vagueness of some philosophical concepts. Then we will propose a semi-algorithmic method and we will test it in our case study with the aim of finding some general character of naturalness, using some historical examples and concluding that the object of our study manifests both a dynamic component, instead of a static one², and a prescriptive component, instead of a descriptive one.

In the end of our analysis we will take a stand with respect to the issue of the role of the common sense in the philosophical inquire of words that have - also - an intuitive meaning; as it is the case of naturalness. Indeed, when calling something natural we rhetorically evoke the idea of “pertaining to nature”, even if the causal context does not apply to mathematics.

This semantic ambiguity of the notion of naturalness, that we encounter in our work, brings us to the third difficulty in any attempt to give a coherent investigation on how and why naturalness is so much used in mathematics: its intrinsic tension towards different poles. Indeed the fast development of a more and more abstract and artificial mathematics, in the last century, seems to diverge from a natural point of view. Hence how to fit this historical phenomenon with the increasing appeal to natural components of mathematical discourse? Moreover, granting that there is a grain of truth in our thesis that dynamics, referential and contextual aspects are fundamental in forming judgments of naturalness, how it is possible to match this discovery with the allusion to nature, that inevitably makes more static and objective what is called natural in mathematics? Just to counter some obvious criticism that can possibly rise, we acknowledge that there are pieces of mathematics that are so clearly stable

²We will clear the meaning we assign to the dynamic-static dichotomy in a moment.

in their naturalness, that any attempt to give a philosophical account of their natural character would necessarily deal with some sociological, or cognitive, or transcendental aspects of our doing mathematics; as, for examples, the natural numbers. On the same par we maintain that there are so unnatural phenomena that would always be outliers with respects to any reasonable account of naturalness in mathematics. As a matter of fact, what interests us here is not the stable components that we can find at the extremities of the dichotomy natural-unnatural, but the different nuances of gray that we can find in between. We think that it is here that the reference of naturalness diverges more substantially from the common use of this notion. Hence this latter deserves a philosophical inquiry. For what concerns these inner tensions and bivalent characters of the use of some words, we sympathize with Friedrich Waismann.

In all these case you notice that expressions which have a trivial use in everyday life, when made part of a certain trains of thought, lose their triviality, become, as it were trascendentalized, and acquire metaphysical status³.

Far from arguing for a truly metaphysical component of the mathematical work, we will try to make sense of this inner tension of the concept of naturalness, showing that the increasing reference to natural components in mathematics is not philosophically innocent. We will then try to unveil the philosophical positions that lie behind its use and to explain the link between the apparently contradictory aspects of the concept of natural and its philosophical imports. At this level we will find the more relevant distance between our position and Maddy's. To refrain a famous Quinean slogan: the notion of naturalness is philosophy in mathematical clothing.

3.1 How to deal with this concept?

When facing a new conceptual problem, like in this case, where to begin if not from the literal evidences that concepts leave in their use? We do not argue for an holistic point of view towards mathematics and natural language, but we just choose to start from what is more certain and secure, in order to explore what is less known and more obscure. Let us look at the definition of naturalness⁴

³[189], p. 56.

⁴We agree with Bertrand Russell that "The study of grammar, in my opinion, is capable of throwing far more light on philosophical questions than is commonly supposed by philosophers. Although a grammatical distinction cannot be uncritically assumed to correspond to a genuine philosophical difference, yet the one is *prima facie* evidence of the other, and may often be most usefully employed as a source of discovery" (in [147], p. 42).

that we find in the Oxford dictionary: “Existing in or derived from nature; not made or caused by humankind”. By hinting at these possible answers to the question ‘what is naturalness?’ we just want to stress that aside what we will find in our analysis, there is a common-sense meaning of this word that needs also to be considered. But then, is this reference to the dictionary’s meaning enough to grasp the whole semantic of naturalness in mathematics? We doubt it. Consider, for instance, these few occurrences of the term⁵:

The proof presented in Section 4.1 is similar to Kruska’s original proof in (5). However, we add more clarification to it in order to show that the proof is natural and intuitive.⁶

Since the operations TC are commutative, associative, monotonic, continuous in the topology of weak convergence, etc., this shows that there are very natural operations on distribution functions that do not correspond in any simple fashion to operations on random variables.⁷

Semirings, in the general setting as described above or with more restrictive assumptions, arise naturally in such diverse areas of mathematics as combinatorics, functional analysis, topology, graph theory, Euclidean geometry, ring theory including partially ordered rings, optimization theory, automata theory...⁸

To claim that the use of naturalness in these examples is fully captured by our two dictionary entries is by no means conceptually easy. What does it mean for semirings to exist in nature? - and in this case what kind of nature are we talking about? Or, in which sense a class of operations on a topology is not made or caused by humankind? Any of these questions arises a bunch of well-known and hard philosophical issues, so that it is unrealistic to consider them solved by every person who uses the term natural in his mathematical practice. On the contrary, one may simply acknowledge that, although its use is well understood, every such appeal to naturalness in mathematics contains a semantic that is not

⁵We deliberately chose these examples randomly from the mathematical literature.

⁶A. Stegeman, N. D. Sidiropoulos. On Kruska’s uniqueness condition for the Candecom-Parafac decomposition *Linear Algebra and its Applications*, 420 (2-3), pages 540-552 (2007), p.542

⁷A.Sklar. Random variables, joint distribution functions, and copulas.*Kibernetika*, 9 (6), pages 449-460 (1973), p.457

⁸H.J. Weinert. Review of *The theory of semirings (with applications in mathematics and theoretical computer science)* by J.S.Golan. Bulletin of the American Mathematical Society, 30 (2), pages 313-315 (2004), p. 314

philosophically trivial to clarify. When starting this inquiry our feeling is quite similar to that of Augustine about time: “*Quid est ergo tempus? Si nemo ex me quaerat scio; si quaerenti explicare velim, nescio*”⁹. Indeed any appeal to naturalness is perfectly understood by the mathematical community, without any attempt to define and formalize this concept.

But up to now, we have still ignored the question: why dealing with naturalness in mathematics? The answer is clear from a random inspection of any contemporary mathematical journal: the presence of the word naturalness, or natural, or references to naturalness are ubiquitous in the mathematical literature. In order to support this claim we did not feel satisfied with a possibly accidental picture, but we tried to make an informed statistics of the phenomenon.

“Don’t think, but look!” (*Philosophical investigations* 66). Following this Wittgenstein’s remark, we begin our work with a glance at the *American Mathematical Society* database (*MathSciNet*). We want to be clear, since the beginning, that our work is not sociological: we use this evidence to call attention on a general phenomenon that we want to analyze with philosophical tools¹⁰.

The following table gives us the frequency of the use of “natural” and “naturalness” between 1940 and 2009:

⁹“What then is time? If no one asks me, I know what it is. If I wish to explain it to him who asks, I do not know.”, *Confessions*, XI, 14

¹⁰Few words concerning our corpus: the MathSciNet database consists entirely of mathematical reviews. However, we believe that there is not much difference in the prose of a review from that of an article. Moreover, the emergence of the phenomenon we will describe is so strong that any small distortion of the data cannot hide the emergence of the use of naturalness in the mathematical literature. Finally, we would like to stress that there is no *wide* corpus of mathematical text ready for a corpus linguistics analysis. Indeed, consider that the Corpus of Contemporary American English (COCA) counts more than 450 millions words, whereas MathSciNet consists of 2.949.420 reviews. The other attempt to perform a similar linguistic analysis in a mathematical context, known to the authors, has been presented by Lorenz Demey at the ILLC’s Logic Tea, on April 21st, 2009 and it makes use of a small corpus of less than 3 millions words. In conclusion, we believe that our starting point, even if partial, is representative enough for the described phenomenon, although a more detailed analysis would need a much larger corpus. Nonetheless, as it will be clear later, we do not feel that the absence of a such a linguistic tool is a limitation for the goals of our work.

<i>Decade</i>	<i>Total articles (T)</i>	<i>Occurrences (N)</i>	<i>Rate ($\frac{N}{T}$)</i>
1940 – 1949	40538	602	0.014
1950 – 1959	89158	1935	0.021
1960 – 1969	168567	4802	0.028
1970 – 1979	327427	11500	0.035
1980 – 1989	483143	21026	0.043
1990 – 1999	617522	34032	0.055
2000 – 2009	841470	47056	0.056

It shows that during the last decades there had been a clear increasing appeal to this concept. However, data need always to be handled with particular care. Natural embraces a bunch of various and heterogeneous meanings and these uses may be divided in two - absolutely non-exhaustive - classes:

1. natural number, natural deduction, natural proof, natural transformation, natural isomorphism, natural topology, ...
2. natural method, natural way, natural solution, natural explanation, natural argument, natural example, ...

The items on the first list are formal definitions in which natural gains some technical meaning. On the other hand, the second list reflects an informal employment of the term, showing a strong inclination to assign an appearance of naturalness to mathematical practice. The term occurs both on formal and informal sides, spanning a wide semantic domain from being specific to some rigid contexts (list 1.) - once a formal definition is given and so the whole meaning of the expression is fixed - to some other relaxed and variable uses (list 2.). We believe that the contexts of use of the term “natural” are not equivalent and they exhibit a twofold aspect of this concept. In formal definitions, naturalness is involved only in some initial stages; e.g. there are relevant reasons why Gentzen called the system of natural deduction “natural”, but almost every further use is actually unrelated to considerations on naturalness¹¹. Moreover, if we want to understand the meaning of these terms, there is not much to do more than looking for some equivalency results, since their semantics is fixed by the deduction rules. There are attempts in this direction, by Shelah and Hodges, who describe their intention as follows, in [86].

Eilenberg and Mac Lane [...] explained the notion of a ‘natural’ embedding by giving a categorical definition. Starting from their

¹¹The same can be said for the use of the term “natural” in category theory.

examples, we argue that one could equally well explain natural as meaning ‘uniformly definable in set theory’. But do the categorically natural embeddings coincide with the uniformly definable ones?¹²

On the contrary, the picture of informal occurrences is messy and dynamic: to call a portion of a mathematical work “natural” is a meaningful operation that consists in assigning to a definition (an axiom, a proof, a construction etc.) an informal feature; we may already hint that such operation is somehow metalinguistic, in fact it expresses a sort of comment in the margin of our formalization. But there is something more. Statistically, the use of the items on the first list remains almost stable during the decades - or, sometimes, it even decreases. For example, this is the table concerning the string “natural number”¹³.

<i>Decade</i>	<i>Total articles (T)</i>	<i>Occurrences (N)</i>	<i>Rate ($\frac{N}{T}$)</i>
1940 – 1949	40538	92	0.0023
1950 – 1959	89158	401	0.0045
1960 – 1969	168567	1182	0.0070
1970 – 1979	327427	2456	0.0075
1980 – 1989	483143	2565	0.0053
1990 – 1999	617522	2740	0.0044
2000 – 2009	841470	3269	0.0039

While this is the one given by “naturally”.

<i>Decade</i>	<i>Total articles (T)</i>	<i>Occurrences (N)</i>	<i>Rate ($\frac{N}{T}$)</i>
1940 – 1949	40538	106	0.0026
1950 – 1959	89158	305	0.0034
1960 – 1969	168567	702	0.0041
1970 – 1979	327427	1768	0.0053
1980 – 1989	483143	3172	0.0065
1990 – 1999	617522	5187	0.0083
2000 – 2009	841470	7670	0.0091

¹²[86], p. 1.

¹³For the other formal uses the situation is even less significant for the general picture. “Natural deduction”: (decade) 40-49, (occurrences) 0; 50-59, 15; 60-69, 37; 70-79, 150; 80-89 148; 90-99 295; 00-09, 254. “Natural transformation”: (decade) 40-49, (occurrences) 0; 50-59, 3; 60-69, 112; 70-79, 231; 80-89 171; 90-99 241; 00-09, 283. “Natural isomorphism”: (decade) 40-49, (occurrences) 4; 50-59, 32; 60-69, 49; 70-79, 111; 80-89 113; 90-99 180; 00-09, 177. “Natural topology”: (decade) 40-49, (occurrences) 11; 50-59, 27; 60-69, 73; 70-79, 113; 80-89 143; 90-99 147; 00-09, 195.

Then, if one wants to consider the growth highlighted by the table, to focus on the informal side is quite an Hobson's choice. Moreover, it seems reasonable that a good theory of the informal uses of naturalness - with 'good theory' we mean something fairly different from a conclusive answer to the question: "What is naturalness in mathematics?" - would shed light also on the formal employments. However, speaking of naturalness remains somehow obscure, because this concept lies in a wobbly geography of informal notions. In fact, the dichotomy natural/unnatural overlaps - and maybe gathers - a collection of classical oppositions: pure/artificial; simple/complex; primitive/derivative; general/particular; direct/indirect; easy/difficult; essential/contingent ; intrinsic/extrinsic. Here the boundaries are rough: in most cases there is no particular reason for choosing one of these notions over the others, and often the preference is settled by habit. However, there is something peculiar in the case of naturalness, since the statistical weight of the other notions, and their rates give a fairly different picture. Two interesting examples are that of "simple",

<i>Decade</i>	<i>Total articles (T)</i>	<i>Occurrences (N)</i>	<i>Rate ($\frac{N}{T}$)</i>
1940 – 1949	40538	2688	0.066
1950 – 1959	89158	6128	0.068
1960 – 1969	168567	10379	0.061
1970 – 1979	327427	19380	0.059
1980 – 1989	483143	28408	0.058
1990 – 1999	617522	35032	0.056
2000 – 2009	841470	44648	0.053

and that of "essential",

<i>Decade</i>	<i>Total articles (T)</i>	<i>Occurrences (N)</i>	<i>Rate ($\frac{N}{T}$)</i>
1940 – 1949	40538	439	0.0108
1950 – 1959	89158	935	0.0104
1960 – 1969	168567	1702	0.0100
1970 – 1979	327427	3459	0.0105
1980 – 1989	483143	5258	0.0108
1990 – 1999	617522	6464	0.0104
2000 – 2009	841470	8340	0.0099

where we can see a constant decrease in the use of simple and an alternating trend of essential that end with a substantial diminishing of their use. Our guess is that, in both cases, the reason is a partial semantic erosion by the term “natural”.

Since our starting point was the natural language, with the definition of naturalness we can find in the Oxford dictionary, we could wonder if this trend is a global general tendency of the academic language - and in general of the natural language. However the next graphic shows that this is not the case.

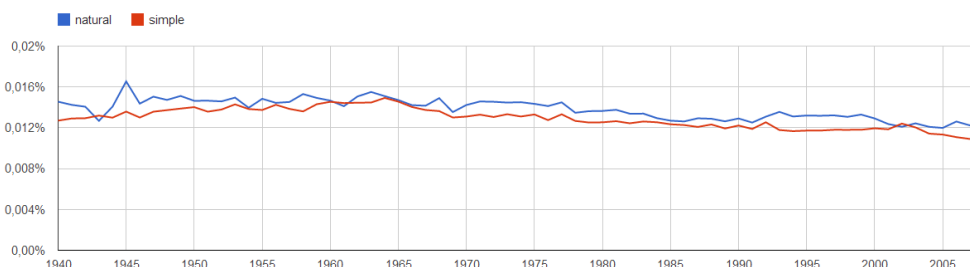


Figure 3.1: graphic Google-ngram

This kind of analysis could in principle proceed with more refined tools - and there are indeed people how did so for the expression “it is easy to see that”¹⁴ - but this approach brings together a radical form of naturalism, that we may call a linguist naturalism, as it has been hinted by Maddy.

Mathematics is a form of human activity, a distinctive linguistic practice, and as such it can be studied like any other such practice [e.g.] by linguistics [...]. Here the naturalist will face questions about the similarities and dissimilarities between mathematical and natural scientific language.¹⁵

However we decide not to pursuit a purely syntactic analysis and to stop here our statistical overview. As a matter of fact, we employed it only in order to point at the problem, but we think that this methodology is not sufficient for its philosophical solution, for three distinct reasons. While explaining our

¹⁴This analysis has been pursued by Lorenz Demey at the ILLC’s Logic Tea, on April 21st, 2009. However his starting point is quite different from ours, because it is in the same path as Corfield’s approach, that tries to avoid the “foundational filter”.

¹⁵[110], p. 453. Of course this is not the end of the story for Maddy’s position. Indeed her naturalism gives much attention also to mathematical practice, but sympathizes with every descriptive philosophical enterprise.

disagreement with a linguistic approach we will clear our position also with respect both to other quantitative styles of argumentation, in the context of an analysis of the informal aspects of mathematical work, and with respect to a naturalist attitude towards the philosophy of mathematics.

1. On the one hand, a mere linguistic description of the occurrences of the term 'natural' is too inclusive, in the sense that it groups together different motivations that lead an author to refer to naturalness. These reasons, may not be conceptual, but also stylistic or idiosyncratic. But when engaged in a linguistic analysis, how can we discern between the relevant and proper use of a term, from the inappropriate ones? This problem is what we may call 'the problem of relevance'. It is not a merely methodological difficulty, because it conceals an important conceptual complication for every inquirer that pretends to start from a philosophically neutral position - as many forms of naturalism aim to be. It is the problem of discerning what is relevant and what is not, in the choice of the data that we are analyzing, without imposing to the bare facts the structure we would like to find within the data. Indeed the recognition of some uses as canonical and others as deviant presupposes a framework where concepts have already been defined and that, in consequence, shapes the results of our analysis. In other words: the answer to a problem is somehow contained in its setting. But is this good philosophy? To what extent can we recognize a statistical discovery as authentic? We do not try to give a full answer to this problem, but we simply point to this difficulty for a linguistic analysis of a philosophical problem. Notice, moreover, that this is a meta-theoretical problem, when engaged in the analysis of the notion of naturalness, because this is exactly the task of the naturalness judgments in a mathematical work.
2. On the other hand a mere linguistic analysis can be seen as too limited, because it misses the implicit uses of naturalness in mathematics. When we undertake the task of understating what does natural mean in mathematics, our goal is mainly semantic and we believe that to presuppose a perfect coincidence between the formal and the informal side of mathematical language, is a too strong philosophical assumption. This aspect has been suggested also by Harvey Friedman in a discussion on this topic on the Foundation Of Mathematics list, in 2006.

One can attempt to formally justify the constant and pervasive use [of naturalness] by taking some major Journals and text-

books, and counting up the number of uses, or counting up the number of *implied uses*.

But again we have to counter the you-find-what-you-look-for objection that comes from considering the problem of relevance. That is, why we cannot count as a solution to our problem the recognition of the criteria for finding the implied uses? The main reason for rejecting such a solution, that we may call a Philosophy First approach, is that the vagueness of the concept and the irregular geography of its semantics, together with its pervasive presence in every field of mathematical research, calls for an analysis of the problem that starts from concrete case studies. Only in this way it is possible to have an objective insight for a possible answer to the question: “what is naturalness in mathematics?” In other words we think that the best framework where to perform a philosophical analysis of this concept, is the philosophy of mathematical practice.

3. Finally we think that a purely linguistic analysis does not explain the reasons of the historical increment of use of the term ‘natural’. As a matter of fact, a perfectly detailed linguistic analysis of the notion of naturalness can give, at best, a tautological description of the phenomenon, that is not able to characterize naturalness if not pointing at what mathematicians call natural. The situation is even more complicated because it is common practice to call something natural with respect to other natural piece of mathematics, without giving a critical account of the notion of naturalness. This is a major hint that a philosophical naturalist approach, that does not debate the internal methodology of mathematics, if not a wrong approach, at least clashes with this mathematical attitude, that seems in need of an investigation able to discuss and clear its methodology. This approach that we may call *implicit* or *recursive* or *dynamic* is well expressed by Chow, in the same thread from where the above quotation by Harvey Friedman was taken.

I would incline towards modeling the space of mathematical statements as something like a graph, with vertices being known theorems and conjectures, and edges representing “similarity” or “relatedness” or some such. Then a statement would be natural if it has high degree and is near the center of a giant connected component, or something like that.

In other words, a statement is likely to be natural if it is similar to many other statements that have been considered before,

and/or if it is conceptually linked with many other natural statements. In contrast, a statement that is easily stated but has a strange form and is not related to other known statements is probably unnatural¹⁶.

This context-dependent and dynamic character of naturalness in mathematics needs to be taken into account, and hints in the direction of the need of some qualitative and global philosophical consideration, for a good description of this phenomenon. Indeed, if the predicate of naturalness is gained in virtue of the connection with other natural pieces of mathematics, it is not the link, but some peculiar quality of the mathematical objects that makes possible this connection. Hence, we do not think that a mere quantitative proposal - the one that, for example, counts the number of edges between two different natural objects - can shed light on such a concept - contrary to what Michel Friedman ([39]) and Philip Kitcher ([92]) tried to do with the notion of explanation. In conclusion we would like to give an answer not only to the question “what is naturalness in mathematics?”, but also, “why mathematicians call a piece of mathematics natural?”.

Then, if we rejected a too strong descriptive naturalism that does not reflect on the internal methodology of mathematical practice, we are left with the following dilemma: we cannot decide from the outset what are the relevant examples, but we need to recognize some specific charter of naturalness in concrete cases?

A philosophical step is needed, in order to start our analysis, that, at the same time, avoids the problem of relevance. This is what we try to do in the next section. Since the beginning we want to say that there is no painless way-out, if not a slightly shift of the problem.

3.2 A tentative methodology

Once we discard an uncritical registration of the use of the term ‘natural’, we reject a Philosophy First approach and we refuse a purely quantitative attitude, where to start our analysis? Since now, we took a stance toward some guidelines of a general naturalist approach, while maintaining that a weak form of naturalism is implicit in the philosophy of mathematical practice methodology. However we still have not excluded a stronger form of naturalism like the one of Penelope

¹⁶Chow, FOM-list on Jan 28, 2006.

Maddy: a position that supports its refutation of philosophical arguments in mathematic, with strong philosophical arguments. We now want to explain our methodology, while distance ourself from Maddy's position, in order to clarify the role that philosophy plays in such an analysis.

The first attempt to find a way out from the dilemma we outlined at the end of the last section could be to look for concrete examples and draw a philosophical moral out of them, without imposing a philosophical prejudice to our scrutiny, but relying only on the mathematical methodology that we encounter in the practice of the mathematical community; this is for example the strategy of Maddy's naturalist philosopher. However, even assuming a consistent and well organized attitude of the mathematical community towards the issue of naturalness, we come upon two problems.

On the one hand we have to accept a methodological naturalism that, in the programmatic absence of any philosophical posture, depends on others' prejudice and runs into the trouble of raising to philosophical maxim the lesson we learn from a finite number of cases, that can never count as a substantial set of examples for a global treatment of the problem. attitude towards mathematics conceptual naturalism. We believe that without a sufficiently clear philosophical framework, it is too ambitious to draw a philosophical lesson from partial examples. Moreover in the case of naturalness in mathematics, we cannot even rely on a sufficiently clear conceptual analysis of this notion from the side of the mathematical community.

On the other hand we have to accept the naturalistic holism implicit in Maddy's work that argues for a strong autonomy of mathematics. In Maddy's work this thesis is interpreted as saying that philosophical considerations do not find place in a mathematical enterprise. Even if she maintains that philosophical consideration can have a role in the inspirational side of the discovery of a theorem - as it is the case for Gödel's realism in the discovery of the coherence of the Continuum Hypothesis - she discards their role in the process of justification of portions of mathematics¹⁷. However, when dealing with a concept like that of naturalness, that manifest a strong informal character, and when the need for its use comes from mathematics itself, are we allowed to discard so easily a philosophical component in the mathematical work? Maddy does not permit any exception.

After uncovering corresponding methodological argumentation in a range of cases, the Second Philosopher concludes that though meta-physical theories on the nature of mathematical truth and existence

¹⁷Cfr. [111], p. 366.

undeniably do turn up in such debates, they are not in fact decisive, they are in fact distractions from the underlying purely mathematical considerations at work. Actual methodological decisions, she sees, are based on a perfectly rational style of means-ends reasoning: the most effective methods available toward the concrete mathematical goals in play are the ones endorsed and adopted. Acting on her assumption that the actual methods of mathematics are the ones that should be followed, she resolves to apply such typically mathematical methodological reasoning to any contemporary debates she might face¹⁸.

But now, the situation is different from the predicative example discussed in the passage before the above quotation. Mathematicians make more and more use of a term with a strong philosophical flavor - as we showed thanks to the definition of the Oxford dictionary - and that pertains to the informal side of mathematical work. Should we then only look for ‘mathematical methodological reasoning’ to address the problem of naturalness, in spite of its intrinsic vague character? We think that the answer is no, and that every Second Philosopher misses the possibility of a real analysis of mathematical practice and of the genuine philosophical problems that arise in it. Indeed we assume the presence of naturalness in mathematics as a surface detector of a much deeper theoretical phenomena, and we acknowledge that this is a major distance from Maddy’s form of naturalism. We believe that we cannot get rid of a philosophical context in the present investigation. We think that our case is even more compelling than those of ‘simplicity’ and ‘fruitfulness’ that motivated the following defense, by Tappenden, of a philosophy of mathematical practice.

The assessments of simplicity or fruitfulness we make would no doubt be different if our brains were wired differently, and this would affect the mathematics and science that we produced, but still the judgements we actually make are too systematically embedded in our actual practices to be simply shrugged off in studies of either scientific or mathematical method¹⁹.

Then we are left with our dilemma. How to escape this impasse? We cannot decide from the outset what are the relevant examples, but we need to recognize some specific charter of a natural piece of mathematics. We admit that there is no way-out of this problem, but at least we are in good company, being this a

¹⁸[111], p. 349.

¹⁹[164], p. 154.

general problem of every philosophical inquire. If we want to avoid any prejudice and we doubt of the feasibility of every argumentative step, how to proceed?

Since it seems to be impossible to proceed in our analysis, let us step back and ask a more fundamental question than ‘what is naturalness in mathematics?’, trying to address the following issue: which form should take a possible answer to the question ‘what is naturalness in mathematics?’. Then we find out that the discussion we did so far is not useless, because we argued in favor of a philosophy of mathematical practice approach to the problem. Once we determine the context, it is surely easier to shape an answer to our original question.

However the context of the philosophy of mathematical practice is still too wide and too vague to help in the solution of our problem. Instead of trying to describe which kind of theoretical features an account of naturalness should have, or avoid to have, in this context - this task would surely be interesting, although far from easy - we could inspect different methodologies, typical of this branch of philosophy of mathematics, looking for hints towards a description of an account of naturalness.

Indeed, on the one hand we believe that the identification of the philosophy of mathematical practice, as the framework of our work, determines the general features of an account of naturalness, while on the other hand we think that the types of approaches that we can find in this context are in part responsible for shaping the answer to a question like “what is x ?”.

Furthermore, in the case of naturalness, this strategy is also motivated by the absence of a literature structured enough to enable us to find general methodological guide-lines for our analysis.

In doing so, two aspects deserve particular attention: first of all, if we import other methodologies from different analyses of mathematical practice, we have to give valid philosophical reasons to explain why we consider them related to the problem of naturalness; secondly, we need to keep with our idea of preserving our point of view as neutral as possible, dismissing the risk of a distortion of the real uses of the term ‘natural’ just for the sake of our argumentative line. The methodological solution to this second concern will be to import into the framework of our inquire of the notion of naturalness a philosophical debate - already well developed - that exists between opposite approaches to a particular problem of the philosophy of the mathematical practice - as we will see: the analysis of the mathematical explanation.

We could then make this argument more sharp and outline a sort of algorithm that can be used to account for a vague notion as that of naturalness in mathematics.

A semi-formal algorithm

(for a philosophical analysis of a mathematical term)

0. Look for empirical evidence of literal uses of the term in the mathematical literature. If its frequency is marginal, then stop. If it calls for an explanation, then proceed. If the term has a common sense, look for its definition in the dictionary.
1. (If necessary) Enforce your analysis with tools from Corpus Linguistics.
2. Find the right philosophical context where to place your analysis and give convincing philosophical reasons to support why it is relevant for understanding the term.
3. Inspect the possible methodologies of the context you found in 2. and look for the philosophical ideas that motivate them.
4. Formulate a (possibly binary) dichotomy, in accordance with the philosophical ideas you found in 3.
5. List historical examples in which the term is involved in some explicit form, or add contexts clearly connected to the dichotomy, where the role of the term is relevant.
6. Test items from 4. using items from 5.
7. Verify plausibility with previous outputs of the algorithm.
8. Connect the horn of the dichotomy, to which the examples point, with the philosophical idea that motivated its proposal.
9. If the term has a common sense, compare your results with the common sense and see it informs the philosophical ideas you found in 3.
10. Go to 1.

Remark 3.2.1. Point 7. is needed in order to allow the possibility that behind the notion of naturalness there is not a philosophically relevant notion. As a matter of fact, even if our goal is to show that there are philosophical ideas in - mostly on the informal side of - the mathematical practice, we do not argue that every informal notion that is extensively used in mathematics is philosophically meaningful and deserve an explanation. But, if there is a consistent notion, then, we can draw a philosophical moral. Then we are in the position to answer the question ‘why mathematicians call something natural?’ explaining the philosophical ideas that motivate this use, and thanks to point 9. explaining

the general intentions that motivate the recognition of natural aspects in the mathematical work.

Remark 3.2.2. Point 10. is used as a means to always come back to the statistical evidence, that, if not a sufficient tool for a philosophical analysis, is nevertheless very useful when facing problems of mathematical practice. For example, if the outcome of an application of this method points in the direction of the similitude between the notion of naturalness and another different one, a statistical analysis can be used to ascertain this fact.

We believe that this algorithm really propose a third way between a too naturalistic approach and a Philosophy-first approach. Against naturalism, we believe that philosophy is needed in order to argue in favor of the right context where to place our analysis. Moreover, instead of following the mathematical practice too closely and uncritically, we propose to inspect the methods we have to analyze it and shrink them down to their philosophical rationale. In doing so, we admit that we need a starting point in any conceptual analysis and that the setting of a problem cannot be wholly philosophically neutral. However, since we want to refute a Philosophy First approach to the problem of naturalness in mathematics, we will inspect different and opposite methodologies, in order to find a binary opposition - as we ask in the algorithm's point 4. - that allows us to keep a balanced point of view. Then it will be the mathematical practice that will speak for itself, tipping the balance in favor of one side or the other of the proposed dichotomy.

This is exactly what we plan to do in the next section, where we will run the proposed algorithm, applying our method of inquire.

3.3 Applying the method

Following our algorithm, what we need now is an appropriate context, conceptually close to that of naturalness, from which we can import a dichotomy of two different features that a naturalness account may exhibit.

This strategy is not new. Indeed, Tappendel proposed to look at how the notion of naturalness has been used in the contemporary metaphysical debate:

It's unlikely that mathematical and non-mathematical reasoning are so disjoint as to exclude interesting points of overlap. In recent decades there has been a revival of old-fashioned metaphysical debates about the reality of universals, the artificial/natural distinction, and cognate topics. It might seem initially promising to draw

on these debates to illuminate the questions appearing in the survey essay²⁰.

Nonetheless, he then shows with convincing arguments that the debate on metaphysical natural properties is not the right context where to find an answer to the question ‘what is naturalness in mathematics?’; even granting the possibility of a coincidence between mathematical and non-mathematical discourse. As it is shown considering Sider’s paper *Naturalness and Arbitrariness* ([155]), often mathematical debates cannot be settled appealing only to metaphysical intuitions, as it is the case for Benacerraf problem of *What numbers could not be*, but, on the contrary, they can find a solution thanks to intra-mathematical, pragmatological reasons, as the reasons for accepting Von Neumann’s identification between sets and numbers show²¹. Then Tappenden hints to a partial coincidence between the notion of naturalness and that of fruitfulness. Thus, he seems to agree with Maddy’s suggestion: to consider only intra-mathematical reasons, without a sufficiently philosophical analysis of the problem. Moreover, we cannot consider this analysis satisfactory, because the arguments in [166] and [165] concentrate on the justification of the naturalness judgments, without trying to address the problem of ‘why mathematicians call something natural?’. This latter question being relevant, once a study of this notion is performed in the context of the philosophy of mathematical practice, as Tappenden seems to admit.

If metaphysics is not the proper context of analysis, where to look? Tappenden himself indicate an alternative route:

It will help us sharpen the issues²² to look for a philosophical niche served up by treatments of explanation and understanding in the natural sciences, since these have been extensively addressed²³.

We propose to take this hint seriously, not only on the ground of a principle of authority, but also for theoretical reasons. There are, indeed, similarities

²⁰[166], p. 3.

²¹See [162] in this respect.

²²In this work Tappenden is not addressing primarily the problem of naturalness, but many problem related to it. This quotation is taken after the presentation of a case study where visualization seems to be a fundamental character of the representation of the multiplication table for octonions. At this point, he is discussing the naturalness of the formulation of a problem, the essentiality of its presentation and its fruitfulness. Then, also considering the relevance that fruitfulness plays, for him, in the context of naturalness - as one sees in [166] -, we believe that this passage is relevant and well placed in this discussion.

²³[164], p. 158.

between the notions of naturalness and that of explanation that call for a similar treatment. First of all, they both pertain to the field of the philosophy of mathematical practice, since they belong to the informal side of a mathematical work; where we can find rational arguments that are not always fully formalized. These two notions help in showing that, even in mathematics, results do not come from nothing, but they conclude a process of discovery, that is only partially formalized in the proof of a theorem. Both naturalness and explanation are able to unveil the presence of elements that not always find a place in the justification of truth, but that enlarge the scientific perspective, considering mathematics as a human activity.

On the other hand, ‘it is natural’ is sometimes used to mean self-explanatory. Then naturalness judgments relieve the necessity of explanation, doing the job in absence of a rational argument. This end-of-the-argumentation character of naturalness is, we believe, an important aspect of this notion and we will see that on this topic, we will diverge from the view of [164].

Then, let us look at what happens in the field of mathematical explanation. We encounter two different and antithetical approach: the so called²⁴ bottom-up and top-down methodologies.

It should be obvious from the above that mathematicians seek explanations. But what form do these explanations take? It is here that two possibilities emerge. One can follow two alternative approaches: top-down or bottom-up. In the former approach one starts with a general model of explanation (perhaps because of its success in the natural sciences) and then tries to see how well it accounts for the practice. In the latter approach one begins by avoiding, as much as possible, any commitment to a particular theoretical/conceptual framework²⁵.

The main methodological opposition that these two alternative approaches evoke is the one between a monistic account versus a pluralistic account. A top-down attitude, as outlined in [117], starts from a general model and argues towards a conformity of the case studies to its standard. The outliers of its analysis are discarded as not pertinent. There is not much room for dissimilarities and this approach has the effect of forcing one to ignore what diverges from its description - following Hegel’s motto “*Desto schlimmer fur die Tatsachen*”²⁶.

²⁴See for example [117], or [126].

²⁵[117], p. 221.

²⁶“So much worse for the facts [if they do not fit the theory]”, attributed to Hegel as an answer to those who noticed that new observations did not fit in the theory formulated in its PhD thesis. See [105] for a reference of this anecdote.

This monistic attitude has a strong static character. The burden of relevance is left on the philosophical arguments that an author presents to defend his theoretical point of view.

On the other hand a bottom-up approach starts from concrete cases and look for insights able to characterize a general notion - as for example the one of explanation - accepting, since the beginning, the possibility that a global and consistent characterization of that notion is not possible. It tries to follow closely the intricate picture of the use of a concept, accepting the possibility that behind a notion there is not only one idea, but a possible cluster of different perspectives. This pluralistic stance, then, aims to recognize the peculiarity of the context where a term is used, and to find a common theme that allows to identify a general pattern. This context-depending attitude then focuses on the dynamic character of a notion. In this case the relevance of an example is given by its concrete character and the main difficulty is that of recognizing similarities in different contexts and that of proposing a coherent philosophical account, able to include divergent outcomes.

Of course there are differences in analyzing naturalness and explanation. As a matter of fact, explanation evokes the idea of a process, while - as we hinted before, outlining the self-explanatory character of a natural property - naturalness point at a more atemporal phenomenon. This dissimilarity can be also found at a more theoretical level, where we acknowledge that explanation pertains more to epistemology, while naturalness to ontology. Recall also the common sense meaning of naturalness that points in the direction of an objective - realistic - character of what is labeled as natural. However, it is also possible to endorse a view of axioms as pattern of proofs, or to account for the naturalness of a proof, but here we want to support the claim that, even granting their similarities, explanation and naturalness are two distinct notions.

Then the theoretical dichotomy we want to test is the one between static and dynamic. This division echoes the two alternative positions expressed by Friedman and Chow on the FOM list, in 2006.

We call *static* view any approach for which naturalness is an inherent and stable property (or class of properties) of the 'object' or 'action' that we call natural - even if it is not possible to characterize it properly.

We call *dynamic* view any approach for which naturalness rests on some contextual properties (as, for example, relational with other natural piece of mathematics - synchronic character of dynamism - or depending to the development of a mathematical theory - diachronic character of dynamism), so that it is not possible to determine what is naturalness, if not appealing to other element of a theoretical picture.

The oscillation between a dynamic and static picture of an account of natu-

ralness is evident also in the treatment that Corfield deserves to the analysis of this notion in his book *Towards a philosophy or real mathematics*, where he acknowledges that its use hides the possibility of different meanings - thus favoring a bottom-up approach, as one would expect from his naturalism.

In sum, a full analysis of the use of the term ‘natural’ by mathematicians through the ages would require a book-length treatment. As used today it possesses several shades of meaning, which blend into each other to some extent, relying as they do on a sense of freedom from arbitrariness and artificiality²⁷.

However, in the pages preceding this observation, Corfield describes the practice of calling a mathematical object natural, whenever its features are compatible with the characters of the class to which the object belongs. This style of argumentation is similar to the context-depending appeal to naturalness we discussed in relation to Chow’s quotation and it can be easily seen to pertain to a dynamic view. On the contrary, discussing the naturalness of the concept of groupoid, Corfield writes that: “[a]nother way of arguing for the naturalness of a concept is in terms of the inevitability of its discovery²⁸”. This latter position is clearly connected with a more static view of naturalness for which it is relevant that different people independently converge on the discovery of a concept - and not its definition.

These examples are meant to show that the dichotomy we are proposing fits very well with the different approaches that it is possible to have toward the notion of naturalness. Hence, we take this as an hint that we are on the right path. Then the question that we will now try to answer, thanks to the case studies, is the following: is naturalness a static or a dynamic notion?

3.4 Case studies

We will now discuss a couple of examples taken from the field of set theory. The motivation to consider this domain comes from the fact that the philosophical reflection we will proposed springs not only from a general knowledge of the subject, but also by taking part in the working scientific community and so by the awareness of the methods of reasoning of contemporary set theory. As a matter of fact, we believe that only being involved in the research of a field it is possible to have first hand knowledge of all the unwritten aspects of a mathematical

²⁷[22], p. 230.

²⁸[22], p. 225.

work. This is of primary importance, if one aims to give a philosophical account of a mathematical phenomenon. On this aspect we agree with Quine that “He [the philosopher] tries to improve, clarify and understand the system [science] from within. He is the busy sailor adrift on Neurath’s boat²⁹”. We believe that the informal side of formal sciences cannot be disregarded and, on the contrary, it occupies a central role in any attempt to understand the mathematical work. Indeed we think that neither the picture of a philosopher who follows too closely the mathematical work, without giving any philosophical advice, and without getting his hands dirty, is useful, nor the picture of a naïve mathematician who only follows internal pragmatic reasons for the development of science, without having a more complex opinion, and without the need for a philosophical picture, is fair.

We will then concentrate our attention on the naturalness of the concept of set and on the naturalness of new axioms in set theory.

3.4.1 The concept of set

We start with the literal evidence that naturalness is normally assumed as a property of the concept of set.

Faced with the inconsistency of naive set theory, one might come to believe that any decision to adopt a system of axioms about sets would be arbitrary in that no explanation could be given why the particular system adopted had any greater claim to describe what we conceive sets and the membership relation to be like than some other system, perhaps incompatible with the one chosen. One might think that no answer could be given to the question: why adopt this particular system rather than that or this other one? One might suppose that any apparently consistent theory of sets would have to be unnatural in some way or fragmentary, and that, if consistent, its consistency would be due to certain provisions that were laid down for the express purpose of avoiding the paradoxes that show naive set theory inconsistent, but that lack any independent motivation. One might imagine all this; but there is another view of sets: the iterative conception of set, as it is sometimes called, which often strikes people as entirely *natural*³⁰, free from artificiality, not at all ad hoc, and one they might perhaps have formulated themselves³¹.

²⁹[144], p. 72.

³⁰My italics.

³¹[18] p.218.

We see, in this quotation by George Boolos, a clear presentation of the so called *iterative conception*, influentially discussed also by Parsons ([135]) and Wang ([186]). This conception stems from the idea of a cumulative hierarchy for the universe of set theory and it is here linked directly to the problem of naturalness. Then following our method we should ask: is the naturalness of the notion of set a static or a dynamic one?

As it is clear from Boolos' quotation, and well known from the history of the discipline, the notion of set changes through history and thus it is a dynamic one. Indeed its first appearance in the history of mathematics was shown to be inconsistent, and moreover it was linked with a different idea, widespread and common in the mathematical community at that time: that of a set as determined by a law. We will now analyze these two notions of sets - the original Cantorian one and the iterative one - trying to understand if the dynamic character of this concept determines a similar aspect of the notion of naturalness.

In Cantor's work, the first definition of set is in 1882, in the third paper of the series of six from the period 1878-1884, bearing the title *Über unendliche, lineare Punktmannichfaltigkeiten*.

I call a manifold (an aggregate [Inbegriff], a set) of elements, which belong to any conceptual sphere, well-defined, if on the basis of its definition and in consequence of the logical principle of excluded middle, it must be recognized that it is internally determined whether an arbitrary object of this conceptual sphere belongs to the manifold or not, and also, whether two objects in the set, in spite of formal differences in the manner in which they are given, are equal or not. In general the relevant distinctions cannot in practice be made with certainty and exactness by the capabilities or methods presently available. But that is not of any concern. The only concern is the internal determination from which in concrete cases, where it is required, an actual (external) determination is to be developed by means of a perfection of resources³².

However, the first relevant one for a conscious history of set theory is the one in the *Grundlagen einer allgemeinen Mannigfaltigkeitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen*, from 1883.

By a 'manifold' or 'set' I understand any multiplicity which can be thought of as one, i.e. any aggregate [Inbegriff] of determinate elements which can be united into a whole by some law³³.

³²[191], p. 150.

³³[191], p. 204.

The idea of a set as a mathematical object determined by a law explains the reason why set theory is commonly considered as a part of logic - where logic is intended to be the general science of the law of thought. As a matter of fact we find here expounded the notion of set as extension of a concept, that, properly formalized, will bring Frege to the failure of its logicistic program. However, Cantor's idea of general law - naïve as it may be - is not limited to some repertory of tools of definition, but it seems opened to any possible - and future - means.

This same idea of sets as concept-extension is also what guided Dedekind in his work on the foundation of number theory: *Was sind und was sollen die Zahlen?* - that will influence Zermelo, together with Cantor's work, in the axiomatization of set theory.

It very frequently happens that different things a, b, c . . . considered for any reason under a common point of view, are collected together in the mind, and one then says that they form a system S; one calls the things a, b, c . . . the elements of the system S, they are contained in S; conversely, S consists of these elements. Such a system S (or a collection, a manifold, a totality), as an object of our thought, is likewise a thing; it is completely determined when, for every thing, it is determined whether it is an element of S or not³⁴.

Then, the question we should ask is: was the first conception of set thought as natural? Cantor had the idea that his notion of set was instrumental for the development of his theories of ordinal number and infinite cardinal numbers. Then, in trying to justify the former he says that the extension from the finite to the infinite was natural and helped him to develop set theory.

I am so dependent on this extension of the number concept that without it I should be unable to take the smallest step forward in the theory of sets [*Mengen*]; this circumstance is the justification (or, if need be, the apology) for the fact that I introduce seemingly exotic ideas into my work. For what is at stake is the extension or continuation of the sequence of integers into the infinite; and daring though this step may seem, I can nevertheless express, not only the hope, but the firm conviction that with time this extension will have to be regarded as thoroughly simple, proper, and *natural*³⁵.

³⁴[25], p. 344.

³⁵My emphasis. In *Grundlagen einer allgemeinen Mannigfaltigkeitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen*, see [25] p. 883.

Moreover Cantor, talking about the properties and laws of the infinite, says that they depends “on the nature of things³⁶”³⁷. But then, how it is possible that a natural notion was transformed into an other different natural notion?

The change in the notion of set comes from Zermelo’s axiomatization in 1908, where the explicit attempt was to keep as ample as possible the concept of set, without running into the paradoxes.

This discipline [set theory] seems to be threatened by certain contradictions, or ”antinomies”, that can be derived from its principle - principles necessarily governing our thinking, it seems - and to which no entirely satisfactory solution has yet been found. In particular, in view of the “Russell antinomy” of the set of all sets that do not contain themselves as elements, it no longer seems admissible today to assign to an arbitrary logically definable notion a set, or class, as its extension. Cantor’s original definition of a set (1895) therefore certainly requires some restrictions; it has not, however, been successfully replaced by one that is just as simple and does not give rise to such reservations. Under these circumstances there is at this point nothing left for us to do but to proceed in the opposite direction and, starting from set theory as it is historically given, to seek out the principles required for establishing the foundations of this mathematical discipline. In solving the problem we must, on the other hand, restrict these principles sufficiently to exclude all contradictions and, on the other, take them sufficiently wide to retain all this valuable in this theory³⁸.

Few comments are needed after this quotation. Zermelo says explicitly that he wants to axiomatize the “theory created by Cantor and Dedekind”, but he likewise explicitly says that the theoretical framework that motivated the founding fathers is not tenable anymore because of the antinomies. The problem is found exactly in the main definition of set, that came so naturally from Cantor’s

³⁶[191], p. 371-372.

³⁷By the way, this opinion may be questioned by the modern development of set theory. Indeed it is important to stress that Cantor’s theory of cardinals is not as “natural” as it could be seen; as a matter of fact it hides an important choice behind it. There are two conflicting ideas: Cantor’s Principle - two sets have the same size if there is a bijection between them - and Aristotle’s Principle - if a set A is a proper subset of another set B , then the size of A is smaller than the size of B . As the development of a theory of numerosity has shown ([16], [15]) the formalization of the infinite does not involve necessarily Cantor’s theory of cardinal numbers

³⁸E. Zermelo, *Investigations in the foundations of set theory I* in *From Frege to Gödel*, Harvard University Press, 1967, pp 199-215

analysis of well-order sets and infinite cardinalities: a too loose use of the idea of sets as concepts extension is dangerous. Then Zermelo's proposal, in the line of Hilbert's school, is to start from an historically given theory and try to arrange its main theorems in a logical order, while implicitly defining the basic notion of the theory. This style of reasoning is very far from Cantor's deduction - in a kantian sense - of the principles of set theory, as he attempted to do in a letter to Hilbert, dated 10 october 1898.

It should also be noted that Zermelo does not appeal to the naturalness of the concept of set as defined by his axioms. Indeed his system is not justified in terms of the concepts involved - even less in terms of Cantor's notion of set - but motivated by pragmatic reasons, with the explicit goal to avoid paradoxes. Then, in 1930, Zermelo, while engaged in the search for a consistency proof for set theory, proved a quasi-categoricity theorem for second order ZF. The context of Zermelo's work is quite far from our modern treatment of the subject³⁹, but the main idea, making use of Von Neumann's contributions to set theory, was to shape a model of ZF thanks to a cumulative hierarchy: a division in levels where the elements of a set lay in levels of the hierarchy that come before the one the set belongs to. These stages were ordered by ordinal numbers and the first level that formed a model for all ZF was indexed by a strong inaccessible cardinal. Subsequently the adoption of the idea of a cumulative hierarchy by Gödel in his proof of the coherence of the Axiom of Choice - where he developed the Constructible Universe - helped in spreading the idea that "set" is an iterative notion.

This concept of set (...) according to which a set is anything obtainable from the integers (or some other well-defined objects) by iterated application of the operation "*set of*"⁴⁰ and not something obtained by dividing the totality of all existing things into two categories, has never led to any antitomy whatsoever; that is, the perfectly "naïve" and uncritical working with this concept of set has so far proved completely self-consistent⁴¹.

From that moment on we could see a progressive shift from the idea of a cumulative hierarchy, for a model of ZF, to an iterative notion for the concept

³⁹Zermelo's work was in the in the context of second order logic and moreover he thought that the definition of a model of set theory had two degree of freedom: height and width - with respect of the *urelemente* to be considered as primitive. While the former stems from the idea of a cumulative hierarchy - and then it is still actual - the latter is not anymore a concern for the mainstream modern research in set theory, that abandoned a theory of sets with *urelemente*.

⁴⁰My italics.

⁴¹Gödel, CW II p.180, 1947 what is the continuum problem.

of set. Indeed, this idea became so linked with the concept of set that people started to inverse the process that lead from ZF to the cumulative hierarchy, trying to justify the axioms in terms of an iterative notion: the conceptual counterpart of the structural, model theoretic conception of a cumulative hierarchy. This is exactly the case of Boolos's arguments in favor of the naturalness of the axioms of ZF. Moreover, notice that this argument can be proposed only after it was possible to give a clear and intuitive picture of the theory that formalizes the notion of set. In this way the axioms that inspired and shaped Zermelo's model(s), in the search of their consistency, are justified in terms of the model(s) itself; but what does this mean, really? It is important to remember that Zermelo's theorem is a quasi-categoricity theorem: it says that a model of second order ZF has just two degrees of freedom, its height and the width, depending on the *urelemente*. Then, since all the possible models of these axioms are build as a cumulative hierarchy, it would seem that there was no need to justify the axioms in terms of their iterative character. To makes sense of this operation, we have to accept that what needs a justification is not the fact that these are axioms for set theory, but the fact that they capture the essence of the concept of set. What is at work here is a hidden thesis that fixes a concept. We could call it the Zermelo-Gödel Thesis: being a set means being an object that belongs to a cumulative hierarchy - and, after Zermelo quasi categoricity theorem: being a set means to be a set in a model of ZF. As in the case of Church-Turing Thesis (CTT) what seems to be the natural choice is, in reality, the stipulation of a relevant aspect of a concept. Then arguments as Boolos' or Parsons' or Wang's are at par with the attempt to prove or justify CTT⁴².

Then we can conclude that the dynamic character of the notion of set informs the notion of naturalness, shaping the latter with a dynamic component - in the diachronic sense we proposed. Moreover, our analysis also showed that a normative component is hidden in the natural character of a mathematical concept⁴³.

⁴²In this discussion we implicitly assumed that the cantorial notion of set, at least the one proposed in the *Grundlagen*, is different from the iterative one. For what concerns the strongest claim that it is not possible to find this notion in Cantor's work we do not take a stand, even if we believe that even the definition presented in the *Beiträge* cannot be considered as cumulative, if not forcing it from our modern perspective. See [34] and [87] in this respect. However, it is fare to say that the iterative conception is not entirely incompatible with the latest reflections of Cantor, even if we believe that it had different conceptual motivations, as it is well shown in [50]. The main possibility of a specification of Cantor's notion of set in terms of an iterative conception does, indeed, sustain our thesis of the prescriptive character of the notion of naturalness.

⁴³Notice that this opinion was proposed quite early, in the development of set theory, contrary to the general idea of a naturalness of the notion of set - as this quote from König shows clearly:

3.4.2 Natural new axioms for ZFC

A second easier example is the discussion on the naturalness of the axioms that extend ZFC. For this case we start from a quotation by Joan Bagaria, who gave a mathematical characterization of the bounded forcing axioms in terms of generic absoluteness ([8]) and then tried to argue for their naturalness ([9]).

All together, the criteria [*Maximality, Fairness, Consistency and Success*] may be regarded as an attempt to define what being a natural axiom of Set Theory actually means⁴⁴.

Let us analyze these criteria, in search for static or dynamic elements in the characterization of naturalness in the context of new axioms for ZFC.

Maximality. This principle is considered useless in absence of further specifications⁴⁵. Then it is exemplified with some of the criteria proposed by Gödel - Reflection, Extensionalization, and Uniformity - whose program is presented as the program “of finding new natural axioms which, added to the ZFC axioms, would settle the continuum problem”. After discussing the issues related to these principles - and acknowledging that the criterion of maximality is not sufficient, alone, to settle CH - Bagaria concludes saying that “Gödel’s principles of Reflection, Extensionalization, and Uniformity arise naturally from the systematic application of the criterion of Maximality⁴⁶.”

Fairness. This criterion is explained as advising his promoters not to discriminate between sentences of the same complexity. Then, the reasons for considering classes of sentences pertaining to sets with the same rank, or to sets with the same hereditary cardinality, is the following: “Now the complexity of a set may be defined in different ways, but the most

“That the word ‘set’ is being used indiscriminately for completely different notions and that this is the source of the apparent paradoxes of this young branch of science, that, moreover, set theory itself can no more dispense with axiomatic assumptions than can any other exact science and that these assumptions, just as in other disciplines, are subject to a certain arbitrariness, even if they lie much deeper here - I do not want to represent any of this as something new.” [55] p. 147.

⁴⁴[9], p. 6.

⁴⁵Notice that also Maddy says something similar: “In both cases, the structure of the counterexamples suggests that the formal criterion will need supplementation by informal considerations of a broader character.” in [109], p. 255. These supplementations are comments like: “This last [$AD^{L(\mathbb{R})}$] is a particularly natural hypothesis, stating that AD is true in the smallest model of ZF containing all ordinals and all reals.”, p. 226.

⁴⁶[9], p. 9.

natural measures of the complexity of a set are its rank and its hereditary cardinality.⁴⁷”

Success. This criterion is easily explained in term of solutions to natural problems. “ A new axiom should not only be natural, but it should also be useful. Now, usefulness may be measured in different ways, but a useful new axiom must be able at least to decide some natural questions left undecided by ZFC. If, in addition, the new axiom provides a clearer picture of the set-theoretic universe, or sheds new light into obscure areas, or provides new simpler proofs of known results, then all the better.⁴⁸”

For what concern consistency, this principle is explicitly considered as a regulative idea that acts only as a necessary condition for new axioms. As a matter of fact, once we are in the context of classical first order logic, this principle can be subsumed under the one of success, because if an axiom is not consistent it allows the proof of every proposition. Hence it is not useful⁴⁹.

In the light of these consideration it is clear that the definition of a natural axiom is not static, but dynamical - in the synchronic sense we proposed - because it depends on the context, on other attempts to define naturalness, on natural ways to consider sentences of the theory, and on the naturalness of other pieces of mathematics.

However, how to make sense of an attempt to *define* - as Bagaria argues - naturalness in terms of naturalness, granting that our goal is to understand the meaning of this notion and the reasons of its use? The appeal to ‘natural questions’ and ‘natural measures’ is sustained by qualitative judgments, on the subject matter of the theory, that pertain to considerations of relevance and of importance, that far from being objective and necessary, gain strength in connection to other naturalness considerations. Here again we find at work normative judgments that stems from subjective or intra-subjective - read scientific community - considerations that aim to shape mathematical work, pointing to what is relevant and what deserves attention and commitment.

3.4.3 Dynamism and the missing ingredient: normativism

As an outcome of our historical examples one may come to the formulation of the thesis according to which naturalness of a mathematical object depends on the sum of all the contextual relations that an object has with respect to other

⁴⁷[9], p. 9.

⁴⁸[9], p. 10.

⁴⁹However nothing is said for what concerns the fact that consistency points in a different direction than maximality and fairness, as long as it limits the set of probable sentences.

pieces of mathematics. This definition is not completely new. It recalls Chow's image of a graph of similarities, that we quoted in the methodological part of this chapter.

Nonetheless, such a solution is at the same time persuasive and frustrating. Indeed, this kind of graph has to admit among its edges a gigantic class of different notions of similarity. But if mathematical ones are extremely hard to classify, most of the other ones are simply too vague to be embedded in a coherent model. Moreover, to establish how these multiple notions of similarities interact with each other is by no means clear. To solve this complication, a possible hint may be found in this comment by Shoenfield:

There is one common feature of the above uses of natural: the assumption that whatever natural means, what is natural is good. I see some signs of this assumption in some of the communications to FOM, which seem to argue that if there is no natural intermediate r.e. degree, then there is something wrong or deficient in the study of r.e. degrees. To justify such a claim, it is not merely necessary to explain the meaning of natural; it is necessary to explain why lack of this type of naturalness is a deficiency in a theory⁵⁰

We assist, here, to a shift analogue to something we already assist to. In our methodology, while debating which kind of tools could prevent preserved us from both naturalism and Philosophy-first approaches, we proposed a subtle but significant change to our main question: from "What is naturalness?" to "What is naturalness for?". Or, rather, "Why mathematicians make use of naturalness?". Shipman's formulation of this very same question can be stated as follows: "Why do we ask for naturalness in our mathematical theories?"

Brought to the extreme conclusion, Shipman's suggestion, deputed of its provocative aspects, is very poignant. In this perspective, naturalness might be considered as a request posed to our theories, rather than an attribute of the elements of the theories themselves. As we can see, even in the presence of our graph, judgements on naturalness are, most of all, requests about the ways in which we do mathematics⁵¹. Consequently, a graph of similarities in which the edges are stably given would be deprived of a crucial element for the comprehension of naturalness, namely normativism.

⁵⁰Shoenfield, FOM list: November 3, 1999.

⁵¹This aspect of naturalness is also linked with a common-sense meaning of naturalness that points to our habits and the familiarity we have, in this context, with some pieces of mathematics. As we will see in the last section of this chapter this temptation to reduce unfamiliar to familiar aspects of our mathematical work goes hand in hand with an even stronger attitude towards mathematics.

It is now clear that our proposal is precisely to consider the normative aspects as decisive. If we admit that calling a portion of mathematics or a mathematical object natural is an operation that includes an element of normativity in itself, then any general request for a criterion to distinguish, once for good, between natural and unnatural characteristics would become meaningless. Thus, even though some very local border of the geography given by our graph comes to be persuasive and accepted enough to be considered stable - recall what we said about natural numbers in the introduction - most of the judgements of naturalness are still the outcome of a continuous process of negotiation; a negotiation that takes place fully inside the practice. Thus, this geography is, essentially, in movement.

Therefore, to call a piece of mathematical discourse natural is a normative operation that consists in connecting it with something already labelled as natural. Whenever this process becomes critical, we have a crisis of naturalness. For naturalness is not to be intended as a specific attribute of our mathematical objects, but rather as an issue about our mathematical practice. Then, these crises can not be solved thanks to the elaboration of a 'stable' theoretical frame. If we want to fix these contradictions, then, we have to go back to the practice.

In conclusion, we suggest that naturalness should be considered as a device of self-regulation within mathematical practice, a device that through a dynamic and communitarian process informs us of the ways in which we want this practice to be performed. At this point, one may believe that we are fallen back into a form of naturalism, for which a mathematical problem, such as the one of naturalness, can not be solved in a philosophical dimension. But it would be a wrong interpretation, since what we tried to show, with the particular analysis of naturalness, is instead precisely that mathematical practice is continuously exposed to philosophical issues, that address it and shape it. This latter aspect is what we want to elucidate in the conclusion of this chapter.

3.5 Philosophy in mathematical clothing

Testing our case studies on the dichotomy statical-dynamical we found that the notion of naturalness has mostly a dynamical character. Moreover, and this is a byproduct of our analysis, we discovered a normative component in the naturalness judgments, that hides the adoption of a criterion of relevance, for conceptual reasons that are not entirely formalized. A piece of mathematics is natural when it fits with a background idea that is chosen to be relevant. In the case of the concept of set, the background ideas was, first, the logical principle of comprehension, for Cantor and Dedekind, and then Zermelo's quasi-categoricity

theorem in terms of a cumulative hierarchy, for Gödel and Boolos. In the case of Bagaria's criteria for natural axioms, the naturalness of the bounded forcing axioms is given with respect to background ideas that inform set-theoretical practice and that determine the relevant problems, focusing on the on-going research.

We think that it is now time to back to the beginning of our analysis and confront what we discovered, thanks to our case studies, with the sense of naturalness as "Existing in or derived from nature; not made or caused by humankind", that we found in the Oxford dictionary. How to fit the image of the use of the word we found in the mathematical practice with its common sense? To rephrase this question in a clearer way, how it is possible that the search for the right definition, the correct method and appropriate concept are justified with reference to the stability and objectivity of a good description of the nature of the mathematical realm, while hiding a normative component and a dynamical character? The two sides of this problem apparently push in opposite directions and seem hard to reconcile.

In trying to give an answer to the problem of "why mathematicians make use of naturalness?" we could believe that we are facing a why-question that is typical of an account of explanation in mathematics. The reason being, at an intuitive level, that naturalness do have an explanatory virtue: that of self-explanation. Then, the appeal to naturalness seems to get to the bedrock of a scientific explanation. This aspect suits perfectly the idea that a piece of mathematics may be called natural in virtue of its evident and clear description of the state of affairs of the mathematical world, but not with the normative character we discovered, whose justification may be the result of a rational process. Also Jamie Tappenden - in [164] - discussing the choice of an axiom system, defends the thesis that the notion of naturalness is the outcome of a reflection on the success and fruitfulness of the axiomatic choices. We agree with him that a prescriptive judgment has a rational basis and that fruitfulness is, of course, one of the elements that a mathematician can use to assess naturalness, but we believe that the recognition of fruitfulness is not the main goal of calling something natural. Tappenden's analysis on the one hand does not explain the dynamical - recursive - aspects of naturalness in mathematics, and on the other fails to account for the divergence of the use of this notion in mathematics, from the one that pertains to common sense.

The solution to this apparent incompatibility is to be found, we believe, in the philosophical goal that motivates the recognition of naturalness in mathematics. This goal is orthogonal to the features we described. Indeed, the dynamic and prescriptive characters of naturalness point in the direction, as we hinted before, of a non-evident character of the notion. In other words, when something is said

to be natural, the reasons for it are to be found in rational arguments, that do not gain their strength by the inevitability of a definition, or by the clearness of a concept, but on the ground of the context where the appeal to naturalness arise. Of course success and fruitfulness play a role, but to put in Cantor's words: "every mathematical concept carries within itself the necessary corrective: if it is fruitless or unsuited to its purpose, then that appears very soon through its uselessness and it will be abandoned for lack of success."⁵² So which are the rational arguments, besides success, and the philosophical reasons that occur in the widespread use of the notion of naturalness?

We believe that answering this question may reconcile the opposition between mathematical naturalness and common-sense naturalness.

The common-sense reference to the order of nature, in contexts where it is not evident which is the nature of the objects involved, hides, in its normative aspect, the choice for a realist point of view in philosophy of mathematics. The appeal to the common-sense meaning of naturalness provides stability to definitions or concepts that, in principle, are not so. In the context of a departure of abstract mathematics from the evidence of the common sense, and due to the lack of an intuition able to ground the stability of mathematical notions, the mathematician tries to pull back its abstract knowledge to some more concrete ways of thinking. The latter ways are found in the analogy with the real world: a causal world, where things have a stable and objective existence, and where, like in physics, science have a normative task. Then, to sharpen the tentative definition we gave at the beginning of this section: a piece of mathematics is natural when it fits with a background idea that was chosen as relevant, in the context of a realist perspective.

These philosophical ideas, far from being metaphysical non-sense, inform mathematical practice. They show that a wide conceptual framework is essential to mathematics and they call for a global approach in order to understand it.

⁵²[191], p. 206.

Chapter 4

The case of Forcing Axioms

In this chapter we resume our arguments so far and we come back to the problem that motivated this work, trying to see if we gained some insight towards its solution. The inspiring question was: which are the reasons for accepting a new axiom in set theory? Before taking up the understating of the problem we collected so far, let us clear from the outset that we want to make the former question more precise, testing our considerations on the Forcing Axioms.

Our analysis started from Hilbert's conception of axiom, seen as the first conscious instance of this notion in a modern axiomatic setting. The focus on the Completeness Axiom has outlined its *sufficient* character, with respect to 1) a collection of other axioms, and 2) a set of proposition which are the subject of the axiomatization. This same character of sufficiency has been found for the Forcing Axioms, in particular MM^{+++} , thanks to the Viale's theorem ([182]) presented in the second chapter, with respect to 1) $\text{ZFC} +$ large cardinals, and 2) the theory of $H(\aleph_2)$.

Moreover, the study of the Axiom of Completeness singled out two properties of the sufficient axioms. The first one is *Completeness* with respect to an intuitive theory, able to 1) prove all the relevant fact of that theory, and 2) close the domain of the theory with all the allowed methods of definition of new object; i.e. closing the theory with respect to the method of ideal elements. The second one was *Prescriptiveness* of the subject matter of the intuitive domain. For what concerns Forcing Axioms, their Completeness was shown comes by Viale's theorem, that shows that it is possible to prove all relevant facts about the structure of $H(\aleph_2)$ - except Gödel's sentences and the consistency of the theory. Thus it is possible to define what does it mean to be true in $H(\aleph_2)$ in terms of forcability with a stationary preserving forcing. Thus, this latter aspect amounts in the prescriptive character of the Forcing Axiom. Moreover, the

completeness was argued also by asking closure with respects to ideal elements, in a consistent way. This is indeed the intuitive content of the Forcing Axiom: given a notion of forcing and a collection of dense sets, there is a filter generic for all these dense sets. In other word, if it was possible to define a new object in a generic extension, thanks to a generic filter that intersects the relevant dense sets, then this object can be already found in the ground model. Then if we define a set in $H(\aleph_2)$ as being an object whose existence can be proved by means of the axioms of ZFC + large cardinals, together with all the accepted methods of definition - and we recognize that the method of forcing is a method for defining new ideal sets that live in the universe outside the model we are considering - we can rephrase MM^{+++} as follows.

Definition 4.0.1. (*Completeness for $H(\aleph_2)$*). *The sets in $H(\aleph_2)$ form a system of things that, compatibly with the other axioms ZFC + “There is a class many super huge cardinals”, can not be extended; i.e. it is not possible to add to the system of sets in $H(\aleph_2)$ another system of things in such a way that in the resulting system all the axioms ZFC + “There is a class many super huge cardinals” are satisfied.*

The meaning of extension here is to be intended literally as adding new mathematical objects by means of a stationary set preserving forcing, that preserves MM^{+++} . Then if we limit ourself to all methods of definition known - what else could we do instead? - we can say that the rationale behind the Forcing Axiom and in particular MM^{+++} can be considered complete in Hilbert’s sense.

A couple of philosophical considerations are needed now, before dealing with the third criterion of *Unification*. The shift from considering the method of forcing as a tools for independence proof - or as a tool for proving theorems, as it is the case for Viale’s result - to the proposal of granting that forcing is a method of definition of new “ideal objects” has a philosophical price: that of considering objects prior to their domain of definition, or existence. Is this realist stand acceptable in the context of the notion of set exemplified by contemporary axiomatic set theory and for what concerns the methodology of the set-theoretical inquiry?

4.1 A realist stand

Let us try, first, to answer the second part of this concern. By Cohen’s theorem we know that given a countable transitive model, we can force over it and add new sets that lay in the real universe of sets V . Thus, granting that a collection of things is a set laying in V , we do not have a priority of mathematical objects

over the domain of all sets. However, this is not how the method of forcing is usually applied in practice. As a matter of fact, normally the arguments run as follows: take a poset \mathbb{P} with some particular property and force with it over V . Then, the implicit assumption is that outside the universe of all possible collections, there are ideal objects-sets that can be defined and added. This realist point of view - i.e. sets, as long as mathematical objects, exist independently of their universe of existence - is particularly relevant for those position that accept that there are many different universes of set theory, as, for example, the multiverse view proposed by Hamkins. Indeed, such a position appears to be far more platonic than a position that takes a realist stand toward the truth value of every mathematical problem, and it should offer a definition of set independent from the definition of a model for set theory. However, this problem is not our concern here. The point we want to stress is the presence in the set-theoretic practice - both in the multiverse view side and in the search for the right universe of sets - of a realist stand that accepts that sets exist prior to, and independent from, their definitions.

For what concerns the notion of set, let us come back to its different definitions, between Cantor's and Zermelo's, in order to understand if the priority of the objects - over domains - is compatible with the contemporary axiomatic presentation of set theory. It is important to notice that in the first of the two Cantor's definitions, from 1882, there is reference to a "conceptual sphere", that disappears in the subsequent definitions. Moreover, in Cantor's mature work there is no reference to the universe of sets because it is the context of discourse, outside which nothing can exist. Then we can say that in Cantor's conception the notion of set depends on its domain of existence; local, before, and global, after. On the other hand, in 1908, Zermelo begins its axiomatization of set theory, in line with Hilbert's school, saying that "Set theory is concerned with a *domain* \mathcal{B} of individuals, which we shall call simply *objects* and among which are the *sets*¹." The difference between Cantor's late definition and Zermelo's one, from 1908, is to be found not only, as we explained in the last chapter, in the absence, in 1908, of a concept-extension notion of set, but also in a local conception of set theory - Zermelo - instead of global one - Cantor. The notion of domain will remain constant in Zermelo's reflection and will culminate in 1930 with the work *On boundary numbers and domains of sets* where a quasi-categoricity theorem for ZF will be proved. In this article, where it can be found the birth of a modern approach to the universe of sets in terms of a cumulative hierarchy, the difference between sets and domains became less marked, because: "every categorically determined domain can also be conceived of as a 'set' in some way or

¹[55], p. 201.

another²”. Then, together with the consequent change of the underlying notion of set, the difference between sets and domains became a matter of context. As a consequence, the possibility that sets may exist outside a domain is granted by the indefinite possibility to extend a domain, turning it into a set³. The key ingredients, in Zermelo’s quasi-categoricity result, are the Power set axiom and the class of ordinals. But, even if a domain can become a set, the notion of all possible subsets is not relativized, nor limited, because - as Zermelo says - a domain contains both objects and sets.

The role of the powerset axiom in the axiomatization of set theory is also a major difference between Cantor’s and Zermelo’s approach. Moreover, thanks to the solutions given to the conceptual problems related to this topic, it is possible to outline a theory of sets that makes the priority of objects over definable sets its distinctive character.

It is interesting to note that although the powerset operation is the backbone of Zermelo’s model for set theory, it does not find a major place in Cantor’s theory. In a letter to Hilbert, as late as 1898, Cantor showed that some basic principles of set theory could derive from the definition of set - clearly against Hilbert’s conception of axioms. Among them there was the powerset axiom, but few days later Cantor wrote again to Hilbert explaining him that his proof was incorrect, because the elements of the multiplicity of all subsets, of a given set, partially overlap and, in order to prove the existence of such a multiplicity, it is necessary to establish the premise of the separation and the independent existence of these elements.

Unter Bezugnahme auf mein Schreiben v. 10^{ten}, stellt sich bei genauerer Erwägung heraus, dass der Beweis des Satzes IV keineswegs so leicht geht. Der Umstand, dass die *Elemente* der “Vielheit aller *Theilmengen* einer fertigen Menge” sich theilweise decken, macht ihn illusorisch. In die Definition der fert. Menge wird die Voraussetzung des *Getrenntseins* resp. *Unabhängigseins* der Elemente als *wesentlich* aufzunehmen sein⁴.

As we see, the power set operation is, for Cantor, problematic, exactly because it is not clear what is the set of all the subsets of a given set. In other words, following Cantor’s doubt: why should we be allowed to consider the subsets of a set as existent prior to the existence of their collection - i.e. without

²[192], p. 429.

³Notice that we are claiming that, in the cumulative hierarchy context, there is a priority of the notion of objects-sets, over domains.

⁴Letter from Cantor to Hilbert, October 12th 1898, in [133], p. 398.

appealing to the power set operation itself - if we cannot even distinguish one from the other? This concern hides the worry that it may be a too strong assumption to give for granted the existence of all the subsets of a sets, before knowing the existence of their collection, because we are not able to characterize and distinguish its elements, in some way or another. We could imagine that, for Cantor, the indefiniteness of the sets was only proper of the universe of *all* sets, while in the restricted case of all the subsets of a sets we should be able, in principle, to have a more definite picture of them. Then, although for different reasons, for both the universe of set theory and the powerset axiom, Cantor was against the priority of objects over domains, and in some cases, also over sets.

We see here an anxiety, of Cantor, towards what is normally called the quasi-combinatorial conception of sets, in the context of its application to the power set operation. This position has been named by Bernays in 1935 ([17]).

But analysis is not content with this modest variety of platonism [to take the collection of all numbers as given]; it reflects it to a stronger degree with respect to the following notions: set of numbers, senquece of numbers, and function. It abstracts from the possibility of giving definitions of sets, sequences, and functions. These notion are used in a 'quasi-combinatorial' sense, by which I mean: in the sense of an analogy of the infinite to the finite.

Consider, for example, the different functions which assign to each member of the finite series $1, 2, \dots, n$ a number of the same series. There are n^n functions of this sort, and each of them is obtained by n independent determinations. Passing to the infinite case, we imagine functions engendered by an infinity of independent determinations which assign to each integer an integer, and we reason about the totality of these functions.

In the same way, one views a set of integers as the result of infinitely many independent acts deciding for each number whether it should be included or excluded. We add to this the idea of the totality of these sets. Sequences of real numbers and sets of real numbers are envisaged in an analogous manner. From this point of view, constructive definitions of specific functions, sequences, and sets are only ways to pick out an object which exists independently of, and prior to, the construction. The axiom of choice is an immediate application of the quasi-combinatorial concepts in question⁵.

⁵[17], p. 259-260.

In Bernays's description the quasi-combinatorial idea is perfectly linked to the priority of the object over their construction - or definition. Moreover, even if Cantor agrees with this point of view, for what concerns the Axiom of Choice - considering it a logical principle - he seems to disagree with Bernays on the matter of the powerset axiom. This may be the reason why Cantor has never placed his theorem on the uncountability of \mathbb{R} , at the hearth of its system - as it is clear from its absence in the *Beiträge* - nor he ever phrased it in terms of the power set axiom.

Then, for what concerns the answer to the question whether forcing - considered as a means to define new ideal object - is consistent with the modern concept of set, we can say that, as long as a quasi-combinatoric attitude is concerned, the priority that forcing gives to objects, over domains, is compatible with the conceptual framework that motivates the contemporary axiomatization of set theory. Indeed, this attitude is common to Zermelo's set theory, but it suffered some conceptual difficulties in Cantor's reflections. Hence, the philosophical prize we have to pay, in considering the conceptual meaning of the Forcing Axioms, causes no harm to the contemporary view on the nature of sets⁶.

4.2 Forcing Axioms as unifying axioms

We can now come back to the third criterion we isolated, elaborating on the foundational role of set theory: *Unification*. As we saw unification and justification are two sides of the same coin, brought together by necessary and a sufficient condition, if possible, in the form of a completeness theorem that ties together closely syntax and semantics.

In the case of the Forcing Axiom, this job is done by Viale's theorem, in the context of the structure $H(\aleph_2)$, thanks to a strengthening of MM, called MM^{+++} . Due to its necessary and sufficient form, Theorem 2.1.21 on one side shows how to complete the structure $H(\aleph_2)$, explaining its theorems in terms of the axioms ZFC + "There are class many super-huge cardinals" + MM^{+++} , on the other hand it justifies these axioms in terms of their unifying role for the

⁶We want to clear that the recognition of a realist stand in contemporary set theory is, by no means, a way to argue in favor of a form of platonism, in the form of an indispensability argument for modern mathematics. As a matter of fact, this realist stand could be seen as a form of methodological platonism, as the one advanced in the first pages of [32]: "A pragmatic if not logical argument in favor of the transcendental attitude is the observation that assuming the existence of an absolute non-denumerable continuum makes mathematics much simpler and easier, just as the outlook of physics is presumably simplified by the hypothesis of the existence of physical bodies."

theory of $H(\aleph_2)$, showing the correctness of these principle for that context.

But in section 2.2.2 an aspect was left to be clarified: that of the intuitive presentation of a theory, that allows to grasp this latter with a unifying act. Whenever this is possible, not only the axioms can show their completeness - as the theorems proved by Woodin and Viale show, respectively for $H(\aleph_1)$ and $H(\aleph_2)$ - but they can also be justified, by means of a matching between intuition and formalization, that, as we saw, was the problem behind Hilbert's foundational efforts.

In chapter 2 we argued that the subject matter of set theory is not clear and distinct enough, in order to develop a global intuition, able to characterize its intended model and isolate its axioms. This aspect of set theory is - more than a philosophical position - a theorem in ZFC: the so called Reflection Principle. It states that any global property of sets can be relativized to an initial segment of the universe V . As a consequence, it is not possible to characterize the universe of set theory⁷, since any characterization would already hold for a proper initial segment and hence it would cease to describe the whole universe. However, even if the Reflection Principle sustains the idea that it is not possible to develop a global intuition of the concept of set, we believe that a local intuition is possible, thanks to the division of V in cumulative levels.

We do not want here to point in the direction of an intuition *à la* Gödel, that allows to grasp the notion of set, as *set of*, and so to specify it, in order to single out the correct axioms for set theory. Quite the contrary, our argument is more empirical than metaphysical: it is not the notion of set that determines the axioms for set theory, but it is the cumulative hierarchy approach - thanks to which we describe the models for ZFC - that stimulates a contextual intuition for a set-like portion of the universe of sets and that allows a more conceptual reflection of quasi-empirical⁸ data, the proposal of axioms, and, finally, their intuitive justification⁹. It is not the notion of set, vague enough to be able to

⁷We are not saying here that it is not possible to define V , as the formula $x = x$ perfectly does.

⁸We refer here to Putnam's definition of quasi-empirical as can be found in [143], p. 62: "By 'quasi-empirical' methods I mean methods that are analogous to the methods of the physical sciences except that the singular statements which are 'generalized by induction', used to test 'theories', etc., are themselves the product of proof or calculation rather than being 'observation reports' in the usual sense. "

⁹Notice that by the same form of Woodin's and Viale's theorems, the notion of being true in a particular structure is "defined" by the axioms for that theory. Then it is somehow tautological to talk about the truth in a structure, in this context. However, as we explained, justification and explanation are two sides of the same coin and this aspect amounts exactly in the prescriptive character of sufficient axioms, as it was the case for Hilbert's Completeness Axioms.

give a foundation to mathematics, that determines the axioms, but our partial and incomplete description of it - i.e. the mathematical theory we have: ZFC and its models presented in terms of the cumulative hierarchy - that makes possible to grasp, step by step, initial segments of the universe of set theory. As the Reflection Principle suggests, a complete description of the universe of set theory will never be possible and so, if we avoid a metaphysical intuition and, alternatively, we assume that the concept of set is implicitly defined by the axiomatization of set theory, any intuition of the notion of set will always be incomplete. This is the reason why the intuition that acts in the context of $H(\aleph_2)$ is different from the intuition of The concept of set. On the contrary, it is sufficiently clear to allow its specification thanks to the Forcing Axioms.

Many different aspects concur in forming this contextual intuition. First of all, the possibility of defying the domain - in Zermelo's sense - of the mathematical objects we want to axiomatize is a good starting point, in order to grasp its structure in a single mental act. Strange enough, and this is one of the more astonishing discovery of Cantor, even if the universe of set theory cannot be defined, it is possible to climb it up indefinitely by means of the ordinals. So, thanks to the cumulative hierarchy conception, it is possible to define every initial segment of the universe of sets.

Another aspect to be considered is the large amount of work that has been done, trying to understand the structure $H(\aleph_2)$, by means of Forcing Axiom. This has helped in developing a sufficiently reliable intuition on the theory $H(\aleph_2)$. Due to this work it was empirically noticed that any hypothesis formulable in $H(\aleph_2)$ either had a solution by means of Forcing Axioms, or the opposite solution could be find directly in ZFC. Moreover, and this brings us back to the unification power of the axioms of set theory in terms of arguments patterns, useful set-theoretical principles - consequences of the Forcing Axioms - were isolated in order to show explicitly the combinatorial argument behind different theorems. Two examples for many, that will be presented in details in what follows, are the Opening Coloring Axiom (OCA) and the P-Ideal Dichotomy (PID). A hint that these principles could be appropriate set-theoretical tools in set theory consists in the fact that, in the case of OCA, new theorems of ZFC could be discovered. Indeed the following result, due to Todorćević, is in the same argumentative pattern of OCA.

Theorem 4.2.1. *Let $X \subseteq \mathbb{R}$ be a Σ_1^1 set and let $K \subseteq [\mathbb{R}]^2 = \{(x, y) : x > y\}$ be an open subset. Then just one of the following holds:*

- $\exists P \subseteq X$ perfect set that is homogeneous for K (i.e. $[P]^2 \subseteq K$),
- X is covered by countably many sets homogeneous for $[\mathbb{R}]^2 \setminus K$.

Then arguing for an intuitive unifying grasp of the structure $H(\aleph_2)$ we can say that the Forcing Axioms are justified thanks to Viale's theorem. However a question is still open: are they natural?

4.3 Forcing Axioms as natural axioms

As we showed in the last chapter, the naturalness of an axiom or of a concept amounts in the matching between its meaning and import together with the aspects of the underlying theory that are chosen to be relevant, in the context of a realist philosophical attitude towards mathematics.

As we hope to have shown, a realist stand is grounded in the genetic makeup of the Forcing Axiom and this aspect fits perfectly with both set-theoretical practice and the conceptual prerequisites for a set theory presented in terms of a cumulative hierarchy. Then, the recognition of this aspects, together with the fact that the notion of set is as vague as the notion of mathematics, account, at the same time, for the naturalness of a set-theoretical foundation of mathematics and for the naturalness of the Forcing Axioms. However, there is a subtle point here to be cleared: with respect to what Forcing Axioms are natural, if Viale's result is presented in terms of the structure $H(\aleph_2)$ - that is a model of ZFC – Powerset Axiom¹⁰ - but we maintained that one of the major realist aspects of the cumulative hierarchy comes from the central place that the Powerset Axiom occupies - together with the sequence of the ordinals - in his definition?

As a matter of fact the two cumulative hierarchies given by the $\mathcal{P}(\alpha)$'s and the $H(\theta)$'s do not normally coincide, except for ω and the strong limit inaccessible cardinals. Then, if we cannot control the cardinal exponentiation that links the two hierarchies, we cannot say, from Viale's theorem that Forcing Axioms are natural axioms for the structure $H(\aleph_2)$, but just that they are natural for ZFC, shaped in a cumulative hierarchy. This conclusion, even if not problematic for a general account of naturalness in set theory, would seem quite strange if we consider the axiom MM^{+++} locally. But fortunately Forcing Axioms do have an impact on cardinal arithmetic and indeed MM^{+++} decides the cardinality of the continuum, implying that $2^{\aleph_0} = \aleph_2$. Then, we have that $V_{\omega+1} \subseteq H(\aleph_2)$ and so we can argue that MM^{+++} is a natural axiom for $V_{\omega+1}$. This latter is the structure where we can find all the subsets of the natural numbers, and so all the reals. Thus, for what concerns naturalness, we can say that MM^{+++} is indeed natural with respect to the realist attitude towards the subsets of the natural

¹⁰As it has been shown in [43], we assume that $H(\theta)$ models, together with the other Zermelo's axioms the comprehension schema and not the replacement, to avoid bad behavior of the structure.

numbers that is implicit in the presentation of the structure $\mathcal{P}(V_\omega)$. This attitude maintains that besides the definable subsets of \mathbb{N} , there are also *arbitrary* sets, whose existence is granted by Cantor's theorem on the non-denumerability of \mathbb{R} . As it is well explained in [30], this notion motivated the works of the founding fathers of set theory, being at the background of their first steps in the field. This conception emerged first in the theory of natural numbers and of the reals and was then extended, thanks to the widespread use of the Powerset Axiom, to the entire mathematics. The acceptance of the existence of arbitrary sets of natural number - i.e. sets that cannot be defined in first order logic, being our language countable - is not only at the base of the naturalness of Forcing Axioms, but it is also sharpened and made more precise by the method of forcing. As it is argued in [30], the axioms of ZFC fall short in capturing this notion, but we believe that the methods of definition, that are introduced thanks to the method of forcing, make this notion more precise, allowing the possibility to talk and thus 'define' object that live outside a given domain of mathematical objects. From the point of view of a countable transitive model of ZFC, an arbitrary set of natural number is a set whose existence can be forced by means of a notion of forcing. Indeed, this is a sharpening of the idea of arbitrary sets, that amounts in the prescriptive rationale along which arbitrary sets are generic sets.

We are now in the position to better understand better the distinction between *intrinsic* and *extrinsic* justifications for axioms of set theory. As the analysis of Woodin's and Viale's results show, on the one hand the fruitfulness of an axiom presupposes the intuitive definition of a theory, for which the axioms should show its naturalness, on the other hand the intrinsic reasons presuppose a unifying power of the axioms that explain mathematical theorems in terms of argument patterns showed in their proof. Thus, when we try to apply this distinction to concrete cases, we find that the dividing line is not so neat and the two aspect interacts closely.

In the end we want to stress again that our goal was not to show that Forcing Axioms are natural because they agree with the concept of set. We believe that if a form of justification can be accepted, it does not make reference to the notion of set but to the formal presentation of set theory, in terms of a cumulative hierarchy. This is indeed a right form of justification, but when it comes to naturalness, besides these reasons, there is a realist stand that need to be checked prior to the conceptual relevance of Forcing Axioms for the axiomatization of the modern notion of set. To put it more directly: Forcing Axioms are justified in terms of the cumulative hierarchy view, and can be considered natural as long as they appeal to a realist attitude, that is shared by the underlying presentation of set theory.

In [87], Jané suggested that there may be two different ways to approach set theory.

For we have to distinguish two aspects in set theory, which for want of better names we call the *conceptual* aspect and the strictly *mathematical* aspect. The iterative description of the set-theoretical universe belongs in the conceptual aspect. [...] When we go from the conceptual to the mathematical aspect, we change our perspective and our basic notions. In mathematics proper there is no room for the ideal closure of an open plurality of relative sets, as there isn't either for the generating rules or for the notion of an indefinitely extensible concept¹¹.

Contrary to this last quotation we do not believe in a clear separation between the conceptual and the mathematical side of set theory. We tried to argue, thanks to the inquire on naturalness in mathematics, that mathematicians not only are personally involved in a foundational enterprise with a strong conceptual character, but also that it can be found in the mathematical literature a masked but strong need for philosophical considerations, in terms of reference to naturalness. Even if the notion of set is too vague to inspire a precise mathematical treatment, we believe not only that a constant dialogue between mathematicians and philosophers is useful, but also that there is a deep influence between the two sides. The conceptual and the formal side need to be distinguished but not to kept apart, being, as we argued, two sides of the same coin.

¹¹[87], pp. 19-20.

Chapter 5

The side conditions

In this chapter starts the more mathematical part of this work. Since now, we argued in favor of the acceptance of the Forcing Axioms, trying to explain their naturalness. We now want to get our hands dirty presenting them in some details and showing their effects on mathematics. In this task our guiding line will be a technique, closely related to properness - a property of posets that names one of the more important Forcing Axiom: the Proper Forcing Axiom (PFA) - that amounts in using models of set theory as part of the definition of a poset: as side conditions. This idea was first introduced by Stevo Todorčević at the beginning of the Nineties, in [170]. In order to explain this idea, we need to introduce some definitions.

5.1 Some technical background

Both when defining properness and when building proper forcings, it will be convenient to work with set models of a sufficiently strong fragment of ZFC. The most appropriate fragment in the present setting is ZFC with the power set axiom omitted. Models of this theory are provided by the structures $(H(\theta), \in)$, where θ is an uncountable regular cardinal and $H(\theta)$ is the collection of all sets of hereditary cardinality less than θ . Note in particular that if A is in $H(\theta)$, then $\mathcal{P}(A) \subseteq H(\theta)$ and if A and B are in $H(\theta)$, then so is $A \times B$. In particular, $H(\theta) \models |A| \leq |B|$ if and only if $|A| \leq |B|$. Since we will frequently be working with elementary substructures of some fixed $H(\theta)$, it will be useful to also fix a well ordering \triangleleft of $H(\theta)$. This provides a nice set of Skolem functions for $H(\theta)$.

In what follows, we will only be interested in countable elementary submodels of $H(\theta)$ and will write $M \prec H(\theta)$ to mean that M is a countable elementary submodel of $H(\theta)$. We will say that θ is *sufficiently large* with respect to X if

$\mathcal{P}(X)$ is in $H(\theta)$. A *suitable model* for X is an $M \prec H(\theta)$ where X is in M and θ is sufficiently large for X .

Definition 5.1.1. If \mathbb{Q} is a forcing notion and M is a suitable model for \mathbb{Q} , then a condition $q \in \mathbb{Q}$ is (M, \mathbb{Q}) -generic if whenever $r \leq q$ and $D \subseteq \mathbb{Q}$ is dense and in M , there is a $s \in D \cap M$, such that s and r are compatible (i.e. $D \cap M$ is predense below q).

The above definition is equivalent to saying that

$$q \Vdash \dot{G} \cap \check{M} \text{ is } \check{M}\text{-generic}$$

where \dot{G} is the \mathbb{Q} -name for the generic filter.

Definition 5.1.2. A forcing notion \mathbb{Q} is proper if whenever M is a suitable model for \mathbb{Q} and q is in $\mathbb{Q} \cap M$, q has an extension which is (M, \mathbb{Q}) -generic.

The following are useful facts about countable elementary submodels.

Fact 5.1.3. If X is definable in $H(\theta)$ from parameters in $M \prec H(\theta)$, then $X \in M$.

Fact 5.1.4. If $X \in M \prec H(\theta)$, then X is countable if and only if $X \subseteq M$

Proof. Only the reverse implication requires argument. Observe that since elements of ω are definable, $\omega \subseteq M$. If X is countable, then, by elementarity, M will contain a function from X into ω . Since ω is contained in X , its preimage under this function is as well and hence $X \subseteq M$. \square

Fact 5.1.5. If $M \prec H(\theta)$, then $M \cap \omega_1$ is a countable ordinal.

Proof. If $\alpha \in M \cap \omega_1$, then $\alpha \subseteq M$ by the previous fact. Hence $M \cap \omega_1$ is a transitive set of ordinals and therefore an ordinal. \square

The goal of a forcing with side conditions is to simplify the proof that a condition is (M, \mathbb{P}) -generic, incorporating the model M itself in the main definition of the forcing \mathbb{P} . Hence, asking from the outset some compatibility requirements between models and conditions.

5.2 The pure side conditions poset \mathbb{M}_θ^1

One of the consequences of constructing a scaffolding of models around a poset is the preservation of ω_1 . The reason is the properness of the pure side condition poset that we now present. In the next chapters we will also introduce a pure side condition poset that preserves ω_1 and ω_2 .

Definition 5.2.1. Given a regular cardinal θ , we define \mathcal{S}^θ as the set of countable $M \prec H(\theta)$.

Notice that \mathcal{S}^θ is a club in $[H(\theta)]^{\aleph_0}$.

Definition 5.2.2. Given a regular cardinal θ , we let \mathbb{M}_θ^1 be the poset consisting of conditions $p = \mathcal{M}_p$ such that

1. \mathcal{M}_p is a finite \in -chain
2. every $M \in \mathcal{M}_p$ is an element of \mathcal{S}^θ .

We say that $p \leq q$ if $\mathcal{M}_q \subseteq \mathcal{M}_p$.

With an abuse of notation we will confound a chain with the set of its models. Now we want to show that \mathbb{M}_θ^1 is proper in a stronger sense.

Let \mathbb{P} be a forcing notion. We say that a set M is *adequate* for \mathbb{P} if for every $p, q \in M \cap \mathbb{P}$ if p and q are compatible then there is $r \in \mathbb{P} \cap M$ such that $r \leq p, q$. Note that we do not require that \mathbb{P} belongs to M . In the forcing notions we consider if two conditions p and q are compatible then this will be witnessed by a condition r which is Σ_0 -definable from p and q . Thus, all elements of \mathcal{S}_0^θ will be adequate for the appropriate forcing notions.

Definition 5.2.3. Suppose \mathbb{P} is a forcing notion and M is adequate for \mathbb{P} . We say that a condition p is (M, \mathbb{P}) -strongly generic if p forces that $\dot{G} \cap M$ is a V -generic subset of $\mathbb{P} \cap M$, where \dot{G} is the canonical name for the V -generic filter over \mathbb{P} .

The difference with properness is the fact that an (M, \mathbb{P}) -strongly generic condition not only forces that $\dot{G} \cap M$ is an M -generic subset of $\mathbb{P} \cap M$, but also that it is a V -generic subset of $\mathbb{P} \cap M$. Hence, it is clear that strong properness implies properness. In order to check that a condition is strongly generic over a set M we can use the following characterization, see [124] for a proof.

Fact 5.2.4. Suppose \mathbb{P} a notion of forcing and M is adequate for \mathbb{P} . A condition p is (M, \mathbb{P}) -strongly generic if and only if for every $r \leq p$ in \mathbb{P} there is a condition $r|_M \in \mathbb{P} \cap M$ such that any condition $q \leq r|_M$ in M is compatible with r . \square

Definition 5.2.5. Suppose \mathbb{P} is a forcing notion and \mathcal{S} is a collection of sets adequate for \mathbb{P} . We say that \mathbb{P} is \mathcal{S} -strongly proper, if for every $M \in \mathcal{S}$, every condition $p \in \mathbb{P} \cap M$ can be extended to an (M, \mathbb{P}) -strongly generic condition q .

Fact 5.2.6. Given a notion of forcing \mathbb{P} and a set $\mathcal{T} \subseteq [H(\theta)]^{\aleph_0}$ of models, that are adequate for \mathbb{P} , if \mathbb{P} is \mathcal{T} -strongly proper and if \mathcal{T} is stationary in $[H(\theta)]^{\aleph_0}$, then \mathbb{P} , preserves ω_1 , with respect to \mathcal{T} ; i.e. if $N \in \mathcal{T}$, then ω_1 is not collapsed in $N[G]$, with G a generic filter over \mathbb{P} . \square

5.3 Strong properness of \mathbb{M}_θ^1

In order to prove that \mathbb{M}_θ^1 is \mathcal{S}^θ -strongly proper, we need the following lemmas.

Lemma 5.3.1. *Suppose $M \in \mathcal{S}^\theta$ and $p \in \mathbb{M}_\theta^1 \cap M$. Then there is a new condition p^M , which is the smallest element of \mathbb{M}_θ^1 extending p and containing M as an element.*

Proof. Notice that $p \in M$, implies, by finiteness of p , that $p \subseteq M$. Then every model in p belongs to M , and in particular its top model. Hence, if $p = M_0 \in \dots \in M_k$, for some $k \in \omega$, we have that $p^M = M_0 \in \dots \in M_k \in M$ is a condition in \mathbb{M}_θ^1 extending p and containing M . \square

Lemma 5.3.2. *Suppose $r \in \mathbb{M}_\theta^1$ and $M \in \mathcal{M}_r$. Let $q \in M$ be such that $q \leq r \cap M$. Then q and r are compatible.*

Proof. Notice that $q \in M$ implies $q \subseteq M$. Then the top model of q belongs to M . Letting $q = M_0^q \in \dots \in M_n^q$, and $(r \setminus M) \setminus M = M_0^r \in \dots \in M_k^r$, we have that $M \in M_0^r$ and so that $r^* = q \cup \{M\} \cup (r \setminus M) \setminus M$: i.e.

$$M_0^q \in \dots \in M_n^q \in M \in M_0^r \in \dots \in M_k^r,$$

is an \in -chain. Moreover, since $q \leq r \cap M$, r^* extends both r and q . \square

As an immediate consequence of Lemma 5.3.2 we have the following.

Theorem 5.3.3. *\mathbb{M}_θ^1 is \mathcal{S}^θ -strongly proper.*

Proof. Suppose $M \in \mathcal{S}^\theta$ and $p \in M \cap \mathbb{M}_\theta^1$. We shall show that p^M is (M, \mathbb{M}_θ^1) -strongly generic. To see this we for every condition $r \leq p^M$ we have to define a condition $r|M \in \mathbb{M}_\theta^1 \cap M$ such that for every $q \in \mathbb{M}_\theta^1 \cap M$ if $q \leq r|M$ then q and r are compatible. If we let $r|M$ simply be $r \cap M$ this is precisely the statement of Lemma 5.3.2. \square

Corollary 5.3.4. *The forcing \mathbb{M}_θ^1 is strongly proper and so preserves ω_1 .* \square

However, this is not true for the cardinals between ω_1 and θ^+ .

Theorem 5.3.5. *The forcing \mathbb{M}_θ^1 collapses θ to ω_1 .*

Proof. For every $x \in H(\theta)$, let

$$D_x = \{p \in \mathbb{M}_\theta^1 : \exists M \in \mathcal{M}_p \wedge x \in M\},$$

and notice that is a dense subset of \mathbb{M}_θ^0 . Then if we let G be a V -generic filter in \mathbb{M}_θ^1 and \mathcal{M}_G be a generic \in -chain we have that, in $V[G]$, the structure $H(\theta)$ is covered by \mathcal{M}_G . Moreover, since \mathcal{M}_G is also an \subseteq -chain, because \mathcal{M}_G is a transitive chain, we have that its length is ω_1 . Hence the set

$$\{\alpha_M : M \in \mathcal{M}_G \wedge \alpha_M = \text{sup}(M \cap \theta)\}$$

has cardinality \aleph_1 and it is cofinal in θ . □

Chapter 6

PFA and some of its consequences

In these chapter we will present an exposition of the Proper Forcing Axiom (PFA). We will first discuss examples of the consequences of PFA. We will then present two proper partial orders which are used to force two combinatorial principles which follow from PFA: The P-Ideal Dichotomy (PID) and Todorcevic's formulation of the Open Coloring Axiom. On the one hand, these posets and the proofs of their properness are quite typical of direct applications of PFA. On the other hand, these principles already capture a large number of the consequences of PFA and do not use any terminology or technical tools from the theory of forcing.

6.1 Consequences of PFA

We start with the definition of PFA.

Definition 6.1.1. *PFA holds if, given a proper poset \mathbb{Q} and a collection \mathcal{D} of dense subsets of \mathbb{Q} , with $|\mathcal{D}| \leq \aleph_1$, then there is a filter $G \subseteq \mathbb{Q}$ such that $G \cap D \neq \emptyset$ for all $D \in \mathcal{D}$.*

It is not hard to see that property of properness is a weakening of the *countable chain condition (c.c.c.)*. In particular, PFA is a strengthening of $\text{MA}(\aleph_1)$.

The following two theorems predated PFA and were important in its formulation.

Theorem 6.1.2. *(Solovay, Tennenbaum [161]) Souslin Hypothesis is consistent with ZFC.*

Theorem 6.1.3. (*Baumgartner [11]*) *The following statement is consistent with ZFC: every pair of \aleph_1 -dense¹ subsets of \mathbb{R} are isomorphic.*

Martin observed that Solovay and Tennenbaum’s proof of the consistency of Souslin’s Hypothesis could be adapted to prove the consistency of a stronger statement, now known as $\text{MA}(\aleph_1)$. The model which Baumgartner constructed to establish the above theorem is, like Solovay and Tennenbaum’s model, obtained by a finite support iteration of forcings which satisfy the c.c.c.. Unlike Souslin’s Hypothesis, however, the conclusion of Baumgartner’s theorem did not apparently follow from $\text{MA}(\aleph_1)$ (this was later confirmed by work of Abraham and Shelah [2]).

Let us now consider two principles which we will later demonstrate are consequences of PFA. Recall that an ideal $\mathcal{I} \subseteq [S]^\omega$ is a *P-ideal* if

- I contains every finite subsets of S ,
- for every family $\{X_n : n \in \omega\} \subseteq I$ there is an $X \in I$ such that X_n is contained in X modulo a finite set for every $n \in \omega$.

Definition 6.1.4. (*P-Ideal Dichotomy (PID)*) *If S is a set and $\mathcal{I} \subseteq [S]^\omega$ is a P-ideal, then either*

- *there is an uncountable $Z \subseteq S$ such that $[Z]^\omega \subseteq \mathcal{I}$, or*
- *$S = \bigcup_{n \in \omega} S_n$ such that no infinite subset of any S_n is in \mathcal{I} .*

Definition 6.1.5. (*Open Coloring Axiom (OCA)*) *If G is an open graph on a separable metric space X , then either*

- *there is an uncountable $Z \subseteq X$ such that $[Z]^2 \subseteq G$, or*
- *$X = \bigcup_{n \in \omega} X_n$ such that $[X_n]^2 \cap G = \emptyset$, for every $n \in \omega$.*

Remark 6.1.6. The formulation of this principle is due to Todorcevic and is based on the different formulations of OCA presented in [1].

We will now turn to some of the consequences of these principles.

Theorem 6.1.7. (*Todorcevic [172]*) *PFA implies that if X is a Banach space of density \aleph_1 , then X has a quotient with a basis of length ω_1 .*

In fact for Todorcevic’s result, the conjunction of PID and $\text{MA}(\aleph_1)$ is sufficient to derive the desired conclusion.

¹A set $X \subseteq \mathbb{R}$ is said to be κ -dense, if every interval meets X in κ points.

Theorem 6.1.8. (Farah [26]) *OCA implies that if H is a separable infinite dimensional Hilbert space, then all automorphisms of $\mathcal{B}(H)/\mathcal{K}(H)$ are inner.*

Theorem 6.1.9. (Moore [128]) *PFA implies that every uncountable linear order contains an isomorphic copy of one of the following: $X, \omega_1, -\omega_1, C, -C$. Here X is a set of reals of cardinality \aleph_1 and C is a Countryman line.*

Theorem 6.1.10. (Moore [129]) *PFA implies that η_C is universal for the Aronszajn lines. Here η_C is the direct limit of the finite lexicographic products of the form $C \times (-C) \times \dots \times (\pm C)$.*

Theorem 6.1.11. (Martinez [120]) *PFA implies that the Aronszajn lines are well quasi ordered by embeddability.*

Theorem 6.1.12. (Todorćević, Velićković [13] [179]) *PFA implies that $2^{\aleph_0} = \aleph_2$.*

Theorem 6.1.13. (Todorćević [168]) *PFA implies that for every regular $\kappa > \aleph_1$, $\square(\kappa)$ fails.*

Theorem 6.1.14. (Viale [185]) *PFA implies $2^\mu = \mu^+$ whenever μ is a singular strong limit cardinal.*

The two last theorems are consequences of PID [171] [183].

Theorem 6.1.15. (Todorćević [169]) *OCA implies that every (κ, λ^*) -gap in ω^ω/fin is of the form $\kappa = \lambda = \omega_1$ or $\min(\kappa, \lambda) = \omega$ and $\max(\kappa, \lambda) \geq \omega_2$.*

6.2 Forcings associated to OCA and PID

We will now define posets associated to OCA and to PID and prove that they are proper.

6.2.1 The poset for OCA

Suppose that G is an open graph on X . Let $\mathbb{Q}_{X,G}$ be the poset consisting of all the pairs $q = (H_q, \mathcal{N}_q)$ such that:

1. $H_q \subseteq X$ is a finite clique in G ;
2. \mathcal{N}_q is an increasing finite \in -chain of elementary submodels of $H(2^{\aleph_0}^+)$ which each contain (X, G) ;
3. if $x \neq y$ are in H_q , then there is an N in \mathcal{N}_q such that $|N \cap \{x, y\}| = 1$;

4. if N is in \mathcal{N}_q and $E \subseteq X$ is in N with $[E]^2 \cap G = \emptyset$, then $E \cap H_q \subseteq N$.

Define the order on $\mathbb{Q}_{X,G}$ by $p \leq q$ if $H_q \subseteq H_p$ and $\mathcal{N}_q \subseteq \mathcal{N}_p$. Notice that, by *condition 3*, \mathcal{N}_q induces an order on H_q corresponding to the number of elements of \mathcal{N}_q which do not contain a given element of H_q . It will be convenient to let $x_q(i)$ denote the i^{th} element of H_q in this enumeration.

Remark 6.2.1. If CH holds, then can modify the forcing $\mathbb{Q}_{X,G}$ (and $\mathbb{Q}_{X,G}^*$ below) as follows. Let \mathcal{N} be a continuous \in -chain of length ω_1 of countable elementary submodels of $H(2^{\aleph_0}^+)$, each containing the relevant objects. Define, e.g., $\mathbb{Q}_{X,G}$ to be all H_q such that (H_q, \mathcal{N}_q) is in $\mathbb{Q}_{X,G}$ whenever \mathcal{N}_q is a finite subset of \mathcal{N} . One can verify that this modified forcing in fact satisfies the countable chain condition. The role of CH here can be explained as follows: if X is a separable metric space and if M and N are two countable elementary submodels of $(H(2^{\aleph_0}^+), \in, X)$ such that $M \cap \omega_1 = N \cap \omega_1$, then CH implies that $M \cap H(\aleph_1) = N \cap H(\aleph_1)$ and hence M and N have the same intersection with the closed subsets of X . Notice that in *condition 4*, the set E can be assumed to be closed since if $[E]^2 \cap G$ is empty, the same is true for the closure of E .

6.2.2 The poset for PID

Let I be a P-ideal on a set S and suppose that θ is a regular cardinal such that I is in $H(\theta)$. For each countable subset X of I , let I_X be an element of I such that $I \subseteq_* I_X$ for every I in X . Here $I \subseteq_* J$ means that the set $I \setminus J$ is finite. If M is a countable elementary submodel of $H(\theta)$, will let I_M denote $I_M \cap \mathcal{I}$. Define I^\perp to be the collection of all subsets of S which have finite intersection with every element of I .

Let $\mathbb{Q}_{\mathcal{I}}$ be the poset consisting of all the pairs $q = (Z_q, \mathcal{N}_q)$ such that

1. $Z_q \subseteq S$ is finite;
2. \mathcal{N}_q is an increasing finite \in -chain of elementary submodels of $H(\theta)$ which each contain I ;
3. if $x, y \in S$, then there is an N in \mathcal{N}_q such that $|N \cap \{x, y\}| = 1$;
4. if N is in \mathcal{N}_q and X is in $N \cap \mathcal{I}^\perp$, then $X \cap Z_q \subseteq N$.

Define the order on $\mathbb{Q}_{\mathcal{I}}$ by:

1. $p \leq q$ if $Z_q \subseteq Z_p$;
2. $\mathcal{N}_q \subseteq \mathcal{N}_p$;
3. $(Z_p \setminus Z_q) \cap N \subseteq I_N$, whenever N is in $N \in \mathcal{N}_q$.

6.3 Verification of properness

The above examples of forcings all have a common form. For instance, their elements consist of pairs $q = (X_q, \mathcal{N}_q)$ where X_q is a finite approximation of some desired object and \mathcal{N}_q is a finite \in -chain of countable elementary submodels of some suitably chosen $H(\theta)$. In order to verify properness of such forcings, the general strategy is to argue that if M is a suitable model for \mathbb{Q} and $M \cap H(\theta)$ is in \mathcal{N}_q , then q is (M, \mathbb{Q}) -generic. The forcing \mathbb{Q} is defined in such a way that it is trivial to verify that if q_0 is in $M \cap \mathbb{Q}$, then

$$q = (X_{q_0}, \mathcal{N}_{q_0} \cup \{M \cap H(\theta)\})$$

a condition in \mathbb{Q} which extends q_0 .

In order to verify that $M \cap H(\theta) \in \mathcal{N}_q$ implies the (M, \mathbb{Q}) -genericity of q , one usually argues that:

(*) if $D \subseteq \mathbb{Q}$ is in M and q is in D , then q is compatible with an element of $D \cap M$.

Notice that this implies genericity: if $D \subseteq \mathbb{Q}$ is dense and in M , then we can first extend q to an element r of D and then appeal to (*) to find a condition s in $D \cap M$ which is compatible with r and hence with q .

We now show that $\mathbb{Q}_{X,G}$ and $\mathbb{Q}_{\mathcal{I}}$ are proper and argue that $\mathbb{Q}_{A,B}$ can be viewed as a special case of a modified version of $\mathbb{Q}_{X,G}$.

Theorem 6.3.1. $\mathbb{Q}_{X,G}$ is proper.

Proof. Suppose that M is suitable for $\mathbb{Q}_{X,G}$, r is in $\mathbb{Q}_{X,G}$, and $N = M \cap H(2^{\aleph_0}^+)$ is in \mathcal{N}_r . It is sufficient to verify that r is $(M, \mathbb{Q}_{X,G})$ -generic and this will be done by verifying (*). To this end, let $D \subseteq \mathbb{Q}_{X,G}$ be in M and contain r . Fix disjoint open sets U_i ($i < n$) which are in N such that $x_r(i)$ is in U_i and $U_i \times U_j \subseteq G$ whenever $i \neq j < n$.

By replacing D with a subset if necessary, we may assume that if s is in D , then

- $s \leq r_0$,
- for some $N_s \in \mathcal{N}_s$, $r_0 = (H_s \cap N_s, \mathcal{N}_s \cap N_s)$,
- $|H_s| = |H_r| = n$, and
- $x_s(i) \in U_i$ for all $i < n$.

Let $r_0 = r \cap M$ and note that $r_0 \in N$.

To find a condition s in $D \cap M$ compatible with r we have to check that $[H_r \cup H_s]^2 \subseteq G$. By construction of the set D , we just need to show that

$$\{x_r(i), x_s(j)\} \in G \text{ for } i = j,$$

because for $i \neq j$, U_i and U_j already witness the fact that $\{x_r(i), x_s(j)\} \in G$.

The following is now the key lemma.

Lemma 6.3.2. *Suppose that:*

- $N \prec H(2^{\aleph_0}^+)$ with $(X, G) \in N$;
- $A \subseteq X^n$ is in N ;
- $\bar{x} \in \text{cl}(A)$ and $\bar{x} \upharpoonright n-1 \in M$;
- for every $E \subseteq X$ in M with $[E]^2 \cap G = \emptyset$, $x(n-1)$ is not in $\text{cl}(E)$.

Then there is a basic open $U \subseteq X$ not containing $x(n-1)$ such that

$$\{x(n-1), y\} \in G \text{ for every } y \in U,$$

$$\bar{x} \upharpoonright n-1 \in \text{cl}(\{\bar{y} \upharpoonright (n-1) : \bar{y} \in A \wedge y(n-1) \in U\}).$$

Proof. Let $\{W_i : i \in \omega\}$ be a neighborhood base for $\bar{x} \upharpoonright (n-1)$ which is in M and define

$$E_0 = \bigcap_i \text{cl}(\{\bar{y}(n-1) : \bar{y} \in A \text{ and } \bar{y} \upharpoonright (n-1) \in W_i\}).$$

Notice that $x(n-1) \in E_0$. Observe that E_0 is in M since it is definable from parameters in M . Similarly,

$$E = \{z \in E_0 : \forall y \in E_0 \{z, y\} \notin G\}$$

is in M . By the fourth condition in the definition of $\mathbb{Q}_{X,G}$, $x(n-1) \notin E$. Pick then $y \in E_0$ such that $\{x(n-1), y\} \in G$. Since G is open, we can find a basic open neighborhood U of y which does not contain x such that $\{x(n-1)\} \times U \subseteq G$.

We now must show that

$$\bar{x} \upharpoonright n-1 \in \text{cl}(\{\bar{y} \upharpoonright (n-1) : \bar{y} \in A \wedge y(n-1) \in U\}).$$

Let $j \in \omega$ be given. Since U is open and y is in E_0 , there is a \bar{z} in A such that $\bar{z} \upharpoonright (n-1)$ is in W_j and $z(n-1)$ is in U . Since $\{W_i : i \in \omega\}$ is a neighborhood base at $\bar{x} \upharpoonright (n-1)$ and j was arbitrary, we have the desired conclusion. \square

We are now ready to complete the proof that $\mathbb{Q}_{X,G}$ is proper. Set $A = \{\bar{x}_s : s \in D\}$ and, using the above lemma, inductively construct a sequence of sets A_i ($i \leq n$) in N such that:

- $A_0 = A$ and $A_i \subseteq A_{i-1}$ if $i > 0$;
- $x_r \upharpoonright i$ is an accumulation point of $\{y \upharpoonright i : y \in A_i\}$;
- if $i < j < n$ and y is in A_i , then $\{x(j), y(j)\} \in G$.

Now if s is in N and \bar{x}_s is in A_n , it follows that s is compatible with r . □

Finally we turn to the proof that the PID forcing is proper.

Lemma 6.3.3. \mathbb{Q}_I is proper.

Proof. Let M be a suitable model for \mathbb{Q}_I and suppose that r is in \mathbb{Q}_I with $M \cap H(\theta)$ in \mathcal{N}_r . Define $r_0 = (Z_r \cap M, \mathcal{N}_r \cap M)$ and let $n = |Z_r \setminus M|$. We will verify that (*) holds. To this end, suppose that $D \subseteq \mathbb{Q}_I$ with D in M and r in D . By replacing D with a subset if necessary, we may assume that if s is in D then:

- there is an N_s in \mathcal{N}_s such that $r_0 = (Z_s \cap N_s, \mathcal{N}_s \cap N_s)$;
- $|Z_s \setminus N_s| = n$.

By definition of extension in \mathbb{Q}_I , our goal will be to find an $s \in D \cap M$ such that $Z_s \setminus N_s \subseteq I_P$, for each $P \in \mathcal{N}_r \setminus M$. This is sufficient, since then $(Z_s \cup Z_r, \mathcal{N}_s \cup \mathcal{N}_r)$ is in \mathbb{Q}_I and is a common extension of r and s .

Define $T = \{t_s : t_s = Z_s \setminus N_s \text{ and } s \in D\}$.

Claim 6.3.4. *There is a $T' \subseteq T$ which is \mathcal{J}^+ -splitting, where \mathcal{J} is the σ -ideal generated by \mathcal{I}^\perp .*

Proof. If $U \subseteq S^n$, define ∂U to be the set of all u in U such that there is no $k < n$ with

$$\{v(k) : (v \in U) \wedge (u \upharpoonright k = v \upharpoonright k)\} \in \mathcal{J}.$$

By definition, U is \mathcal{J}^+ -splitting if $\partial U = U$ and U is non empty. Observe that for any $U \subseteq S^n$, if u is in $\partial^{i+1}U \setminus \partial^i U$, then the k which witnesses this must be less than $n - i$. In particular, $\partial^{n+1}T = \partial^n T$. Observe that $T' = \partial^n T$ is in N and contains $t_r = Z_r \setminus N$. □

Now we inductively build $\sigma_0, \sigma_1, \dots, \sigma_n \in M$ such that $\sigma_0 = \emptyset$ and each σ_i is the initial part of some $s \in T'$ with σ_{i+1} having σ_i as an initial part whenever $i < n$. Given σ_i consider

$$\{x \in S : \sigma_i \hat{\ } x \text{ has an extension in } T'\}.$$

This set is in M , since it is definable from parameters which are in M . By elementarity of M and the definition of \mathcal{J} , the above set contains a countably infinite subset H in M that is in I . In particular, $H \subseteq_* I_P$ for all $P \in \mathcal{N}_r$. Hence there is an element $x \in H \cap \bigcap_{P \in \mathcal{N}_r} I_P$ such that $\sigma_{i+1} = \sigma_i \hat{\ } x$ has an extension in T' . Finally, at stage n , $\sigma_n = \sigma = Z_s \setminus N_s \subseteq \bigcap_{P \in \mathcal{N}_r} I_P$, for some $s \in D$. Such an s is now compatible with r . We have therefore established that \mathbb{Q}_I is proper. \square

6.4 Density arguments

While the main difficulty in proving implications such as *PFA implies PID* lies in the verification that the relevant forcing is proper, some additional argument is usually required to show that certain sets are dense. In the case of PID and OCA, this verification is straightforward. In order to illustrate the argument, we will go through the remainder of the proof that PFA implies PID.

Theorem 6.4.1. (PFA) *PFA implies PID.*

Proof. Suppose that $\mathcal{I} \subseteq [S]^\omega$ is a P-ideal. If S can be covered by countably many sets in \mathcal{I}^\perp , then there is nothing to show. Suppose that this is not the case and let M be a suitable model for $\mathbb{Q}_{\mathcal{I}}$ and $x \in X$ be outside of every element of $\mathcal{I}^\perp \cap M$. Define $q = (\{x\}, \{M \cap H(\theta)\})$. We have established that q is $(M, \mathbb{Q}_{\mathcal{I}})$ -generic and therefore q forces that

$$\dot{Z} = \bigcup_{p \in \dot{G}} Z_p$$

is uncountable. Also observe that q forces that every countable subset of S is contained in N for some N in $\dot{\mathcal{N}} = \bigcup_{p \in \dot{G}} N_p$. This is easily seen for countable sets in V and holds for sets in $V[G]$ as well since $\mathbb{Q}_{\mathcal{I}}$ is proper. Furthermore, q forces that if N is in $\dot{\mathcal{N}}$, then $N \cap \dot{Z} \subseteq_* I_N$ and hence is in \mathcal{I} .

Let \dot{f} be a $\mathbb{Q}_{\mathcal{I}}$ -name for an injection from ω_1 into \dot{Z} and let \dot{g} be a $\mathbb{Q}_{\mathcal{I}}$ -name for a function from ω_1 into \mathcal{I} such that q forces $\forall \delta < \omega_1 (\dot{f}'' \delta \subseteq \dot{g}(\delta))$.

If we let D_ξ be the set of all p such that p decides $\dot{f}(\xi)$ and $\dot{g}(\xi)$, then any filter G meeting D_ξ for all $\xi < \omega_1$ must satisfy that $Z = \dot{Z}[G]$ is an uncountable set such that every countable subset of Z is in \mathcal{I} . \square

6.5 Some application of PID

We now present some consequences of PID.

Definition 6.5.1. An ideal \mathcal{K} on a set S is countably generated in \mathcal{X} if there is a countable family $\{K_n : n \in \omega\} \subseteq \mathcal{X}$ such that for all $K \in \mathcal{K}$, there is an n such that $K \subseteq K_n$. We will say that \mathcal{K} is countably generated if it is countably generated in \mathcal{K} .

In what follows we will use the notational convention that $\mathcal{K} \upharpoonright X = \mathcal{K} \cap \mathcal{P}(X)$.

Theorem 6.5.2. PID implies that if \mathcal{K} is an ideal on S that is not countably generated in $\mathcal{K}^{\perp\perp}$, then there is an $X \subseteq S$ of size less or equal to \aleph_1 such that $\mathcal{K} \upharpoonright X$ is not countably generated in $\mathcal{K}^{\perp\perp}$.

Proof. Define $I = \mathcal{K}^\perp$ and assume that, for every countable $X \subseteq S$, $\mathcal{K} \upharpoonright X$ is countably generated in $\mathcal{K}^{\perp\perp}$.

Claim 6.5.3. I is a P -ideal

Proof. Let $I_n \in I$, for $n \in \omega$ and set $X = \bigcup_n I_n$. Since $X \subseteq S$ is countable, $\mathcal{K} \upharpoonright X$ is countably generated in $\mathcal{K}^{\perp\perp}$, and so fix a family $\{K_i : i \in \omega\} \subseteq \mathcal{K}^{\perp\perp}$ which witnesses this. Without loss of generality we can also suppose that $K_i \subseteq K_{i+1}$.

Now define

$$I_* = \bigcup_i (I_i \setminus K_i).$$

To see that this set is in \mathcal{I} , it is sufficient to show that it has finite intersection with K_n for each $n < \omega$. For a given n , the intersection of I_* with K_n is contained in $\bigcup_{i < n} I_i$, a set which is in \mathcal{I} and has finite intersection with K_n . Hence I_* is in I as desired. Moreover $I_n \subseteq_* I_*$, since $I_n \setminus I_* \subseteq K_n \cap I_n$, which is finite. \square

We can now apply PID. If $S = \bigcup_n S_n$ such that, for all n , $S_n \in I^\perp$, then $\{S_n : n \in \omega\}$ witnesses that \mathcal{K} is countably generated in $\mathcal{K}^{\perp\perp}$.

On the other hand, if $Z \subseteq S$ is uncountable and $[Z]^\omega \subseteq I$, then $Z \cap K$ is finite, for every $K \in \mathcal{K}$, because if $Z \cap K$, were infinite, then there would be a countable $Y \subseteq Z \cap K$, but then $Y \in [Z]^\omega \subseteq I$ and hence $Y \cap K$ would be finite. Z witnesses that $\mathcal{K} \upharpoonright Z$ is not countably generated in $\mathcal{K}^{\perp\perp}$, because there is no family $\{K_n\}_{n \in \omega}$ in $\mathcal{K}^{\perp\perp}$, such that, for all $K \in \mathcal{K}$, there is an n such that $K \subseteq K_n$. \square

Corollary 6.5.4. PID implies SCH.

Proof. We can state SCH in the following way: for every strong limit cardinal μ , with $\text{cof}(\mu) < \mu$, $2^\mu = \mu^+$. Suppose that SCH fails and let μ be the least witness to this. By Silver's theorem, $\text{cof}(\mu) = \omega$. Fix an increasing sequence $\{\mu_n\}_{n \in \omega}$ that converges to μ . For each $\beta < \mu^+$, find $K_{\beta,n} \subseteq \beta$ such that

- $|K_{\beta,n}| \leq \mu_n$,
- $\bigcup_n K_{\beta,n} = \beta$,
- $\forall \beta < \beta'$, for every m there exists an n such that $K_{\beta,m} \subseteq K_{\beta',n}$

The key point here is that if $X \subseteq \mu^+$ is countable, then there is a β such that for every $\beta' > \beta$, the ideal on X generated by $\{K_{\beta,n} \cap X : n \in \omega\}$ equals the ideal on X generated by $\{K_{\beta',n} : n \in \omega\}$. This follows from the fact that $2^{2^{\aleph_1}} < \mu^+$.

The above fact implies that, if \mathcal{K} is the ideal on μ^+ generated by $\{K_{\beta,n} : \beta < \mu^+, n \in \omega\}$, then $\mathcal{K} \upharpoonright X$ is countably generated $\forall X \in [\mu^+]^{\aleph_1}$.

So, by Theorem 6.5.2 and by PID, \mathcal{K} is countably generated in $\mathcal{K}^{\perp\perp}$. In particular there is a $Z \subseteq \mu^+$ cofinal in μ^+ such that all countable subsets of Z are contained in $K_{\beta,n}$, for some β and n .

Thus $|Z^\omega| = \mu^+ = |\mu^\omega|$ and so $2^\mu = \mu^+$, since any subset $X \subseteq \mu$ can be seen as $X = \bigcup_{n \in \omega} X \cap \mu_n$. \square

Corollary 6.5.5. *PID the failure of $\square(\kappa)$, for every regular $\kappa > \aleph_1$.*

Proof. We just sketch the proof. Recall that $\square(\kappa)$ is the following principle: there is a sequence $\langle C_\alpha : \alpha < \kappa \rangle$ such that:

- $C_\alpha \subseteq \alpha$ is club in α ;
- if α is a limit point of C_β , then $C_\alpha = C_\beta \cap \alpha$,
- there is no club $C \subseteq \kappa$ such that, for all limit points α of C , $C \cap \alpha = C_\alpha$.

Now, for $\alpha < \beta < \kappa$ inductively define

$$\varrho_2(\alpha, \beta) = 1 + \varrho_2(\alpha, \min(C_\beta \setminus \alpha)),$$

$$\varrho_2(\alpha, \alpha) = 0.$$

Thus $\varrho_2(\alpha, \beta)$ is the length of the walk from β to α .

We define, for $\beta < \kappa$,

$$K_{\beta,n} = \{\alpha < \beta : \varrho_2(\alpha, \beta) \leq n\}.$$

It can be verified that ϱ_2 has the following properties:

- if $\beta < \beta'$, then there is an n such that $|\varrho_2(\alpha, \beta) - \varrho_2(\alpha, \beta')| \leq n$ for all $\alpha < \beta$;
- if $X \subseteq \kappa$ is unbounded, then there is a β such that $\{\varrho_2(\alpha, \beta) : \alpha \in X \cap \beta\}$ is infinite.

It follows that the ideal \mathcal{K} generated by $\{K_{\beta, n} : (\beta < \kappa) \wedge (n < \omega)\}$ is not countably generated in $\mathcal{K}^{\perp\perp}$ and yet $\mathcal{K} \upharpoonright \beta$ is countably generated in $\mathcal{K}^{\perp\perp}$ for all $\beta < \kappa$. \square

Chapter 7

The five element basis theorem

In [128] Moore showed that PFA implies that the class of the uncountable linear orders has a five element basis, i.e., that there is a list of five uncountable linear orders such that every uncountable linear order contains an isomorphic copy of one of them. This basis consists of X , ω_1 , ω_1^* , C , and C^* , where X is any suborder of the reals of cardinality ω_1 and C is any Countryman line¹. It was previously known from the work of Baumgartner [11] and Abraham-Shelah [3], that, assuming a rather weak forcing axiom, the existence of a five element linear basis for uncountable linear orderings is equivalent to the following statement, called the *Coloring Axiom for Trees* (CAT):

There is an Aronszajn tree T such that for every $K \subseteq T$ there is an uncountable antichain $X \subseteq T$ such that $\wedge(X)$ is either contained in or disjoint from K .

Here $\wedge(X)$ denotes the set of all pairwise meets of elements of X .

One feature of the argument from [128] is that it relied crucially on the Mapping Reflection Principle (MRP), a strong combinatorial principle previously introduced by Moore in [127], in order to prove the properness of the appropriate forcing notion. It was shown in [127] that MRP implies the failure of \square_κ , for all $\kappa \geq \omega_1$, and therefore its consistency requires very large cardinal axioms. However, it was not clear if any large cardinals were needed for the relative consistency of CAT. Progress on this question was made by König, Larson, Moore and Veličković in [98] who reduced considerably the large cardinal assumptions in Moore's proof. They considered a statement φ which is a

¹Recall that a *Countryman line* is an uncountable linear order whose square is the union of countably many non-decreasing relations. The existence of such a linear order was proved by Shelah in [?].

form of saturation of Aronszajn trees and showed that it can be used instead of MRP in the proof of the Key Lemma (Lemma 5.29) from [128]. Moreover, they showed that for the consistency of BPFA together with φ it is sufficient to assume the existence of a reflecting Mahlo cardinal. If one is only interested in the consistency of the existence of a five element basis for the uncountable linear orderings then even an smaller large cardinal assumption is sufficient (see [128] for details).

The purpose of this chapter is to present a direct proof of CAT, and therefore the existence of a five element linear basis, assuming the conjunction of BPFA and φ . The argument is much simpler than the original proof from [128]. It is our hope that by further understanding this forcing one will be able to determine if any large cardinal assumptions are needed for the consistency of CAT.

The chapter is organized as follows. In Section 1 we present the background material on Aronszajn trees and the combinatorial principles ψ and φ . In Section 2 we start with a coherent special Aronszajn tree T and a subset K of T , define a new coloring of finite subsets of T and prove some technical lemmas. In Section 3 we define the main forcing notion $\partial^*(K)$ and show that it is proper. In Section 4 we complete the proof that BPFA together with φ implies CAT.

7.1 Saturation of Aronszajn trees

By an *Aronszajn tree* or simply an *A-tree* we mean an uncountable tree in which all levels and chains are countable. A *subtree* of an A-tree T is an uncountable downward closed subset of T . All our trees will be subtrees of $2^{<\omega_1}$ or products of such trees. If T is such a tree, $t \in T$ and $\xi < \text{ht}(t)$ then $t \upharpoonright \xi$ is the predecessor of t on level ξ of T . We start by discussing the notion of saturation of an Aronszajn tree.

Definition 7.1.1. *An Aronszajn tree T is saturated if whenever \mathcal{A} is a collection of subtrees T which have pairwise countable intersection, \mathcal{A} has cardinality at most ω_1 .*

This statement follows from the stronger assertion shown by Baumgartner in [12] to hold after Levy collapsing an inaccessible cardinal to ω_2 .

For every Aronszajn tree T , there is a collection \mathcal{B} of subtrees of T such that \mathcal{B} has cardinality ω_1 and every subtree of T contains an element of \mathcal{B} .

However, in Baumgartner's model CH holds and we need to have saturation of Aronszajn trees together with BPFA. It is for this reason that a different

approach was taken in [98]. We now recall the relevant definitions from this paper.

If \mathcal{F} is a collection of subtrees of T , then \mathcal{F}^\perp is the collection of all subtrees B of T such that for every A in \mathcal{F} , $A \cap B$ is countable. If \mathcal{F}^\perp is empty, then \mathcal{F} is said to be *predense*. For \mathcal{F} a collection of subtrees of an Aronszajn tree T , we consider the following statements:

$\psi_0(\mathcal{F})$ There is a closed unbounded set $E \subseteq \omega_1$ and a continuous chain $\langle N_\nu : \nu \in E \rangle$ of countable subsets of \mathcal{F} such that for every ν in E and t in T_ν there is a $\nu_t < \nu$ such that if $\xi \in (\nu_t, \nu) \cap E$, then there is $A \in \mathcal{F} \cap N_\xi$ such that $t \upharpoonright \xi$ is in A .

$\varphi_0(\mathcal{F})$ There is a closed unbounded set $E \subseteq \omega_1$ and a continuous chain $\langle N_\nu : \nu \in E \rangle$ of countable subsets of $\mathcal{F} \cup \mathcal{F}^\perp$ such that for every ν in E and t in T_ν either

1. there is a $\nu_t < \nu$ such that if $\xi \in (\nu_t, \nu) \cap E$, then there is $A \in \mathcal{F} \cap N_\xi$ such that $t \upharpoonright \xi$ is in A , or
2. there is a B in $\mathcal{F}^\perp \cap N_\nu$ such that t is in B .

It is not difficult to show that $\psi_0(\mathcal{F})$ implies that \mathcal{F} is predense. It is also clear that $\psi_0(\mathcal{F})$ is a Σ_1 -formula in the parameters \mathcal{F} and T . While $\varphi_0(\mathcal{F})$ and $\psi_0(\mathcal{F})$ are equivalent if \mathcal{F} is predense, $\varphi_0(\mathcal{F})$ is in general not a Σ_1 -formula in \mathcal{F} and T . Let φ be the assertion that whenever T is an A-tree and \mathcal{F} is a family of subtrees T , $\varphi_0(\mathcal{F})$ holds and let ψ be the analogous assertion but with quantification only over \mathcal{F} which are predense. As noted, φ implies ψ .

The following was proved as Corollary 3.9 in [98].

Proposition 7.1.2. *For a given family \mathcal{F} of subtrees of an Aronszajn tree T , there is a proper forcing extension which satisfies $\varphi_0(\mathcal{F})$. \square*

Remark 7.1.3. If we want to force φ it is natural to start with an inaccessible cardinal κ and do a countable support iteration of proper forcing notions $\langle \mathcal{P}_\alpha, \dot{Q}_\beta; \alpha \leq \kappa, \beta < \kappa \rangle$. At stage α we can use \diamond_κ to guess an Aronszajn tree \dot{T}_α and a family $\dot{\mathcal{F}}_\alpha$ of subtrees of \dot{T}_α in the model $V^{\mathcal{P}_\alpha}$ and let \dot{Q}_α be a \mathcal{P}_α -name for the proper poset which forces $\varphi_0(\dot{\mathcal{F}}_\alpha)$. Suppose in the final model $V^{\mathcal{P}_\kappa}$ we have an Aronszajn tree \dot{T} and a family $\dot{\mathcal{F}} \in V^{\mathcal{P}_\kappa}$ of subtrees of \dot{T} . In order to ensure that $\varphi_0(\dot{\mathcal{F}})$ holds in $V^{\mathcal{P}_\kappa}$ we need to find a stage α of the iteration at which \dot{T} and $\dot{\mathcal{F}}$ are guessed, i.e. $\dot{T}_\alpha = \dot{T}$ and $\dot{\mathcal{F}} \upharpoonright V^{\mathcal{P}_\alpha} = \dot{\mathcal{F}}_\alpha$ and moreover such that

$$(\dot{\mathcal{F}}^\perp)^{V^{\mathcal{P}_\kappa}} \upharpoonright V^{\mathcal{P}_\alpha} = (\dot{\mathcal{F}}_\alpha^\perp)^{V^{\mathcal{P}_\alpha}}.$$

This is the reason why a Mahlo cardinal is used in the following theorem from [98].

Theorem 7.1.4. *If there is a cardinal which is both reflecting and Mahlo, then there is a proper forcing extension of L which satisfies the conjunction of BPFA and φ . In particular the forcing extension satisfies that the uncountable linear orders have a five element basis.* \square

If one is interested only in the consistency that the uncountable linear order have a five element basis it was observed in [98] then a somewhat smaller large cardinal is sufficient. Indeed, for the desired conclusion one does not need the full strength of BPFA and one only needs $\varphi_0(\mathcal{F})$ for certain families of subtrees of an Aronszajn tree T which are Σ_1 -definable using a subset of ω_1 as a parameter. The precise large cardinal assumption is that there is an inaccessible cardinal κ such that for every $\kappa_0 < \kappa$, there is an inaccessible cardinal $\delta < \kappa$ such that κ_0 is in $H(\delta)$ and $H(\delta)$ satisfies there are two reflecting cardinals which are greater than κ_0 .

7.2 Colorings of Aronszajn trees

For the remainder of the paper we fix an Aronszajn tree $T \subseteq 2^{<\omega_1}$ which is coherent, special, and closed under finite modifications. The tree $T(\varrho_3)$ from [173] is such an example. Recall that T is *coherent* if for every $s, t \in T$ of the same height, say α , the set

$$D(s, t) = \{\xi < \alpha : s(\xi) \neq t(\xi)\}$$

is finite. By T_α we denote the α -th level of T , i.e. the set of nodes of height α . For $A \subseteq \omega_1$ we set $T \upharpoonright A = \bigcup_{\alpha \in A} T_\alpha$. If s and t are incomparable nodes in T , i.e. if $D(s, t)$ is non empty, we let

$$\Delta(s, t) = \min D(s, t).$$

We also let $s \wedge t$ denote the largest common initial segment of s and t , i.e. $s \upharpoonright \Delta(s, t)$. Given a subset X of T we let

$$\wedge(X) = \{s \wedge t : s, t \in X, s \text{ and } t \text{ incomparable}\}.$$

Note that if R_X is the tree induced by X , i.e. the set of all initial segments of elements of X , then $\wedge(R_X) = \wedge(X)$. We also let

$$\pi(X) = \{t \upharpoonright \text{ht}(s) : s, t \in X \text{ and } \text{ht}(s) \leq \text{ht}(t)\}.$$

We let $\text{lev}(X) = \{\text{ht}(t) : t \in X\}$. If $\alpha \in \text{lev}(X)$ we let $\pi_\alpha(X) = \pi(X) \cap T_\alpha$.

We will also need to consider finite powers of our tree T . Given an integer n and a level T_α of T we let

$$T_\alpha^{[n]} = \{\tau \in T_\alpha^n : i < j \rightarrow \tau(i) \leq_{lex} \tau(j)\}$$

where \leq_{lex} denotes the lexicographic ordering of T . We let $T^{[n]} = \bigcup_\alpha T_\alpha^{[n]}$. Morally, elements of $T^{[n]}$ are n -element subsets of T of the same height. In order to ensure that $T^{[n]}$ is closed under taking restrictions, it is necessary to allow for n -element sets with repetition and the above definition is a formal means to accommodate this. We will abuse notation and identify elements of $T^{[n]}$ which have distinct coordinates with the set of their coordinates. In our arguments, only the range of these sequences will be relevant.

If $\sigma \in T_\alpha^{[n]}$ and $\tau \in T_\alpha^{[m]}$, for some α , then, by abusing notation, we will write $\sigma \cup \tau$ is the sequence of length $n+m$ which enumerates the coordinates of σ and τ in \leq_{lex} -increasing order counting repetitions. We will also write $\sigma \subseteq \tau$ if σ is a subsequence of τ . $T^{[n]}$ will be considered as a tree with the coordinate-wise partial order induced by T . If $\sigma \in T^{[n]}$ and $\alpha < \text{ht}(\sigma)$ we write $\sigma \upharpoonright \alpha$ for the sequence $(\sigma(i) \upharpoonright \alpha : i < n)$. If $\sigma, \tau \in T^{[n]}$ are incomparable we will let

$$\Delta(\sigma, \tau) = \min\{\alpha : \sigma(i)(\alpha) \neq \tau(i)(\alpha), \text{ for some } i < n\}$$

and we'll write $\sigma \wedge \tau$ for $\sigma \upharpoonright \Delta(\sigma, \tau)$.

For $\sigma \in T^{[n]}$ let

$$D_\sigma = \langle D(\sigma(i), \sigma(0)) : i < n \rangle$$

Suppose $\sigma, \tau \in T^{[n]}$ and $\text{ht}(\sigma) \leq \text{ht}(\tau)$. We say that the pair $\{\sigma, \tau\}$ is *regular* if $D_\tau \upharpoonright \text{ht}(\sigma) = D_\sigma$, i.e. for all $i < n$,

$$D(\tau(i), \tau(0)) \cap \text{ht}(\sigma) = D(\sigma(i), \sigma(0)).$$

Note that in this case, for all $i, j < n$,

$$\Delta(\tau(i), \sigma(i)) = \Delta(\tau(j), \sigma(j)).$$

We say that a subset X of $T^{[n]}$ is *regular* if every pair of elements of X is regular. Note that if X is regular then so is the tree R_X generated by X . A *level sequence* of $T^{[n]}$ is a sequence $\{\sigma_\alpha : \alpha \in A\}$ where A is a subset of ω_1 and $\sigma_\alpha \in T_\alpha^{[n]}$, for all $\alpha \in A$. The following is a simple application of the Δ -system lemma and the Pressing Down Lemma.

Fact 7.2.1. *Let $\mathcal{A} = \{\sigma_\alpha : \alpha \in A\}$ be a level sequence. If A is uncountable (stationary) then there is an uncountable (stationary) subset B of A such that $\mathcal{B} = \{\sigma_\alpha : \alpha \in B\}$ is regular. \square*

From now on we assume the conjunction of BPFA and φ . We are given a subset K of T and we want to find an uncountable antichain X in T such that $\wedge(X) \subseteq K$ or $\wedge(X) \cap K = \emptyset$. We will refer to K as a *coloring* of T . We first note that, for every integer n , K induces a coloring $K^{[n]}$ of $T^{[n]}$ defined by

$$K^{[n]} = K^n \cap T^{[n]}.$$

We let \mathcal{F}_n be the collection of regular subtrees R of $T^{[n]}$ such that $\wedge(R) \cap K^{[n]} = \emptyset$. The following fact is immediate by using Fact 7.2.1.

Fact 7.2.2. *If $R \in \mathcal{F}_n^\perp$ then for every uncountable antichain $X \subseteq R$ there are $\sigma, \tau \in X$ such that $\sigma \wedge \tau \in K^{[n]}$.* \square

By $\varphi_0(\mathcal{F}_n)$ we can find a club C_n in ω_1 and a continuous increasing chain $\langle N_\xi^n : \xi \in C_n \rangle$ of countable subsets of $\mathcal{F}_n \cup (\mathcal{F}_n)^\perp$ witnessing $\varphi_0(\mathcal{F}_n)$. By replacing each of the C_n by their intersection we may assume that the C_n are all the same and equal to say C . We now define a new coloring of $T^{[n]} \upharpoonright C$ as follows.

Definition 7.2.3. *A node $\sigma \in T^{[n]} \upharpoonright C$ is in $K_\varphi^{[n]}$ if, letting α be the height of σ , there exists $R \in (\mathcal{F}_n)^\perp \cap N_\alpha^n$ such that $\sigma \in R$, i.e. if σ is in case (2) of the dichotomy for $\varphi_0(\mathcal{F}_n)$. We denote $(T^{[n]} \upharpoonright C) \setminus K_\varphi^{[n]}$ by $L_\varphi^{[n]}$. We let $K_\varphi = \bigcup_n K_\varphi^{[n]}$ and $L_\varphi = \bigcup_n L_\varphi^{[n]}$.*

Remark 7.2.4. The induced coloring $T^{[n]} = K_\varphi^{[n]} \cup L_\varphi^{[n]}$, for $n < \omega$, is our analog of the notions of *acceptance* and *rejection* from [128]. The main difference is that these notions are defined [128] relative to a given countable elementary submodel of $H(\omega_2)$ whereas our colorings do not make reference to any such model. This simplifies considerably the proof of properness of the main forcing notion we define in §3.

We now note some useful facts about these induced colorings.

Fact 7.2.5. *If there is a node t in $L_\varphi^{[1]}$ whose height is a limit point of C then there is an uncountable antichain X in T such that $\wedge(X) \cap K = \emptyset$, i.e. X is homogenous for $T \setminus K$.*

Proof. Assume t is such a node and let α be the height of t . By our assumption, case (1) of the dichotomy for $\varphi_0(\mathcal{F}_1)$ holds for t . Therefore, there exists $\eta < \alpha$ such that for every $\xi \in (\eta, \alpha) \cap C$ there is a $R \in \mathcal{F}_1 \cap N_\xi^1$ such that $t \upharpoonright \xi \in R$. Since $(\eta, \alpha) \cap C$ is non empty it follows that \mathcal{F}_1 is non empty as well. Since members of \mathcal{F}_1 are precisely uncountable trees which are homogenous for $T \setminus K$, and every such tree contains an uncountable antichain, the conclusion follows. \square

Fact 7.2.6. K_φ is closed for subsequences and L_φ is closed for supersequences which are in $\bigcup_n T^{[n]} \upharpoonright C$. \square

Fact 7.2.7. If $\sigma \in K_\varphi$ and $\alpha = \text{ht}(\sigma)$ is a limit of C then there is $\eta < \alpha$ such that $\sigma \upharpoonright \xi \in K_\varphi$, for all $\xi \in (\eta, \alpha) \cap C$. Similarly for L_φ . We refer to this property as continuity of the induced coloring. \square

Fact 7.2.8. Suppose S is a stationary subset of C and $\mathcal{S} = \{\sigma_\xi : \xi \in S\}$ is a level sequence in $T^{[n]}$ consisting of elements of $K_\varphi^{[n]}$. Then there exist distinct $\xi, \eta \in S$ such that $\sigma_\xi \wedge \sigma_\eta \in K^{[n]}$.

Proof. By using Fact 7.2.1 and shrinking S we may assume that \mathcal{S} is regular. For $\xi \in S$ since $\sigma_\xi \in K_\varphi^{[n]}$ there exists a tree $R_\xi \in N_\xi^n \cap (\mathcal{F}_n)^\perp$ such that $\sigma_\xi \in R_\xi$. By the Pressing Down Lemma and shrinking S again we may assume that the trees R_ξ are all the same and equal to say R . This simply means that $\mathcal{S} \subseteq R$. Since $R \in (\mathcal{F}_n)^\perp$, by Fact 7.2.2, there are $\xi \neq \eta \in S$ such that $\sigma_\eta \wedge \sigma_\xi \in K^{[n]}$, as required. \square

Definition 7.2.9. Let $\mathcal{S} = \{\sigma_\xi : \xi \in A\}$ be a regular level sequence in $T^{[n]}$, for some integer n . Then $\mathcal{P}_\mathcal{S}$ is the poset consisting of finite subsets p of \mathcal{S} such that $\wedge(p) \cap K^{[n]} = \emptyset$, ordered by reverse inclusion.

The following lemma is the main technical result of this section.

Lemma 7.2.10. Let $\mathcal{S} = \{\sigma_\alpha : \alpha \in S\}$ and $\mathcal{Z} = \{\tau_\gamma : \gamma \in Z\}$ be two regular level sequences in $T^{[n]}$ and $T^{[m]}$ respectively such that S is a stationary subset of C and $\mathcal{S} \subseteq K_\varphi^{[n]}$. Assume that, for every $\alpha \in S$ and $\gamma \in Z$, if $\alpha < \gamma$ then

$$\sigma_\alpha \cup \tau_\gamma \upharpoonright \alpha \in L_\varphi^{[n+m]}.$$

Then $\mathcal{P}_\mathcal{Z}$ is c.c.c.

Proof. Before starting the proof, notice that by using the Pressing Down Lemma and shrinking S if necessary we may assume that there is a fixed tree $R_0 \in (\mathcal{F}_n)^\perp$ witnessing that σ_α is in $K_\varphi^{[n]}$, for all $\alpha \in S$. By shrinking S and Z if necessary we may moreover assume that for every $\alpha \in S$, $\gamma \in Z$, every $i < n$ and $j < m$, $\sigma_\alpha(i)$ and $\tau_\gamma(j)$ are incomparable in T .

Now, assume \mathcal{A} is an uncountable subset of $\mathcal{P}_\mathcal{Z}$. We need to find distinct p and q in \mathcal{A} which are compatible, i.e. such that $\wedge(p \cup q) \cap K^{[m]} = \emptyset$. By a standard Δ -system argument we can assume that all elements of \mathcal{A} have a fixed size k and are mutually disjoint. For each $\alpha \in S$ we pick an element p_α of \mathcal{A} such that $\text{ht}(\tau) \geq \alpha$, for all $\tau \in p_\alpha$.

Fix for a moment one such α . Since by our assumption $\sigma_\alpha \cup \tau \upharpoonright \alpha \in L_\varphi^{[n+m]}$, for all $\tau \in p_\alpha$, we can fix an ordinal $\eta_\alpha < \alpha$ such that for every $\xi \in (\eta_\alpha, \alpha) \cap C$ and every $\tau \in p_\alpha$ there is a tree $R \in \mathcal{F}_{n+m} \cap N_\xi^{n+m}$ such that $\sigma_\alpha \upharpoonright \xi \cup \tau \upharpoonright \xi \in R$. By applying the Pressing Down Lemma and shrinking S again we may assume that all the ordinals η_α are equal to some η_0 .

Now, for each $\alpha \in S$, fix an enumeration $\{v_\alpha^0, \dots, v_\alpha^{l_\alpha-1}\}$ of distinct elements of $\{\tau \upharpoonright \alpha : \tau \in p_\alpha\}$. We may assume that there is a fixed integer l such that $l_\alpha = l$, for all $\alpha \in S$. Moreover, by shrinking S further, we may assume that if $\alpha, \beta \in S$ are distinct then v_α^i and v_β^j are incomparable, for all $i, j < l$. For each $\alpha \in S$ let

$$F_\alpha = \{\sigma_\alpha(j) : j < n\} \cup \{v_\alpha^i(j) : i < l, j < m\}$$

and let

$$D_\alpha = \bigcup \{D(s, t) : s, t \in F_\alpha\}.$$

Then D_α is finite, so if α is a limit ordinal and we let $\xi_\alpha = \max(D_\alpha) + 1$ then $\xi_\alpha < \alpha$. By the Pressing Down Lemma and shrinking S yet again we may assume that there exists a fixed ordinal ξ , a sequence $\sigma \in T_\xi^{[n]}$, and sequences $v^i \in T_\xi^{[m]}$, for $i < l$, such that, for each $\alpha \in S$, we have:

1. $\xi_\alpha = \xi$,
2. $\sigma_\alpha \upharpoonright \xi = \sigma$,
3. $v_\alpha^i \upharpoonright \xi = v^i$, for $i < l$.

Now, notice that if $\alpha, \beta \in S$ are distinct then, for every $i < l$,

$$\Delta(v_\alpha^i, v_\beta^i) = \Delta(\sigma_\alpha, \sigma_\beta).$$

Moreover, if i and j are distinct then

$$v_\alpha^i \wedge v_\beta^j = v_\alpha^i \wedge v_\alpha^j = v_\beta^i \wedge v_\beta^j \notin K^{[m]}.$$

Therefore, p_α and p_β are compatible in \mathcal{P}_Z provided $v_\alpha^i \wedge v_\beta^i \notin K^{[m]}$, for all $i < l$. We have finally set the stage for the proof of the lemma.

Fix a sufficiently large regular cardinal θ and a countable elementary submodel M of $H(\theta)$ containing all the relevant objects and such that $\delta = M \cap \omega_1$ belongs to S . Working in M fix a countable elementary submodel N of $H(\omega_2)$ containing all the relevant objects and let $\zeta = N \cap \omega_1$. Since, by our assumption $\eta_0, C \in N$, C is a club, and $N \in M$ we have that $\zeta \in (\eta_0, \delta) \cap C$. Therefore for

each $i < l$ there exists a tree $A_i \in \mathcal{F}_{n+m} \cap N_\zeta^{n+m}$ such that $\sigma_\delta \upharpoonright \zeta \cup v_\delta^i \upharpoonright \zeta \in A_i$. Since $N_\zeta^{n+m} \subseteq N$ we know that $A_i \in N$, for all i . Let

$$H = \{\eta : \exists \alpha \in S[\alpha > \eta \wedge \forall i < l(\sigma_\alpha \upharpoonright \eta \cup v_\alpha^i \upharpoonright \eta \in A_i)]\}$$

Since all the parameters in the definition of H are in N , by elementarity of N it follows that $H \in N$. On the other hand $\zeta \in H \setminus N$, therefore H is uncountable. Fix a 1 – 1 function $f : H \rightarrow S$ with $f \in N$ such that for every $\eta \in H$, $f(\eta)$ witnesses that $\eta \in H$. Then the set $X = \{\sigma_{f(\eta)} : \eta \in H\}$ belongs to N . We also know that X is an uncountable subset of R_0 . Since $T^{[n]}$ is a special tree, by shrinking H we may assume that $Y = \{\sigma_{f(\eta)} \upharpoonright \eta : \eta \in H\}$ is an antichain in $T^{[n]}$. Since $Y \subseteq R_0$ and $R_0 \in (\mathcal{F}_n)^\perp$, by Fact 7.2.2, there are distinct $\eta, \rho \in H$ such that:

$$\sigma_{f(\eta)} \wedge \sigma_{f(\rho)} = \sigma_{f(\eta)} \upharpoonright \eta \wedge \sigma_{f(\rho)} \upharpoonright \rho \in K^{[n]}.$$

Let $\alpha = f(\eta)$ and $\beta = f(\rho)$. We claim that p_α and p_β are compatible in \mathcal{P}_Z . To see this, consider some $i < l$. We know that $\sigma_\alpha \upharpoonright \eta \cup v_\alpha^i \upharpoonright \eta$ and $\sigma_\beta \upharpoonright \rho \cup v_\beta^i \upharpoonright \rho$ belong to A_i . Therefore,

$$(\sigma_\alpha \cup v_\alpha^i) \wedge (\sigma_\beta \cup v_\beta^i) = (\sigma_\alpha \upharpoonright \eta \cup v_\alpha^i \upharpoonright \eta) \wedge (\sigma_\beta \upharpoonright \rho \cup v_\beta^i \upharpoonright \rho) \notin K^{[n+m]}.$$

Since $\sigma_\alpha \wedge \sigma_\beta \in K^{[n]}$ it follows that

$$v_\alpha^i \wedge v_\beta^i \notin K^{[m]}.$$

Since this is true for all i it follows that p_α and p_β are compatible. \square

Lemma 7.2.11 (MA_{\aleph_1}). *Let $\mathcal{S} = \{\sigma_\alpha : \alpha \in S\}$ and $\mathcal{Z} = \{\tau_\gamma : \gamma \in Z\}$ be two regular level sequences in $T^{[n]}$ and $T^{[m]}$ respectively such that S and Z are stationary subsets of C . Assume $\mathcal{S} \subseteq K_\varphi^{[n]}$ and $\mathcal{Z} \subseteq K_\varphi^{[m]}$. Then there exist $\alpha \in S$ and $\gamma \in Z$ such that $\alpha < \gamma$ and*

$$\sigma_\alpha \cup \tau_\gamma \upharpoonright \alpha \in K_\varphi^{[n+m]}.$$

Proof. For every $\gamma \in Z$ fix a tree $R_\gamma \in (\mathcal{F}_m)^\perp \cap N_\gamma^m$ such that $\tau_\gamma \in R_\gamma$, i.e. witnessing that $\tau_\gamma \in K^{[m]}$. Since Z is stationary by the Pressing Down Lemma and shrinking Z if necessary we may assume that all the R_γ are equal to some tree R . Assume towards contradiction that for every $\alpha \in S$ and $\gamma \in Z$, if $\alpha < \gamma$ then $\sigma_\alpha \cup \tau_\gamma \upharpoonright \alpha \in L_\varphi^{[n+m]}$. By Lemma 7.2.10 \mathcal{P}_Z is c.c.c. By MA_{\aleph_1} we can find an uncountable subset Y of Z such that $\tau_\alpha \wedge \tau_\beta \notin K^{[m]}$, for every distinct $\alpha, \beta \in Y$. This means that the tree R^* generated by $\{\tau_\alpha : \alpha \in Y\}$ belongs to \mathcal{F}_m . However, $R^* \subseteq R$ and R is orthogonal to all trees in \mathcal{F}_m , a contradiction. \square

7.3 The forcing $\partial^*(K)$

In this section we define a notion of forcing $\partial^*(K)$ and prove that it is proper. We then show that either there is an uncountable subset Y of T such that $\wedge(Y) \cap K = \emptyset$ or forcing with $\partial^*(K)$ adds an uncountable subset X of $T \upharpoonright C$ such that $\pi(X) = X$ and X is homogenous for K_φ . Then it will be easy to force again and obtain an uncountable subset Z of X such that $\wedge(Z) \subseteq K$. Before we start it will be convenient to define a certain club of countable elementary submodels of $H(\omega_2)$. Fix, for each $\delta < \omega_1$, a bijection $e_\delta : \omega \rightarrow T_\delta^{[n]}$.

Definition 7.3.1. \mathcal{E} is the collection of all countable elementary submodels M of $H(\omega_2)$ such that $T, C, K, \langle e_\delta : \delta < \omega_1 \rangle$ as well as $\langle N_\xi^n : \xi \in C \rangle$, for $n < \omega$, all belong to M .

We are now in the position to define the partial order $\partial^*(K)$.

Definition 7.3.2. $\partial^*(K)$ consists of all pairs (X_p, \mathcal{M}_p) such that:

1. X_p is a finite subset of $T \upharpoonright C$, $\pi(X_p) = X_p$, and $X_p \cap T_\alpha \in K_\varphi$, for all $\alpha \in \text{lev}(X_p)^2$.
2. \mathcal{M}_p is a finite \in -chain of elements of \mathcal{E} such that for every $x \in X_p$ there is $M \in \mathcal{M}_p$ such that $\text{ht}(x) = M \cap \omega_1$.

The order of $\partial^*(K)$ is the coordinatewise reverse inclusion, i.e. $q \leq p$ iff $X_p \subseteq X_q$ and $\mathcal{M}_p \subseteq \mathcal{M}_q$.

In what follows, for $p \in \partial^*(K)$, M_p^i denotes the i -th model in \mathcal{M}_p , in the enumeration induced by the heights of the models.

Theorem 7.3.3. $\partial^*(K)$ is a proper forcing notion.

Proof. Fix a countable $M \prec H(2^{|\partial^*(K)|^+})$ such that $\partial^*(K), \mathcal{E} \in M$. Given a condition $p = (X_p, \mathcal{M}_p) \in M$, we need to find $q \leq p$ that is $(\partial^*(K), M)$ -generic. Set

$$q = (X_p, \mathcal{M}_p \cup \{M \cap H(\omega_2)\}).$$

We claim that q is desired. To see this, fix a dense set $D \in M$ and a condition $r \leq q$. We need to find $s \in D \cap M$ which is compatible with r . By replacing r with a stronger condition we may assume that $r \in D$. Define

$$r' = (X_r \cap M, \mathcal{M}_r \cap M)$$

²Here, of course, we identify $X_p \cap T_\alpha$ with its \leq_{lex} -increasing enumeration.

and

$$r^* = (X_r \setminus M, \mathcal{M}_r \setminus M)$$

and suppose that $|\mathcal{M}_r \setminus M| = l$. For every $i < l$ let $\delta_i = M_{r^*}^i \cap \omega_1$, let n_i be such that $e_{\delta_i}(n_i) = X_{r^*} \cap T_{\delta_i}$ and let $k_i = |X_{r^*} \cap T_{\delta_i}|$. We now define formulas θ_i , for $i < l$, by reverse induction on i .

$\theta_l(\xi_0, \dots, \xi_{l-1})$ holds if there is a condition $s = (X_s, \mathcal{M}_s) \in \partial^*(K)$ such that:

1. $\mathcal{M}_s = \{M_s^0, \dots, M_s^{l-1}\}$,
2. $M_s^i \cap \omega_1 = \xi_i$, for all $i < l$,
3. $X_s \cap T_{\xi_i} = e_{\xi_i}(n_i)$ and $|X_s \cap T_{\xi_i}| = k_i$, for all $i < l$,
4. $(X_{r'} \cup X_s, \mathcal{M}_{r'} \cup \mathcal{M}_s) \in D$.

Suppose θ_{i+1} has been defined for some $i < l$. Then

$$\theta_i(\xi_0, \dots, \xi_{i-1}) \text{ iff } Q\eta \theta_{i+1}(\xi_0, \dots, \xi_i, \eta).$$

Here $Q\eta \theta(\eta)$ means "there are stationary many η such that $\theta(\eta)$ holds".

Remark 7.3.4. Notice that the parameters of each $\theta_i(\xi_0, \dots, \xi_i)$ are in M , so if $\xi_0, \dots, \xi_{i-1} \in M$ then $\theta_i(\xi_0, \dots, \xi_{i-1})$ holds iff it holds in M . Thus, if W_i be the set of tuples $(\xi_0, \dots, \xi_{i-1})$ such that $\theta_i(\xi_0, \dots, \xi_{i-1})$ holds then $W_i \in M$, for all $i \leq l$. We set $W = \bigcup_{i \leq l} W_i$.

Notice also that if $\theta_i(\xi_0, \dots, \xi_{i-1})$ holds then $e_{\xi_j}(n_j) \in K_\varphi^{[k_j]}$, for all $j < i$.

Claim 7.3.5. $\theta_i(\delta_0, \dots, \delta_{i-1})$ holds, for all $i \leq l$.

Proof. We prove this by reverse induction on i . Notice that $\theta_l(\delta_0, \dots, \delta_{l-1})$ holds as witnessed by the condition r^* . Suppose we have established $\theta_{i+1}(\delta_0, \dots, \delta_i)$. Since $W_{i+1} \in M$ and $M \cap H(\omega_2) = M_{r^*}^0 \subseteq M_{r^*}^i$ it follows that the set

$$Z = \{\eta : (\delta_0, \dots, \delta_{i-1}, \eta) \in W_{i+1}\}$$

also belongs to $M_{r^*}^i$. If Z were non stationary, by elementarity, there would be a club $E \in M_{r^*}^i$ disjoint from it, but $\delta_i \in Z$ and δ_i belongs to any club which is in $M_{r^*}^i$, a contradiction. \square

By Claim 7.3.5 we can pick a *stationary splitting tree* $U \subseteq W$. This means that $U \subseteq (\omega_1)^{\leq l}$ is a tree and for every node $t = (\xi_0, \dots, \xi_{i-1}) \in U$ of height $i < l$ the set

$$S_t = \{\eta : (\xi_0, \dots, \xi_{i-1}, \eta) \in U\}$$

is stationary. We can moreover assume that $U \in M$. We now build by induction an increasing sequence $(\xi_i : i < l)$ of ordinals in M such that:

1. $(\xi_0, \dots, \xi_i) \in U$, for all i ,
2. $e_{\xi_i}(n_i) \cup e_{\delta_0}(n_0) \in K_\varphi^{[k_i+k_0]}$, for all i .

Suppose $j < l$ and we have picked ξ_i , for all $i < j$. Consider the set $S_j = \{\eta : (\xi_0, \dots, \xi_{j-1}, \eta) \in U\}$.

Claim 7.3.6. *There is $\xi \in S_j \cap M$ such that $e_\xi(n_j) \cup e_{\delta_0}(n_0) \in K_\varphi^{[k_j+k_0]}$.*

Proof. Assume otherwise. We know that S_j is stationary and that the level sequence $\mathcal{S}_j = \{e_\eta(n_j) : \eta \in S_j\}$ is contained in $K_\varphi^{[k_j]}$. By shrinking S_j we may also assume that \mathcal{S}_j is regular. Let

$$Z = \{\eta : e_\eta(n_0) \in K_\varphi^{[k_0]} \wedge \forall \xi \in S_j \cap \eta [e_\xi(k_j) \cup e_\eta(k_0) \notin K_\varphi^{[k_j+k_0]}\}\}.$$

Then $Z \in M$ and since we assumed that $\delta_0 \in Z$ it follows that Z is stationary. By shrinking Z we may assume that the level sequence $\mathcal{Z} = \{e_\eta(k_0) : \eta \in Z\}$ is regular. Now, by Lemma 7.2.11 and MA_{\aleph_1} we obtain a contradiction. \square

Suppose $(\xi_0, \dots, \xi_{l-1})$ has been constructed. Since $(\xi, \dots, \xi_{l-1}) \in U \cap M$, by elementarity there is a condition $s \in \partial^*(K) \cap M$ witnessing this fact. Let

$$\bar{s} = (X_{r'} \cup X_s, \mathcal{M}_{r'} \cup \mathcal{M}_s).$$

Then by (4) in the statement of $\theta_l(\xi_0, \dots, \xi_{l-1})$ we know that $\bar{s} \in D$. Since s and r' are both in M so is \bar{s} . We claim that \bar{s} is compatible with r . To see this we define a condition u as follows. Let

$$X_u = \pi(X_r \cup X_s).$$

Note that $\text{lev}(X_u) = \text{lev}(X_{r'}) \cup \text{lev}(X_s) \cup \text{lev}(X_{r^*})$ and we have

$$X_u \cap T_\alpha = \begin{cases} e_{\delta_i}(n_i) & \text{if } \alpha = \delta_i, \text{ for some } i < l, \\ e_{\xi_i}(n_i) \cup e_{\delta_0}(n_0) & \text{if } \alpha = \xi_i, \text{ for some } i < l, \\ X_{r'} \cap T_\alpha & \text{if } \alpha \in \text{lev}(X_{r'}). \end{cases}$$

In all cases we have that $X_u \cap T_\alpha \in K_\varphi$. We let $\mathcal{M}_u = \mathcal{M}_{r'} \cup \mathcal{M}_s \cup \mathcal{M}_{r^*}$. It follows that $u \leq \bar{s}, r$. This completes the proof of Theorem 7.3.3. \square

7.4 The main theorem

In this section we complete the proof of the main theorem saying that the conjunction of BPFA and φ implies CAT. Let G be V -generic over the poset $\partial^*(K)$ and define in $V[G]$:

$$X_G = \bigcup \{X_p : p \in G\}.$$

Note that $\pi(X_G) = X_G$, $\text{lev}(X) \subseteq C$, and every finite subset of X_G contained in one level of T is in K_φ . Let \dot{X}_G be a canonical $\partial^*(K)$ -name for X_G . We first establish the following fact in the ground model V .

Lemma 7.4.1. *Suppose there is no uncountable antichain Y in T such that $\wedge(Y) \cap T = \emptyset$. Then there is a condition $p \in \partial^*(K)$ which forces that \dot{X}_G is uncountable.*

Proof. Suppose the maximal condition forces that \dot{X}_G is countable. Let θ be a sufficiently large regular cardinal and let M be a countable elementary submodel of $H(\theta)$ containing all the relevant objects. As shown in Theorem 7.3.3 $q = (\emptyset, \{M \cap H(\omega_2)\})$ is an $(M, \partial^*(K))$ -generic condition. Therefore, $q \Vdash \dot{X}_G \subseteq M$. If there is a node t in $K_\varphi^{[1]} \cap T_\delta^{[1]}$ then $r = (\{t\}, \{M \cap H(\omega_2)\})$ is a condition stronger than q and $r \Vdash t \in \dot{X}_G$, a contradiction. Assume now that $T_\delta^{[1]} \subseteq L_\varphi^{[1]}$. Since δ is a limit point of C , by Fact 7.2.5 there is an uncountable antichain Y in T such that $\wedge(Y) \cap K = \emptyset$, as desired. \square

Now, assume there is no uncountable antichain $Y \subseteq T$ such that $\wedge(Y) \cap K = \emptyset$ and fix a V -generic G over $\partial^*(K)$ containing a condition as in Lemma 7.4.1. We work in $V[G]$. We can show that $\text{lev}(X_G)$ is a club, but this is not necessary. Namely, let \bar{X}_G be the closure of X_G in the tree topology. Then by Fact 7.2.7 all finite subsets of \bar{X}_G contained in one level of T are in K_φ . Moreover, $\text{lev}(\bar{X}_G)$ is equal to the closure of $\text{lev}(X_G)$ in the order topology and is a club. Clearly, we also have $\pi(\bar{X}_G) = \bar{X}_G$.

Remark 7.4.2. Before continuing it is important to note a certain amount of absoluteness between V and $V[G]$. In V we defined $(\mathcal{F}_n)^V$ to be the collection of subtrees R of $T^{[n]}$ such that $\wedge(R) \cap K^{[n]} = \emptyset$. The same definition in $V[G]$ gives a larger collection $\mathcal{F}_n^{V[G]}$ of subtrees of $T^{[n]}$. Nevertheless, the definition of \mathcal{F}_n is Σ_1 with parameters T and K . Since BPFA holds in V , if a certain tree $A \in V$ is in $(\mathcal{F}_n^V)^\perp$ then it is also in $(\mathcal{F}_n^{V[G]})^\perp$. It follows that the same sequences $\langle N_\xi^n : \xi \in C \rangle$ witness $\varphi(\mathcal{F}_n)$ in V and in $V[G]$, for all n . Therefore the definitions of the induced colorings $K_\varphi^{[n]}$, for $n < \omega$, are also absolute between V and $V[G]$.

Definition 7.4.3. *The poset \mathcal{Q} consists of finite antichains p in \bar{X}_G such that $\bigwedge(p) \subseteq K$, ordered by reverse inclusion.*

Claim 7.4.4. *\mathcal{Q} is a c.c.c. poset.*

Proof. Suppose \mathcal{A} is an uncountable subset of \mathcal{Q} . We need to find two elements of \mathcal{A} which are compatible. By a standard Δ -system argument we may assume that the elements of \mathcal{A} are disjoint and have the same size. For each $\alpha \in \text{lev}(\bar{X}_G)$ choose $p_\alpha \in \mathcal{A}$ such that $\text{ht}(t) \geq \alpha$, for all $t \in p$. We can assume that the p_α are distinct. Let σ_α be the enumeration in \leq_{lex} -increasing order of distinct elements of $\{t \upharpoonright \alpha : t \in p_\alpha\}$. There is a stationary subset S of $\text{lev}(\bar{X}_G)$ and an integer n such that σ_α has size n , for all $\alpha \in S$. Note that $\sigma_\alpha \in K_\varphi^{[n]}$, for all $\alpha \in S$. By shrinking S further we may assume that $\{\sigma_\alpha : \alpha \in S\}$ is a regular level sequence and that for every $\alpha, \beta \in S$ and every distinct $i, j < n$

$$\sigma_\alpha(i) \wedge \sigma_\beta(j) = \sigma_\alpha(i) \wedge \sigma_\alpha(j) \in K.$$

Now, by Fact 7.2.8 we can find distinct $\alpha, \beta \in S$ such that $\sigma_\alpha \wedge \sigma_\beta \in K^{[n]}$, i.e. $\sigma_\alpha(i) \wedge \sigma_\beta(i) \in K$, for all $i < n$. It follows that p_α and p_β are compatible in \mathcal{Q} . \square

By a standard argument there is a condition $q \in \mathcal{Q}$ which forces the generic H to be uncountable. Therefore, by forcing with \mathcal{Q} below q over $V[G]$ we obtain an uncountable antichain H of T such that $\bigwedge(H) \subseteq K$. Since $\partial^*(K) * \mathcal{Q}$ is proper, by BPFA, we have such an antichain in V . Thus, we have proved the main theorem which we now state.

Theorem 7.4.5. *Assume BPFA and φ . Then CAT holds and hence there is a five element basis for the uncountable linear orders.* \square

Chapter 8

Generalized side conditions

We present a generalization of the method of model as side conditions. Generally speaking a poset that uses models as side conditions is a notion of forcing whose elements are pairs, consisting of a working part which is some partial information about the object we wish to add and a finite \in -chain of countable elementary substructures of $H(\theta)$, for some cardinal θ i.e. the structure consisting of sets whose transitive closure has cardinality less than θ . The models in the side condition are used to control the extension of the working part. This is crucial in showing some general property of the forcing such as properness.

The generalization we now present amounts to allowing also certain uncountable models in the side conditions. This allows us to preserve both \aleph_1 and \aleph_2 . This approach was introduced by Neeman [130] who used it to give an alternative proof of the consistency of PFA and also to obtain generalizations of PFA to higher cardinals. In Section 1 we present the two-type poset of pure side conditions from [130], in the case of countable models and approachable models of size ω_1 , and work out the details of some of its main properties that were mentioned in [130]. The remainder of the paper is devoted to applications. We will be primarily interested in adding certain combinatorial objects of size \aleph_2 . These results were known by other methods but we believe that the present method is more efficient and will have other applications. In Section 2 we present a version of the forcing for adding a club in ω_2 with finite conditions, preserving ω_1 and ω_2 . This fact has been shown to be consistent with ZFC independently by Friedman ([40]) and Mitchell ([123]) using more complicated notions of forcing. In Section 3 we show how to add a chain of length ω_2 in the structure $(\omega_1^{\omega_1}, <_{\text{fin}})$. This result is originally due to Koszmider [100]. In Section 4 we give another proof of a result of Baumgartner and Shelah [10] by using side condition forcing to add a thin very tall superatomic Boolean algebra. Finally, in Section 5 we

show how to force an ω_2 -Souslin tree with finite condition.

8.1 The forcing \mathbb{M}

In this section we present the forcing consisting of pure side conditions. Our presentation follows [130], but we only consider side conditions consisting of models which are either countable or of size \aleph_1 . We consider the structure $(H(\aleph_2), \in, \trianglelefteq)$ equipped with a fixed well-ordering \trianglelefteq . In this way we have definable Skolem functions, so if M and N are elementary submodels of $H(\aleph_2)$ then so is $M \cap N$.

Definition 8.1.1. *Let P an elementary submodel of $H(\aleph_2)$ of size \aleph_1 . We say that P is internally approachable if it can be written as the union of an increasing continuous \in -chain $\langle P_\xi : \xi < \omega_1 \rangle$ of countable elementary submodels of $H(\aleph_2)$ such that $\langle P_\xi : \xi < \eta \rangle \in P_{\eta+1}$, for every ordinal $\eta < \omega_1$.*

If P is internally approachable of size \aleph_1 we let \vec{P} denote the least \trianglelefteq -chain witnessing this fact and we write P_ξ for the ξ -th element of this chain. Note also that in this case $\omega_1 \subseteq P$.

Definition 8.1.2. *We let \mathcal{E}_0^2 denote the collection of all countable elementary submodels of $H(\aleph_2)$ and \mathcal{E}_1^2 the collection of all internally approachable elementary submodels of $H(\aleph_2)$ of size \aleph_1 . We let $\mathcal{E}^2 = \mathcal{E}_0^2 \cup \mathcal{E}_1^2$.*

The following fact is well known.

Fact 8.1.3. *The set \mathcal{E}_1^2 is stationary in $[H(\aleph_2)]^{\aleph_1}$. □*

We are now ready to define the forcing notion \mathbb{M} consisting of pure side conditions.

Definition 8.1.4. *The forcing notion \mathbb{M} consists of finite \in -chains $p = \mathcal{M}_p$ of models in \mathcal{E}^2 closed under intersection. The order on \mathbb{M} is reverse inclusion, i.e. $q \leq p$ if $\mathcal{M}_p \subseteq \mathcal{M}_q$.*

Suppose M and N are elements of \mathcal{E}^2 with $M \in N$. If $|M| \leq |N|$ then $M \subseteq N$. However, if M is of size \aleph_1 and N is countable then the \trianglelefteq -least chain \vec{M} witnessing that M is internally approachable belongs to N and so $M \cap N = M_{\delta_N}$, where $\delta_N = N \cap \omega_1$ and M_{δ_N} is the δ_N -th member of \vec{M} .

We can split every condition in \mathbb{M} in two parts: the models of size \aleph_0 and the models of size \aleph_1 .

Definition 8.1.5. *For $p \in \mathbb{M}$ let $\pi_0(p) = p \cap \mathcal{E}_0^2$ and $\pi_1(p) = p \cap \mathcal{E}_1^2$.*

Let us see some structural property of the elements of \mathbb{M} . First, let \in^* be the transitive closure of the \in relation, i.e. $x \in^* y$ if $x \in \text{tcl}(y)$. Clearly, if $p \in \mathbb{M}$ then \in^* is a total ordering on \mathcal{M}_p . Given $M, N \in \mathcal{M}_p \cup \{\emptyset, H(\aleph_2)\}$ with $M \in^* N$ let

$$(M, N)_p = \{P \in \mathcal{M}_p : M \in^* P \in^* N\}.$$

We let $(M, N]_p = (M, N)_p \cup \{N\}$, $[M, N)_p = (M, N)_p \cup \{M\}$ and $[M, N]_p = (M, N)_p \cup \{M, N\}$. Given a condition $p \in \mathbb{M}$ and $M \in p$ we let $p \upharpoonright M$ denote the restriction of p to M , i.e. $\mathcal{M}_p \cap M$.

Fact 8.1.6. *Suppose $p \in \mathbb{M}$ and $N \in \pi_1(p)$. Then $\mathcal{M}_p \cap N = (\emptyset, N)_p$. Therefore, $p \cap N \in \mathbb{M}$. \square*

Fact 8.1.7. *Suppose $p \in \mathbb{M}$ and $M \in \pi_0(p)$. Then*

$$\mathcal{M}_p \cap M = \mathcal{M}_p \setminus \bigcup \{[M \cap N, N)_p : N \in (\pi_1(p) \cap M) \cup \{H(\aleph_2)\}\}.$$

Therefore, $p \cap M \in \mathbb{M}$. \square

The next lemma will be used in the proof of properness of \mathbb{M} .

Lemma 8.1.8. *Suppose $M \in \mathcal{E}^2$ and $p \in \mathbb{M} \cap M$. Then there is a new condition p^M , which is the smallest element of \mathbb{M} extending p and containing M as an element.*

Proof. If $M \in \mathcal{E}_1^2$ we can simply let

$$p^M = \mathcal{M}_{p^M} = \mathcal{M}_p \cup \{M\}.$$

If $M \in \mathcal{E}_0^2$ we close $\mathcal{M}_p \cup \{M\}$ under intersections and show that it is still an \in -chain. First of all notice that, since p is finite and belongs to M , we have $\mathcal{M}_p \subseteq M$. For this reason if $P \in \pi_0(p)$, then $P \cap M = P$. On the other hand, if $P \in \pi_1(p)$, by the internal approachability of P and the fact that $P \in M$ we have that $P \cap M \in P$. Now, if $N \in P$ is the \in^* -greatest element of \mathcal{M}_p below P , then $N \in P \cap M$, since $\mathcal{M}_p \subseteq M$. Finally the \in^* -greatest element of \mathcal{M}_p belongs to M , since \mathcal{M}_p does. \square

In the forcing notions we consider in this chapter if two conditions p and q are compatible then this will be witnessed by a condition r which is Σ_0 -definable from p and q . Thus, all elements of \mathcal{E}^2 will be adequate for the appropriate forcing notions. Our goal is to show that \mathbb{M} is \mathcal{E}^2 -strongly proper. We will need the following.

Lemma 8.1.9. *Suppose $r \in \mathbb{M}$ and $M \in \mathcal{M}_r$. Let $q \in M$ be such that $q \leq r \cap M$. Then q and r are compatible.*

Proof. If M is uncountable then one can easily check that $\mathcal{M}_s = \mathcal{M}_q \cup \mathcal{M}_r$ is an \in -chain which is closed under intersection. Therefore $s = \mathcal{M}_s$ is a common extension of q and r . Suppose now M is countable. We first check that $\mathcal{M}_q \cup \mathcal{M}_r$ is an \in -chain, then we close this chain under intersections and show that the resulting set is still an \in -chain.

Claim 8.1.10. *The set $\mathcal{M}_q \cup \mathcal{M}_r$ is an \in -chain.*

Proof. Note that any model of $\mathcal{M}_r \setminus M$ is either in $[M, H(\aleph_2))_r$ or belongs to an interval of the form $[N \cap M, N)_r$, for some $N \in \pi_1(r \upharpoonright M)$. Consider one such interval $[N \cap M, N)_r$. Since $N \in r \upharpoonright M$ and $q \leq r \upharpoonright M$ we have that $N \in \mathcal{M}_q$. The models in $\mathcal{M}_r \cap [N \cap M, N)_r$ are an \in -chain. The least model on this chain is $N \cap M$ and the last one belongs to N . Consider the \in^* -largest model P of \mathcal{M}_q below N . Since $q \in M$ we have that $P \in M$. Moreover, since \mathcal{M}_q is an \in -chain we have that also $P \in N$, therefore $P \in N \cap M$. Similarly, the least model of \mathcal{M}_r in $[M, H(\aleph_2))_r$ is M and it contains the top model of \mathcal{M}_q . Therefore, $\mathcal{M}_q \cup \mathcal{M}_r$ is an \in -chain. \square

We now close $\mathcal{M}_q \cup \mathcal{M}_r$ under intersections and check that it is still an \in -chain. We let $Q \in \mathcal{M}_q \setminus \mathcal{M}_r$ and consider models of the form $Q \cap R$, for $R \in \mathcal{M}_r$.

Case 1: $Q \in \pi_0(q)$. We show by \in^* -induction on R that $Q \cap R$ is already on the chain \mathcal{M}_q . Since $Q \in M$ and Q is countable we have that $Q \subseteq M$. Therefore, $Q \cap R = Q \cap (R \cap M)$. We know that $R, M \in \mathcal{M}_r$ and \mathcal{M}_r is closed under intersections, so $R \cap M \in \mathcal{M}_r$. By replacing R by $R \cap M$ we may assume that R is countable and below M in \mathcal{M}_r . If $R \in M$ then $R \in \mathcal{M}_q$ and \mathcal{M}_q is closed under intersection, so $Q \cap R \in \mathcal{M}_q$. If $R \in \mathcal{M}_r \setminus M$ then it belongs to an interval of the form $[N \cap M, N)_r$, for some $N \in \pi_1(r \upharpoonright M)$. Since N is uncountable and $R \in^* N$ it follows that $R \subseteq N$. If there is no uncountable model in the interval $[N \cap M, R)_r$ then we have that $N \cap M \subseteq R \subseteq N$. It follows that

$$Q \cap (N \cap M) \subseteq Q \cap R \subseteq Q \cap N.$$

However, Q is a subset of M and so $Q \cap (N \cap M) = Q \cap N$. Therefore, $Q \cap R = Q \cap N$ and since $Q, N \in \mathcal{M}_q$ we have again that $Q \cap N \in \mathcal{M}_q$. Now, suppose there is an uncountable model in $[N \cap M, R)_r$ and let S be the \in^* -largest such model. Since all the models in the interval $(S, R)_r$ are countable we have that $S \in R$. On the other hand, S is uncountable and above $N \cap M$ in \mathcal{M}_r . It follows

that $N \cap M \subseteq S$. Now, consider the model $R^* = R \cap S$. It is below S in \mathcal{M}_r . We claim that $Q \cap R = Q \cap R^*$. To see this note that, since $Q \subseteq M$ and $R \subseteq N$, we have

$$Q \cap R \subseteq Q \cap (N \cap M) \subseteq Q \cap S.$$

Therefore, $Q \cap R^* = Q \cap (R \cap S) = Q \cap R$. Since R^* is below R in \mathcal{M}_r , by the inductive assumption, we have that $Q \cap R^* \in \mathcal{M}_q$.

Case 2: $Q \in \pi_1(q)$. We first show that the largest element of $\mathcal{M}_q \cup \mathcal{M}_r$ below Q is in \mathcal{M}_q . To see this note that by Fact 8.1.7 any model, say S , in $\mathcal{M}_r \setminus M$ which is below M under \in^* belongs to an interval of the form $[N \cap M, N)_r$, for some $N \in \pi_1(r \upharpoonright M)$. By our assumption, $Q \in \mathcal{M}_q \setminus \mathcal{M}_r$ so N is distinct from Q . Since $N, Q \in \mathcal{M}_q$ and they are both uncountable it follows that either $Q \in N$ or $N \in Q$. In the first case, $Q \in N \cap M$, i.e. Q is \in^* -below S . In the second case, $S \in^* N \in^* Q$ and $N \in M$.

We now consider models of the form $Q \cap R$, for $R \in \mathcal{M}_r$. If R is uncountable then either $Q \subseteq R$ or $R \subseteq Q$ so $Q \cap R$ is in $\mathcal{M}_q \cup \mathcal{M}_r$. If R is countable and below Q on the chain $\mathcal{M}_q \cup \mathcal{M}_r$ then $R \subseteq Q$, so $Q \cap R = R$. If $R \in \mathcal{M}_r \cap M$ then $R \in \mathcal{M}_q$ and since \mathcal{M}_q is closed under intersections we have that $Q \cap R \in \mathcal{M}_q$. So, suppose $R \in \pi_0(r) \setminus M$. By Fact 8.1.7 we know that R is either in $[M, H(\aleph_2))_r$ or in $[N \cap M, N)_r$, for some $N \in \pi_1(r \upharpoonright M)$. We show by \in^* -induction that $Q \cap R$ is either in $\mathcal{M}_q \cup \mathcal{M}_r$ or is equal to Q_{δ_R} and moreover $\delta_R \geq \delta_M$. Consider the case $R \in [M, H(\aleph_2))_r$. If there is no uncountable S in the interval $(M, R)_r$ then $M \subseteq R$. Therefore, $Q \in R$ and $\delta_R \geq \delta_M$. Since $Q \in R$ then $Q \cap R = Q_{\delta_R}$. If there is an uncountable model in the interval $(M, R)_r$ let S be the largest such model. Since Q is below S in the $\mathcal{M}_q \cup \mathcal{M}_r$ chain we have $Q \subseteq S$, so if we let $R^* = R \cap S$, then $Q \cap R^* = Q \cap R$, and moreover $\delta_{R^*} = \delta_R$. Therefore, we can use the inductive hypothesis for R^* . The case when R belongs to an interval of the form $[N \cap M, N)_r$, for some $N \in \pi_1(r \upharpoonright M) \cup \{H(\aleph_2)\}$ is treated in the same way.

The upshot of all of this is that when we close $\mathcal{M}_q \cup \mathcal{M}_r$ under intersections the only new models we add are of the form Q_ξ , for $Q \in \pi_1(\mathcal{M}_q \setminus \mathcal{M}_r)$, and finitely many countable ordinals $\xi \geq \delta_M$. These models form an \in -chain, say C_Q . In particular, the case $R = M$ falls under the last case of the previous paragraph, therefore $Q_{\delta_M} = Q \cap M$ is the \in^* -least member of C_Q . Moreover, if Q' is the predecessor of Q in $\mathcal{M}_q \cup \mathcal{M}_r$, then Q' belongs to both Q and M and hence it belongs to Q_{δ_M} . The largest member of C_Q is a member of Q since it is of the form Q_ξ , for some countable ξ . Thus, adding all these chains to $\mathcal{M}_q \cup \mathcal{M}_r$ we preserve the fact that we have an \in -chain. \square

As an immediate consequence of Lemma 8.1.9 we have the following.

Theorem 8.1.11. \mathbb{M} is \mathcal{E}^2 -strongly proper.

Proof. Suppose $M \in \mathcal{E}^2$ and $p \in M \cap \mathbb{M}$. We shall show that p^M is (M, \mathbb{M}) -strongly generic. To see this we for every condition $r \leq p^M$ we have to define a condition $r|M \in \mathbb{M} \cap M$ such that for every $q \in \mathbb{M} \cap M$ if $q \leq r|M$ then q and r are compatible. If we let $r|M$ simply be $r \cap M$ this is precisely the statement of Lemma 8.1.9. \square

Corollary 8.1.12. The forcing \mathbb{M} is proper and preserves ω_2 . \square

8.2 Adding a club in ω_2 with finite conditions

We now present a version of the Friedman-Mitchell (see [40] and [123]) forcing for adding a club to ω_2 with finite conditions. This will be achieved by adding a working part to the side conditions.

Definition 8.2.1. Let \mathbb{M}_2 be the forcing notion whose elements are triples $p = (F_p, A_p, \mathcal{M}_p)$, where $F_p \in [\omega_2]^{<\omega}$, A_p is a finite collection of intervals of the form $(\alpha, \beta]$, for some $\alpha, \beta < \omega_2$, $\mathcal{M}_p \in \mathbb{M}$, and

1. $F_p \cap \bigcup A_p = \emptyset$,
2. if $M \in \mathcal{M}_p$ and $I \in A_p$, then either $I \in M$ or $I \cap M = \emptyset$.

The order on \mathbb{M}_2 is coordinatewise reverse inclusion, i.e. $q \leq p$ if $F_p \subseteq F_q$, $A_p \subseteq A_q$ and $\mathcal{M}_p \subseteq \mathcal{M}_q$.

The information carried by a condition p is the following. The points of F_p are going to be in the generic club, and the intervals in A_p are a partial description of the complement of that club. The side conditions are there to ensure that the forcing is \mathcal{E}^2 -strongly proper. It should be pointed out that a condition r may force some ordinals to be in the generic club even though they are not explicitly in F_r . The reason is that we may not be able to exclude them by intervals which satisfy conditions (1) and (2) of Definition 8.2.1.

Fact 8.2.2. If $p \in \mathbb{M}_2$ and $M \in \mathcal{M}_p$ then $\sup(M \cap \omega_2) \notin \bigcup A_p$.

Proof. Any interval I which contains $\sup(M \cap \omega_2)$ would have to intersect M without being an element of M . This contradicts condition (2) of Definition 8.2.1. \square

Fact 8.2.3. Suppose $p \in \mathbb{M}_2$, $M \in \mathcal{M}_p$ and $\gamma \in F_p \setminus M$. Then

$$\min(M \setminus \gamma), \sup(M \cap \gamma) \notin \bigcup A_p.$$

Proof. Suppose $\gamma \in F_p$ and let $I \in A_p$. Then I is of the form $(\alpha, \beta]$, for some ordinals $\alpha, \beta < \omega_2$. By condition (2) we know that either $I \cap M = \emptyset$ or $I \in M$. If $I \cap M = \emptyset$ then $\sup(M \cap \gamma), \min(M \setminus \gamma) \notin I$. Assume now that $I \in M$. Since $\gamma \notin I$ we have that either $\gamma \leq \alpha$ or $\gamma > \beta$.

Suppose first that $\gamma \leq \alpha$. Since $\alpha \in M$ it follows $\min(M \setminus \gamma) \leq \alpha$ and so $\min(M \setminus \gamma) \notin I$. Clearly, also $\sup(M \cap \gamma) \notin I$. Suppose now $\gamma > \beta$. In that case, clearly, $\min(M \setminus \gamma) \notin I$. Also, since $\beta \in M$ it follows that $\beta < \sup(M \cap \gamma)$ and so $\sup(M \cap \gamma) \notin I$. □

Definition 8.2.4. Suppose $p \in \mathbb{M}_2$ and $M \in \mathcal{M}_p$. We say that p is M -complete if

1. $\sup(N \cap \omega_2) \in F_p$, for all $N \in \mathcal{M}_p$,
2. $\min(M \setminus \gamma), \sup(M \cap \gamma) \in F_p$, for all $\gamma \in F_p$.

We say that p is complete if it is M -complete, for all $M \in \mathcal{M}_p$.

The following is straightforward.

Fact 8.2.5. Suppose $p \in \mathbb{M}_2$ and $M \in \mathcal{M}_p$. Then there is an M -complete condition q which is equivalent to p . We call the least, under inclusion, such condition the M -completion of p .

Proof. First let $F^* = F_p \cup \{\sup(N \cap \omega_2) : N \in \mathcal{M}_p\}$. Then let

$$F_q = F^* \cup \{\sup(M \cap \gamma) : \gamma \in F^*\} \cup \{\min(M \setminus \gamma) : \gamma \in F^*\}.$$

Let $A_q = A_p$ and $\mathcal{M}_q = \mathcal{M}_p$. It is straightforward to check that $q = (F_q, A_q, \mathcal{M}_q)$ is a condition equivalent to p and M -complete. □

Remark 8.2.6. Note that in the above fact q is M -complete for a single $M \in \mathcal{M}_p$. We may not be able find q which is complete, i.e. M -complete, for all $M \in \mathcal{M}_q$. To see this, suppose there are $M, N \in \mathcal{M}_p$ such that

$$\lim(M \cap N \cap \omega_2) \neq \lim(M \cap \omega_2) \cap \lim(N \cap \omega_2).$$

Note that if $\gamma \in M \cap N$ then either $M \cap \gamma \subseteq N$ or $N \cap \gamma \subseteq M$. Therefore, the least common limit of M and N which is not a limit of $M \cap N$ is above $\sup(M \cap N)$. If q is an extension of p which is complete then $\sup(M \cap N) \in F_q$, because $M \cap N \in \mathcal{M}_q$. Now, $\sup(M \cap N) \notin M \cap N$. Let us assume, for concreteness,

that $\sup(M \cap N) \notin M$. We can define inductively a strictly increasing sequence $(\gamma_n)_n$ by setting $\gamma_0 = \sup(M \cap N)$ and

$$\gamma_{n+1} = \begin{cases} \min(M \setminus \gamma_n) & \text{if } n \text{ is even} \\ \min(N \setminus \gamma_n) & \text{if } n \text{ is odd.} \end{cases}$$

Since, q was assumed to be both M -complete and N -complete we would have that $\gamma_n \in F_q$, for all n . This means that F_q would have to be infinite, which is a contradiction. We do not know if such a pair of models can exist in a condition in \mathbb{M} . Nevertheless, we will later present a variation of \mathbb{M}_2 in which this situation does not occur and in which the set of fully complete conditions is dense.

We now come back to Lemma 8.1.8 and observe that it is valid also for \mathbb{M}_2 .

Lemma 8.2.7. *Let $M \in \mathcal{E}^2$ and let $p \in \mathbb{M}_2 \cap M$. Then there is a new condition, which we will call p^M , that is the smallest element of \mathbb{M}_2 extending p such that $M \in \mathcal{M}_{p^M}$.*

Proof. If $M \in \mathcal{E}_1^2$ then simply let $p^M = (F_p, A_p, \mathcal{M}_p \cup \{M\})$. If $M \in \mathcal{E}_0^2$, then, as in Lemma 8.1.8, we let \mathcal{M}_{p^M} be the closure of $\mathcal{M}_p \cup \{M\}$ under intersection. We also let $F_{p^M} = F_p$ and $A_{p^M} = A_p$. We need to check that conditions (1) and (2) of Definition 8.2.1 are satisfied for p^M , but this is straightforward. \square

Our next goal is to show that \mathbb{M}_2 is \mathcal{E}^2 -strongly proper. We first establish the following.

Lemma 8.2.8. *Suppose $p \in \mathbb{M}_2$ and $M \in \mathcal{M}_p$. Then p is (M, \mathbb{M}_2) -strongly generic.*

Proof. We need to define, for each $r \leq p$ a restriction $r|M \in M$ such that for every $q \in M$ if $q \leq r|M$ then q and r are compatible. So, suppose $r \leq p$. By replacing r with its M -completion we may assume that r is M -complete. We define

$$r|M = (F_r \cap M, A_r \cap M, \mathcal{M}_r \cap M).$$

By Facts 8.1.7 or 8.1.6 according to whether M is countable or not we have that $\mathcal{M}_r \cap M \in \mathbb{M}$ and therefore $r|M \in \mathbb{M}_2 \cap M$. We need to show that for every $q \in M$ if $q \leq r|M$ then q and r are compatible.

If $M \in \mathcal{E}_1^2$ we already know that $\mathcal{M}_s = \mathcal{M}_q \cup \mathcal{M}_r$ is an \in -chain closed under intersection. Let $F_s = F_q \cup F_r$ and $A_s = A_q \cup A_r$. Finally, let $s = (F_s, A_s, \mathcal{M}_s)$. It is straightforward to check that s is a condition and $s \leq r, q$.

We now concentrate on the case $M \in \mathcal{E}_0^2$. We define a condition s as follows. We let $F_s = F_q \cup F_r$, $A_s = A_q \cup A_r$ and

$$\mathcal{M}_s = \mathcal{M}_q \cup \mathcal{M}_r \cup \{Q \cap R : Q \in \mathcal{M}_q, R \in \mathcal{M}_r\}.$$

We need to check that $s \in \mathbb{M}_2$. By Lemma 8.1.10 we know that \mathcal{M}_s is an \in -chain closed under intersection. Therefore we only need to check that (1) and (2) of Definition 8.2.1 are satisfied for s . First we check (1).

Claim 8.2.9. $F_s \cap \bigcup A_s = \emptyset$.

Proof. It suffices to check that $F_q \cap \bigcup A_r = \emptyset$ and $F_r \cap \bigcup A_q = \emptyset$. Suppose first $\gamma \in F_q$ and $I \in A_r$. Since $M \in \mathcal{M}_r$ we have, by (2) of Definition 8.2.1, that either $I \cap M = \emptyset$ or $I \in M$. If $I \cap M = \emptyset$ then, since $\gamma \in M$, we have that $\gamma \notin I$. If $I \in M$ then $I \in A_r \cap M$ and, since $q \leq r|_M$, it follows that $I \in A_q$. Now, q is a condition, so $\gamma \notin I$.

Suppose now $\gamma \in F_r$ and $I \in A_q$. If $\gamma \in F_r \cap M$ then $\gamma \in F_q$. Therefore $\gamma \notin I$. Suppose now $\gamma \in F_r \setminus M$. Since r is M -complete $\gamma^* = \min(M \setminus \gamma) \in F_r$. Then $\gamma^* \in F_r \cap M$ and so $\gamma^* \in F_q$. Now, $I \in M$ and so if $\gamma \in I$ then $\gamma^* \in I$, which would be a contradiction. Therefore $\gamma \notin I$. □

We now turn to condition (2) of Definition 8.2.1.

Claim 8.2.10. *If $Q \in \mathcal{M}_q$ and $I \in A_r$ then either $I \in Q$ or $I \cap Q = \emptyset$.*

Proof. Since $M \in \mathcal{M}_r$ we have that either $I \in M$ or $I \cap M = \emptyset$. If $I \in M$ then $I \in A_r \cap M$ and so $I \in A_q$. Since q is a condition we have that either $I \in Q$ or $I \cap Q = \emptyset$. So, suppose $I \cap M = \emptyset$. If $Q \in \mathcal{E}_0^2$ then $Q \subseteq M$ and so $Q \cap I = \emptyset$, as well. If $Q \in \mathcal{E}_1^2$ then $Q \cap \omega_2$ is an initial segment of ω_2 , say γ . Now, if $I \cap Q \neq \emptyset$ and $I \notin Q$ we would have that $\gamma \in I$. Since $\gamma \in M$ this contradicts the fact that $I \cap M = \emptyset$. □

Claim 8.2.11. *If $R \in \mathcal{M}_r$ and $I \in A_q$ then either $I \in R$ or $I \cap R = \emptyset$.*

Proof. Assume first that $R \in \mathcal{E}_1^2$. Then $R \cap \omega_2$ is an initial segment of ω_2 , say γ . If $I \cap R \neq \emptyset$ and $I \notin R$ then $\gamma \in I$. Now, since r is M -complete we have that $\gamma \in F_r$. If $\gamma \in M$ then $\gamma \in F_q$ and this would contradict the fact that q is a condition. If $\gamma \notin M$ let $\gamma^* = \min(M \setminus \gamma)$. Then, again by M -completeness of r , we have that $\gamma^* \in F_r$. However, $\gamma^* \in M$ and therefore $\gamma^* \in F_q$. Since $I \in A_q$ and $q \in M$ we have that $I \in M$. If $\gamma \in I$ we would also have that $\gamma^* \in I$, which contradicts the fact that q is a condition.

We now consider the case $R \in \mathcal{E}_0^2$. We will show by \in^* -induction on the chain \mathcal{M}_r that either $I \cap R = \emptyset$ or $I \in R$. If $R \in M$ then $R \in \mathcal{M}_q$ so this is clear. If $R \notin M$ then R either belongs to $[M, H(\aleph_2))_r$ or else belongs to $[N \cap M, N)_r$, for some uncountable $N \in \mathcal{M}_r \cap M$.

Suppose $R \in [N \cap M, N)_r$, for some $N \in \pi_1(\mathcal{M}_r \cap M)$. Since $I \in A_q$ and $N \in \mathcal{M}_q$ we have that $I \in N$ or $I \cap N = \emptyset$. On the other hand, $R \subseteq N$ so if $I \cap N = \emptyset$ then also $I \cap R = \emptyset$. If $I \in N$ then, since $q \in M$ and $I \in A_q$, we have that $I \in M$ and so $I \in N \cap M$. If there are no uncountable models in the interval $[N \cap M, R)_r$ then $N \cap M \subseteq R$ and so $I \in R$. If there is an uncountable model in this interval let S be the largest such model. Now, $N \cap M \subseteq S$ and so $I \in S$ and $I \subseteq S$. It follows that if $I \cap R \neq \emptyset$ then also $I \cap R \cap S \neq \emptyset$. Let $R^* = R \cap S$. Then $R^* \in \mathcal{M}_r$ and R^* is below R in the \in^* -ordering. By the inductive assumption we would have that $I \in R^*$ and so $I \in R$. The case when $R \in [M, H(\aleph_2))_r$ is treated in the same way. \square

Finally, suppose $Q \in \mathcal{M}_q$, $R \in \mathcal{M}_r$ and $I \in A_q \cup A_r$. Consider the relation between the model $Q \cap R$ and I . If I belongs to both Q and R then it belongs to $Q \cap R$. If I is disjoint from Q or R it is also disjoint from $Q \cap R$. This completes the proof that s is a condition. Since $s \leq q, r$ it follows that q and r are compatible. \square

Now, by Lemmas 8.2.7 and 8.2.8 we have the following.

Theorem 8.2.12. *The forcing \mathbb{M}_2 is \mathcal{E}^2 -strongly proper. Hence it is proper and preserves ω_2 .* \square

Suppose now G is V -generic filter for the forcing notion \mathbb{M}_2 . We can define

$$C_G = \bigcup \{F_p : p \in G\} \text{ and } U_G = \bigcup \bigcup \{A_p : p \in G\}.$$

Then $C_G \cap U_G = \emptyset$. Moreover, by genericity, $C_G \cup U_G = \omega_2$. Since U_G is a union of open intervals it is open in the order topology. Therefore, C_G closed and, again by genericity, it is unbounded in ω_2 . Unfortunately, we cannot say much about the generic club C_G . For reasons explained in Remark 8.2.6, we cannot even say that it does not contain infinite subsets which are in the ground model. In order to circumvent this problem, we now define a variation of the forcing notion \mathbb{M}_2 . We start by some definitions.

Definition 8.2.13. *Suppose $M, N \in \mathcal{E}^2$. We say that M and N are lim-compatible if*

$$\lim(M \cap N \cap \omega_2) = \lim(M \cap \omega_2) \cap \lim(N \cap \omega_2).$$

Remark 8.2.14. Clearly, this conditions is non trivial only if both M and N are countable. We will abuse notation and write $\lim(M)$ for $\lim(M \cap \omega_2)$.

We now define a version of the forcing notion \mathbb{M} .

Definition 8.2.15. *Let \mathbb{M}^* be the suborder of \mathbb{M} consisting of conditions $p = \mathcal{M}_p$ such that any two models in \mathcal{M}_p are lim-compatible.*

We have the following version of Lemma 8.1.8.

Lemma 8.2.16. *Let $M \in \mathcal{E}^2$ and let $p \in \mathbb{M}^* \cap M$. Then there is a new condition, which we will call p^M , that is the smallest element of \mathbb{M}^* extending p such that $M \in \mathcal{M}_{p^M}$.*

Proof. If $M \in \mathcal{E}_1^2$ then simply let $p^M = M_p \cup \{M\}$. If $M \in \mathcal{E}_0^2$, then we let \mathcal{M}_{p^M} be the closure of $\mathcal{M}_p \cup \{M\}$ under intersection. Then, thanks to Lemma 8.1.8, we just need to check that the models in \mathcal{M}_{p^M} are lim-compatible. Suppose $P \in \pi_0(p)$. Then $P \in M$ and hence $P \subseteq M$. Therefore, P and M are lim-compatible. Suppose now $P \in \pi_1(p)$. Then $P \cap \omega_2$ is an initial segment of ω_2 , say γ . Therefore

$$\lim(M \cap P) = \lim(M \cap \gamma) = \lim(M) \cap (\gamma + 1) = \lim(M) \cap \lim(P),$$

and so P and M are lim-compatible. We also need to check that, for any $P, Q \in \mathcal{M}_p$, the models $P \cap M$ and $Q \cap M$, as well as $P \cap M$ and Q are lim-compatible, but this is straightforward. \square

We now have a version of Lemma 8.1.9.

Lemma 8.2.17. *Suppose $r \in \mathbb{M}^*$ and $M \in \mathcal{M}_r$. Let $q \in \mathbb{M}^* \cap M$ be such that $q \leq r \cap M$. Then q and r are compatible in \mathbb{M}^* .*

Proof. If M is uncountable then one can easily check that $\mathcal{M}_s = \mathcal{M}_q \cup \mathcal{M}_r$ is \in -chain closed under intersection and that any two models in \mathcal{M}_s are lim-compatible.

Suppose now M is countable and let

$$\mathcal{M}_s = \mathcal{M}_q \cup \mathcal{M}_r \cup \{Q \cap R : Q \in \mathcal{M}_q, R \in \mathcal{M}_r\}.$$

Thanks to Lemma 8.1.9 we know that \mathcal{M}_s is an \in -chain closed under intersection. It remains to check that any two models in \mathcal{M}_s are lim-compatible.

Claim 8.2.18. *If $Q \in \pi_0(\mathcal{M}_q)$ and $R \in \pi_0(\mathcal{M}_r)$, then Q and R are lim-compatible.*

Proof. We show this by \in^* -induction on R . Since $Q \in \mathcal{M}_q$ then $Q \in M$ and, since Q is countable, we have that $\lim(Q) \subseteq M$. Moreover, since R and M are both in \mathcal{M}_r , we have that $\lim(R \cap M) = \lim(R) \cap \lim(M)$, and so

$$\lim(Q) \cap \lim(R) = \lim(Q) \cap \lim(R) \cap \lim(M) = \lim(Q) \cap \lim(R \cap M).$$

Hence, without loss of generality we can assume R to be \in^* -below M . If $R \in M$ then $R \in \mathcal{M}_q$ and so Q and R are lim-compatible. Assume now, $R \notin M$. Then by Fact 8.1.7 there is $N \in \pi_1(\mathcal{M}_r \cap M)$ such that $R \in [N \cap M, N)_r$. We may also assume Q is \in^* -below N , otherwise we could replace Q by $Q \cap N$. Hence $Q \subseteq N \cap M$. If there are no uncountable model in the interval $[N \cap M, R)_r$, then $N \cap M \subseteq R$ and since $Q \in N \cap M$ we have $Q \in R$. Therefore, Q and R are lim-compatible. Otherwise, let S be the \in^* -largest uncountable model in $[N \cap M, R)_r$. Then $Q \in S$ and $S \cap \omega_2$ is an initial segment of ω_2 . Let $R^* = R \cap S$. It follows that $\lim(R) \cap \lim(Q) = \lim(R^*) \cap \lim(Q)$. By the inductive assumption we have that $\lim(R^*) \cap \lim(Q) = \lim(R^* \cap Q)$ and hence $\lim(R) \cap \lim(Q) = \lim(R \cap Q)$. \square

Now, we need to check that any two models in \mathcal{M}_s are lim-compatible. So, suppose $S, S^* \in \mathcal{M}_s$. We may assume S and S^* are both countable and of the form $S = Q \cap R$, $S^* = Q^* \cap R^*$, for $Q, Q^* \in \mathcal{M}_q$ and $R, R^* \in \mathcal{M}_r$. Then

$$\lim((Q \cap R) \cap (Q^* \cap R^*)) = \lim((Q \cap Q^*) \cap (R \cap R^*))$$

and by Claim 8.2.18

$$\lim((Q \cap Q^*) \cap (R \cap R^*)) = \lim(Q \cap Q^*) \cap \lim(R \cap R^*),$$

because $Q \cap Q^* \in \mathcal{M}_q$ and $R \cap R^* \in \mathcal{M}_r$. Moreover, we have $\lim(Q \cap Q^*) = \lim(Q) \cap \lim(Q^*)$ and $\lim(R \cap R^*) = \lim(R) \cap \lim(R^*)$, since the elements of \mathcal{M}_q , respectively \mathcal{M}_r , are lim-compatible. Finally, again by Claim 8.2.18, we have

$$\lim(Q) \cap \lim(R) \cap \lim(Q^*) \cap \lim(R^*) = \lim(Q \cap R) \cap \lim(Q^* \cap R^*).$$

\square

We now define a variation of the forcing \mathbb{M}_2 which will have some additional properties.

Definition 8.2.19. Let \mathbb{M}_2^* be the forcing notion whose elements are triples $p = (F_p, A_p, \mathcal{M}_p)$, where $F_p \in [\omega_2]^{<\omega}$, A_p is a finite collection of intervals of the form $(\alpha, \beta]$, for some $\alpha, \beta < \omega_2$, $\mathcal{M}_p \in \mathbb{M}^*$, and

1. $F_p \cap \bigcup A_p = \emptyset$,
2. if $M \in \mathcal{M}_p$ and $I \in A_p$, then either $I \in M$ or $I \cap M = \emptyset$,

The order on \mathbb{M}_2^* is coordinatewise reverse inclusion, i.e. $q \leq p$ if $F_p \subseteq F_q$, $A_p \subseteq A_q$ and $\mathcal{M}_p \subseteq \mathcal{M}_q$.

Remark 8.2.20. Note that the only difference between \mathbb{M}_2^* and \mathbb{M}_2 is that for p to be in \mathbb{M}_2^* we require that $\mathcal{M}_p \in \mathbb{M}^*$, i.e. the models in \mathcal{M}_p are pairwise lim-compatible.

We can now use Lemmas 8.2.18 and 8.2.17 to prove the analogs of Lemmas 8.2.7 and 8.2.8 for \mathbb{M}_2^* . We then obtain the following.

Theorem 8.2.21. *The forcing notions \mathbb{M}_2^* is \mathcal{E}^2 -strongly proper. Hence, it is proper and preserves ω_2 . \square*

Let G^* is V -generic filter for \mathbb{M}_2^* . As in the case of the forcing \mathbb{M}_2 , we define

$$C_G^* = \bigcup \{F_p : p \in G^*\} \text{ and } U_G^* = \bigcup \bigcup \{A_p : p \in G^*\}.$$

As before C_G^* is forced to be a club in ω_2 . Our goal now is to show that it does not contain any infinite subset from the ground model. For this we will need the following lemma which explains the reason for the requirement of lim-compatibility for models \mathcal{M}_p , for conditions p in \mathbb{M}_2^* .

Lemma 8.2.22. *The set of complete conditions is dense in \mathbb{M}_2^* .*

Proof. Consider a condition $p \in \mathbb{M}_2^*$. For each $M \in \mathcal{M}_p$ we consider functions $\mu_M, \sigma_M : \omega_2 \rightarrow \omega_2$ defined as follows:

$$\mu_M(\alpha) = \min(M \setminus \alpha) \text{ and } \sigma_M(\alpha) = \sup(M \cap \alpha).$$

To obtain a complete condition extending p we first define:

$$F_p^* = F_p \cup \{\sup(M \cap \omega_2) : M \in \mathcal{M}_p\}.$$

We then let \bar{F}_p be the closure of F_p^* under the functions μ_M and σ_M , for $M \in \mathcal{M}_p$. Then $q = (\bar{F}_p, A_p, \mathcal{M}_p)$ will be the required complete condition extending p . The main point is to show the following.

Claim 8.2.23. *\bar{F}_p is finite.*

Proof. Let $L = \bigcup \{\lim(M) : M \in \mathcal{M}_p\}$. For each $\gamma \in L$ let

$$Y(p, \gamma) = \{M \in \mathcal{M}_p : \gamma \in \lim(M)\}.$$

and let $M(p, \gamma) = \bigcap Y(p, \gamma)$. Then, since \mathcal{M}_p is closed under intersection $M(p, \gamma) \in \mathcal{M}_p$. Since the models in \mathcal{M}_p are lim-compatible it follows that $\gamma \in \lim(M(p, \gamma))$. Thus, $M(p, \gamma)$ is the least (under inclusion) model in \mathcal{M}_p which has γ as its limit point. For each $\gamma \in L$ pick an ordinal $f(\gamma) \in M(p, \gamma) \cap \gamma$ above $\sup(F_p^* \cap \gamma)$ and $\sup(M \cap \gamma)$, for all $M \in \mathcal{M}_p \setminus Y(p, \gamma)$. For a limit $\gamma \in \omega_2 \setminus L$ let

$$f(\gamma) = \sup\{\sup(M \cap \gamma) : M \in \mathcal{M}_p\}.$$

Notice now that for any limit γ and any $M \in \mathcal{M}_p$, if $\xi \notin (f(\gamma), \gamma)$ then $\mu_M(\xi), \sigma_M(\xi) \notin (f(\gamma), \gamma)$. Since \bar{F}_p is the closure of F_p^* under the functions μ_M and σ_M , for $M \in \mathcal{M}_p$, and $F_p^* \cap (f(\gamma), \gamma) = \emptyset$, for all limit γ , it follows that $\bar{F}_p \cap (f(\gamma), \gamma) = \emptyset$, for all limit γ . This means that \bar{F}_p has no limit points and therefore is finite. \square

\square

Lemma 8.2.24. *Let $p \in \mathbb{M}_2^*$ be a complete condition, and let $\gamma \in \omega_2 \setminus F_p$. Then there is a condition $q \leq p$ such that $\gamma \in I$, for some $I \in A_q$.*

Proof. Without loss of generality we can assume that there is an $M \in \mathcal{M}_p$ such that $\sup(M \cap \omega_2) > \gamma$, otherwise we could let

$$q = (F_p, A_p \cup \{(\eta, \gamma]\}, \mathcal{M}_p),$$

for some $\eta < \gamma$ sufficiently large so that $(\eta, \gamma]$ does not intersect any model in \mathcal{M}_p .

Now, since $\sup(M \cap \omega_2) \in F_p$, for every $M \in \mathcal{M}_p$, the set $F_p \setminus \gamma$ is nonempty. Let τ be $\min(F_p \setminus \gamma)$. Notice that for every model $M \in \mathcal{M}_p$ either $\sup(M \cap \tau) < \gamma$, or $\tau \in \lim(M)$, because

$$\gamma < \sup(M \cap \tau) < \tau,$$

would contradict the minimality of τ . Moreover, if $\sup(M \cap \tau) = \gamma$, then γ would be in F_p , contrary to the hypothesis of the lemma.

Let

$$Y = \{M \in \mathcal{M}_p : \tau \in \lim(M)\}.$$

Without loss of generality we can assume $Y \neq \emptyset$, because otherwise we can let

$$q = (F_p, A_p \cup \{(\eta, \gamma]\}, \mathcal{M}_p)$$

for some η sufficiently large so that $(\eta, \gamma]$ avoids $\sup(M \cap \tau)$, for every $M \in \mathcal{M}_p$. Let $M_0 = \bigcap Y$. Since \mathcal{M}_p is closed under intersection $M_0 \in \mathcal{M}_p$. Moreover, since any two models in \mathcal{M}_p are lim-compatible we have that $\tau \in \lim M_0$. Thus, M_0 is itself in Y and is contained in any member of Y . Therefore, if an interval I belongs to M_0 , then it belongs to every model in Y . Let $\eta = \min(M_0 \setminus \gamma)$. Since $\tau \in \lim(M_0)$ we have $\gamma \leq \eta < \tau$. Since τ is the least element of F_p above γ it follows that $\eta \notin F_p$.

Claim 8.2.25. $\sup(M_0 \cap \gamma) > \sup(F_p \cap \gamma)$.

Proof. Suppose ξ is an element of $F_p \cap \gamma$. Since p is M_0 -complete, we also have $\min(M_0 \setminus \xi) \in F_p$. Notice that $\min(M_0 \setminus \xi) \neq \eta$, since $\eta \notin F_p$. Then

$$\xi \leq \min(M_0 \setminus \xi) < \gamma,$$

and so $\sup(M_0 \cap \gamma) > \xi$. □

Consider now some $M \in \mathcal{M}_p \setminus Y$. Then $\tau \notin \lim(M)$ and, since p is M -complete, we have that $\sup(M \cap \tau) \in F_p$. Since τ is the least element of F_p above γ it follows that $\sup(M \cap \tau) \in F_p \cap \gamma$. Now, pick an element $\eta' \in M_0$ above $\sup(F_p \cap \gamma)$ and let $I = (\eta', \eta]$. It follows that $I \in M$, for all $M \in Y$ and $I \cap M = \emptyset$, for all $M \in \mathcal{M}_p \setminus Y$. Therefore,

$$q = (F_p, A_p \cup \{I\}, \mathcal{M}_p)$$

is a condition stronger than p and $\gamma \in I$. Thus, q is as required. □

Corollary 8.2.26. *If G^* is a V -generic filter over \mathbb{M}_2^* , then the generic club C_G^* does not contain any infinite subset which is in V .* □

8.3 Strong chains of uncountable functions

We now consider the partial order $(\omega_1^{\omega_1}, <_{\text{fin}})$ of all functions from ω_1 to ω_1 ordered by $f <_{\text{fin}} g$ iff $\{\xi : f(\xi) \geq g(\xi)\}$ is finite. In [100] Koszmider constructed a forcing notion which preserves cardinals and adds an ω_2 chain in $(\omega_1^{\omega_1}, <_{\text{fin}})$. The construction uses an $(\omega_1, 1)$ -morass which is a stationary coding set and is quite involved. In this section we present a streamlined version of this forcing which uses generalizes side conditions and is based on the presentation of Mitchell [122]. Before that we show that Chang's conjecture implies that there is no such chain. The argument is inspired by a proof of Shelah from [153]. A similar argument appears in [99].

Proposition 8.3.1. *Assume Chang's conjecture. Then there is no chain in $(\omega_1^{\omega_1}, <_{\text{fin}})$ of length ω_2 .*

Proof. Assume towards contradiction that Chang's conjecture holds and $\{f_\alpha : \alpha < \omega_2\}$ is a chain in $(\omega_1^{\omega_1}, <_{\text{fin}})$. Given a function $g : I \rightarrow \omega_1$ and $\eta < \omega_1$ we let $\min(g, \eta)$ be the function defined by:

$$\min(g, \eta)(\zeta) = \min(g(\zeta), \eta).$$

For each $\alpha < \omega_2$, and $\xi, \eta < \omega_1$ we define a function $f_\alpha^{\xi, \eta}$ by:

$$f_\alpha^{\xi, \eta} = \min(f_\alpha \upharpoonright [\xi, \xi + \omega), \eta).$$

Given $\xi, \eta < \omega_1$, the sequence $\{f_\alpha^{\xi, \eta} : \alpha < \omega_2\}$ is \leq_{fin} -increasing. We define a club $C^{\xi, \eta} \subseteq \omega_2$ as follows.

Case 1: If the sequence $\{f_\alpha^{\xi, \eta} : \alpha < \omega_2\}$ eventually stabilizes under $=_{\text{fin}}$ we let $C^{\xi, \eta} = \omega_2 \setminus \mu$, where μ is least such that $f_\nu^{\xi, \eta} =_{\text{fin}} f_\mu^{\xi, \eta}$, for all $\nu \geq \mu$.

Case 2: If the sequence $\{f_\alpha^{\xi, \eta} : \alpha < \omega_2\}$ does not stabilize we let $C^{\xi, \eta}$ be a club in ω_2 such that $f_\alpha^{\xi, \eta} \not\leq_{\text{fin}} f_\beta^{\xi, \eta}$, for all $\alpha, \beta \in C^{\xi, \eta}$ with $\alpha < \beta$. This means that for every such α and β the set

$$\{n : f_\alpha(\xi + n) < f_\beta(\xi + n) \leq \eta\}$$

is infinite.

Let $C = \bigcap \{C^{\xi, \eta} : \xi, \eta < \omega_1\}$. Then C is a club in ω_2 . We define a coloring $c : [C]^2 \rightarrow \omega_1$ by

$$c\{\alpha, \beta\} < = \max\{\xi : f_\alpha(\xi) \geq f_\beta(\xi)\}.$$

By Chang's conjecture we can find an increasing ω_1 sequence $S = \{\alpha_\rho : \rho < \omega_1\}$ of elements of C such that $c[[S]^2]$ is bounded in ω_1 . Let $\xi = \sup(c[[S]^2]) + 1$. Therefore for every $\rho < \tau < \omega_1$ we have

$$f_{\alpha_\rho} \upharpoonright [\xi, \omega_1) < f_{\alpha_\tau} \upharpoonright [\xi, \omega_1).$$

Now, let $\eta = \sup(\text{ran}(f_{\alpha_1} \upharpoonright [\xi, \xi + \omega)))$. It follows that for every n :

$$f_{\alpha_0}(\xi + n) < f_{\alpha_1}(\xi + n) \leq \eta.$$

Since $\alpha_0, \alpha_1 \in C^{\xi, \eta}$ it follows that $C^{\xi, \eta}$ was defined using Case 2. Therefore the sequence $\{f_{\alpha_\rho}^{\xi, \eta} : \rho < \omega_1\}$ is \leq -increasing and $f_{\alpha_\rho} \neq_{\text{fin}} f_{\alpha_\tau}$, for all $\rho < \tau$. For each $\rho < \omega_1$ let n_ρ be the least such that $f_{\alpha_\rho}^{\xi, \eta}(\xi + n_\rho) < f_{\alpha_{\rho+1}}^{\xi, \eta}(\xi + n_\rho)$. Then there is a integer n such that $X = \{\rho < \omega_1 : n_\rho = n\}$ is uncountable. It follows that the sequence $\{f_{\alpha_\rho}(\xi + n) : \rho \in X\}$ is strictly increasing. On the other hand it is included in η which is countable, a contradiction. \square

Therefore, in order to add a strong ω_2 -chain in $(\omega_1^{\omega_1}, <_{\text{fin}})$ we need to assume that Chang's conjecture does not hold. In fact, we will assume that there is an increasing function $g : \omega_1 \rightarrow \omega_1$ such that

1. $g(\xi)$ is indecomposable, for all $\xi < \omega_1$,
2. o.t. $(M \cap \omega_2) < g(\delta_M)$, for all $M \in \mathcal{E}_0^2$.

It is easy to add such a function by a preliminary forcing. For instance, we can add by countable conditions an increasing function g which dominates all the canonical functions c_α , for $\alpha < \omega_2$, and such that $g(\xi)$ is indecomposable, for all ξ . Moreover, we may assume that g is definable in the structure $(H(\aleph_2), \in, \leq)$ and so it belongs to M , for all $M \in \mathcal{E}^2$.

Our plan is to add an ω_2 -chain $\{f_\alpha : \alpha < \omega_2\}$ in $(\omega_1^{\omega_1}, <_{\text{fin}})$ below this function g . We can view this chain as a single function $f : \omega_2 \times \omega_1 \rightarrow \omega_1$. We want to use conditions of the form $p = (f_p, \mathcal{M}_p)$, where $f_p : A_p \times F_p \rightarrow \omega_1$ for some finite $A_p \subseteq \omega_2$ and $F_p \subseteq \omega_1$, and $\mathcal{M}_p \in \mathbb{M}$ is a side condition. Suppose $\alpha, \beta \in A_p$ with $\alpha < \beta$, and $M \in \pi_0(\mathcal{M}_p)$. Then M should localize the disagreement of f_α and f_β , i.e. p should force that the finite set $\{\xi : f_\alpha(\xi) \geq f_\beta(\xi)\}$ is contained in M . This means that if $\xi \in \omega_1 \setminus M$ then p makes the commitment that $f_\alpha(\xi) < f_\beta(\xi)$. Moreover, for every $\eta \in (\alpha, \beta) \cap M$ we should have that $f_\alpha(\xi) < f_\eta(\xi) < f_\beta(\xi)$. Therefore, p imposes that $f_\beta(\xi) \geq f_\alpha(\xi) + \text{o.t.}((\alpha, \beta) \cap M)$. This motivates the definition of the distance function below. Before defining the distance function we need to prove some general properties of side conditions. For a set of ordinals X we let \overline{X} denote the closure of X in the order topology.

Fact 8.3.2. *Suppose $P, Q \in \mathcal{E}_0^2$ and $\delta_P \leq \delta_Q$.*

1. *If $\gamma \in P \cap Q \cap \omega_2$ then $P \cap \gamma \subseteq Q \cap \gamma$.*
2. *If P and Q are lim-compatible and $\gamma \in \overline{P \cap \omega_2} \cap \overline{Q \cap \omega_2}$ then $P \cap \gamma \subseteq Q \cap \gamma$.*

Proof. (1) For each $\alpha < \omega_2$ let e_α be the \leq -least injection from α to ω_1 . Then $P \cap \gamma = e_\gamma^{-1}[\delta_P]$ and $Q \cap \gamma = e_\gamma^{-1}[\delta_Q]$. Since $\delta_P \leq \delta_Q$ we have that $P \cap \gamma \subseteq Q \cap \gamma$. (2) If $\gamma \in P \cap Q$ this is (1). Suppose γ is a limit point of either P or Q then it is also the limit point of the other. Since P and Q are lim-compatible we have that $\gamma \in \lim(P \cap Q)$. Then $P \cap \gamma = \bigcup \{e_\alpha^{-1}[\delta_P] : \alpha \in P \cap Q\}$ and $Q \cap \gamma = \bigcup \{e_\alpha^{-1}[\delta_Q] : \alpha \in P \cap Q\}$. Since $\delta_P \leq \delta_Q$ we conclude that $P \cap \gamma \subseteq Q \cap \gamma$. \square

Fact 8.3.3. *Suppose $p \in \mathbb{M}$ and $P, Q \in \pi_0(\mathcal{M}_p)$. If $\delta_P < \delta_Q$ and $P \subseteq Q$ then $P \in Q$.*

Proof. If there is no uncountable model in the interval $(P, Q)_p$, then $P \in Q$ by transitivity. Otherwise, let S be the \in^* -largest uncountable model below Q and we proceed by \in^* -induction. First note that $S \in Q$ by transitivity and if we let $Q^* = Q \cap S$ then $\delta_{Q^*} = \delta_Q$. Since $P \subseteq S$, we have that $P \subseteq Q^*$ and so Q^* is \in^* -above P . By the inductive assumption we have $P \in Q^* \subseteq Q$, as desired. \square

Definition 8.3.4. Let $p = \mathcal{M}_p \in \mathbb{M}^*$, $\alpha, \beta \in \omega_2$ and let ξ be a countable ordinal. Then the binary relation $L_{p,\xi}(\alpha, \beta)$ holds if there is a $P \in \mathcal{M}_p$, with $\delta_P \leq \xi$, such that $\alpha, \beta \in \overline{P \cap \omega_2}$. In this case we will say that α and β are p, ξ -linked.

Definition 8.3.5. Let $p = \mathcal{M}_p \in \mathbb{M}^*$ and $\xi < \omega_1$. We let $C_{p,\xi}$ be the transitive closure of the relation $L_{p,\xi}$. If $C_{p,\xi}(\alpha, \beta)$ holds we say that α and β are p, ξ -connected. If $\alpha < \beta$ and α and β are p, ξ -connected we write $\alpha <_{p,\xi} \beta$.

From Fact 8.3.2(2) we now have the following.

Fact 8.3.6. Suppose $p = \mathcal{M}_p \in \mathbb{M}^*$ and $\xi < \omega_1$.

1. Suppose $\alpha < \beta < \gamma$ are ordinal in ω_2 . If $L_{p,\xi}(\alpha, \gamma)$ and $L_{p,\xi}(\beta, \gamma)$ hold, then so does $L_{p,\xi}(\alpha, \beta)$.
2. If $\alpha <_{p,\xi} \beta$ then there is a sequence $\alpha = \gamma_0 < \gamma_1 < \dots < \gamma_n = \beta$ such that $L_{p,\xi}(\gamma_i, \gamma_{i+1})$ holds, for all $i < n$.

\square

We now present some properties of the relation $<_{p,\xi}$, in order to define the distance function we will use in the definition of the main forcing.

Fact 8.3.7. Let $p = \mathcal{M}_p \in \mathbb{M}^*$ and $\xi < \omega_1$. Suppose $\alpha < \beta < \gamma < \omega_2$. Then

1. if $\alpha <_{p,\xi} \beta$ and $\beta <_{p,\xi} \gamma$, then $\alpha <_{p,\xi} \gamma$,
2. if $\alpha <_{p,\xi} \gamma$ and $\beta <_{p,\xi} \gamma$, then $\alpha <_{p,\xi} \beta$.

Proof. Part (1) follows directly from the definition of the relation $<_{p,\xi}$. To prove (2) let $\alpha = \gamma_0 < \dots < \gamma_n = \gamma$ witness the p, ξ -connection between α and γ and let $\beta = \delta_0 < \dots < \delta_l = \gamma$ witness the p, ξ -connection between β and γ . We have that $L_{p,\xi}(\gamma_i, \gamma_{i+1})$ holds, for all $i < n$, and $L_{p,\xi}(\delta_j, \delta_{j+1})$ holds, for all $j < l$. We prove that α and β are p, ξ -connected by induction on $n + l$. If $n = l = 1$ this is simply Fact 8.3.6(1). Let now $n, l > 1$. Assume for concreteness that $\delta_{l-1} \leq \gamma_{n-1}$. By Fact 8.3.6(1) $L_{p,\xi}(\delta_{l-1}, \gamma_{n-1})$ holds; so $\alpha <_{p,\xi} \gamma_{n-1}$ and $\beta <_{p,\xi} \gamma_{n-1}$. Now, by the inductive assumption we conclude that α and β are p, ξ -connected, i.e. $\alpha <_{p,\xi} \beta$. The case $\gamma_{n-1} < \delta_{l-1}$ is treated similarly. \square

The above lemma shows in (1) that the relation $<_{p,\xi}$ is transitive and in (2) that the set $(\omega_2, <_{p,\xi})$ has a tree structure. Since for every $M \in \mathcal{E}_0^2$ if $\delta_M \leq \xi$ then $\text{o.t.}(M \cap \omega_2) < g(\xi)$ and $g(\xi)$ is indecomposable we conclude that the height of $(\omega_2, <_{p,\xi})$ is at most $g(\xi)$. For every $\alpha <_{p,\xi} \beta$ we let $(\alpha, \beta)_{p,\xi} = \{\eta : \alpha <_{p,\xi} \eta <_{p,\xi} \beta\}$. We define similarly $[\alpha, \beta)_{p,\xi}$ and $(\alpha, \beta]_{p,\xi}$ and $[\alpha, \beta]_{p,\xi}$. If $0 <_{p,\xi} \beta$, i.e. β belongs to some $M \in \mathcal{M}_p$ with $\delta_M \leq \xi$ we write $(\beta)_{p,\xi}$ for the interval $[0, \beta)_{p,\xi}$. Thus, $(\beta)_{p,\xi}$ is simply the set of predecessors of β in $<_{p,\xi}$. If β does not belong to $\overline{M \cap \omega_2}$ for any $M \in \mathcal{M}_p$ with $\delta_M \leq \xi$ we leave $(\beta)_{p,\xi}$ undefined. Note that when defined $(\beta)_{p,\xi}$ is a closed subset of β in the ordinal topology.

Fact 8.3.8. *Let $p \in \mathbb{M}^*$, $M \in \mathcal{M}_p$, $\xi \in M \cap \omega_1$ and $\beta \in M \cap \omega_2$. Then $(\beta)_{p,\xi} \subseteq M$. Moreover, if we let $p^* = p \cap M$ then $(\beta)_{p,\xi} = (\beta)_{p^*,\xi}$.*

Proof. Let $\alpha <_{p,\xi} \beta$ and fix a sequence $\alpha = \gamma_0 < \gamma_1 < \dots < \gamma_n = \beta$ such that $L_{p,\xi}(\gamma_i, \gamma_{i+1})$ holds, for all $i < n$. We proceed by induction on n . Suppose first $n = 1$ and let P witness that α and β are p, ξ -linked. Since $\delta_P < \delta_M$ we have by Fact 8.3.2 that $P \cap \beta \subseteq M$ and by Fact 8.3.3 that $P \cap M \in M$. Therefore $\alpha, \beta \in \overline{P \cap M \cap \omega_2} \subseteq M$ and so $P \cap M$ witnesses that α and β are p^*, ξ -linked. Consider now the case $n > 1$. By the same argument as in the case $n = 1$ we know that γ_{n-1} and β are p^*, ξ -linked and then by the inductive hypothesis we conclude that α and β are p^*, ξ -connected. \square

Fact 8.3.9. *Let $p \in \mathbb{M}^*$, $M \in \mathcal{M}_p$, $\beta \in \omega_2 \setminus M$ and $\xi \in M \cap \omega_1$. If $(\beta)_{p,\xi} \cap M$ is non empty then it has a largest element, say η . Moreover, there is $Q \in \mathcal{M}_p \setminus M$ with $\delta_Q \leq \xi$ such that $\eta = \sup(Q \cap M \cap \omega_2)$.*

Proof. Assume $(\beta)_{p,\xi} \cap M$ is non empty and let η be its supremum. Note that η is a limit ordinal. Since $(\beta)_{p,\xi}$ is a closed subset of β in the order topology we know that either $\eta <_{p,\xi} \beta$ or $\eta = \beta$. By Fact 8.3.8 $(\beta)_{p,\xi} \cap M = (\beta)_{p,\xi} \cap \eta = (\eta)_{p,\xi}$. For every $\rho \in (\eta)_{p,\xi}$ there is some $P \in \mathcal{M}_p \cap M$ with $\delta_P \leq \xi$ such that $\rho \in \overline{P \cap \omega_2}$. Since $\mathcal{M}_p \cap M$ is finite there is such P with $\eta \in \overline{P \cap \omega_2}$. Since $P \in M$ it follows that $\overline{P} \subseteq M$, so $\eta \in M$ and therefore $\eta < \beta$. Finally, since η and β are p, ξ -connected, there is a chain $\eta = \gamma_0 < \gamma_1 < \dots < \gamma_n = \beta$ such that γ_i and γ_{i+1} are p, ξ -linked, for all i . Let Q witness that $\eta = \gamma_0$ and γ_1 are p, ξ -linked. Then $\delta_Q \leq \xi$ and $\eta = \sup(Q \cap M \cap \omega_2)$. Since $\gamma_1 \in \overline{Q \cap \omega_2} \setminus M$ it follows that $Q \notin M$. Therefore, Q is as required. \square

We are now ready to define the distance function.

Definition 8.3.10. *Let $p \in \mathcal{M}_p \in \mathbb{M}^*$, $\alpha, \beta \in \omega_2$, and $\xi \in \omega_1$. If $\alpha <_{p,\xi} \beta$ we define the p, ξ -distance of α and β as*

$$d_{p,\xi}(\alpha, \beta) = \text{o.t.}([\alpha, \beta)_{p,\xi}).$$

Otherwise we leave $d_{p,\xi}(\alpha, \beta)$ undefined.

Remark 8.3.11. Notice that for every p and ξ the function $d_{p,\xi}$ is additive, i.e. if $\alpha <_{p,\xi} \beta <_{p,\xi} \gamma$ then

$$d_{p,\xi}(\alpha, \gamma) = d_{p,\xi}(\alpha, \beta) + d_{p,\xi}(\beta, \gamma).$$

Moreover, we have that $d_{p,\xi}(\alpha, \beta) < g(\xi)$, for every $\alpha <_{p,\xi} \beta$.

We can now define the notion of forcing which adds an ω_2 chain in $(\omega_1^{\omega_1}, <_{\text{fin}})$ below the function g .

Definition 8.3.12. Let \mathbb{M}_3^* be the forcing notion whose elements are pairs $p = (f_p, \mathcal{M}_p)$, where f_p is a partial function from $\omega_2 \times \omega_1$ to ω_1 , $\text{dom}(f_p)$ is of the form $A_p \times F_p$ where $0 \in A_p \in [\omega_2]^{<\omega}$, $F_p \in [\omega_1]^{<\omega}$, $\mathcal{M}_p \in \mathbb{M}^*$, and for every $\alpha, \beta \in A_p$ with $\alpha < \beta$, every $\xi \in F_p$ and $M \in \mathcal{M}_p$:

1. $f_p(\alpha, \xi) < g(\xi)$,
2. if $\alpha <_{p,\xi} \beta$ then $f_p(\alpha, \xi) + d_{p,\xi}(\alpha, \beta) \leq f_p(\beta, \xi)$,

We let $q \leq p$ if $f_p \subseteq f_q$, $\mathcal{M}_p \subseteq \mathcal{M}_q$ and for every $\alpha, \beta \in A_p$ and $\xi \in F_q \setminus F_p$ if $\alpha < \beta$ then $f_q(\alpha, \xi) < f_q(\beta, \xi)$.

We first show that for any $\alpha < \omega_2$ and $\xi < \omega_1$ any condition $p \in \mathbb{M}_3^*$ can be extended to a condition q such that $\alpha \in A_q$ and $\xi \in F_q$.

Lemma 8.3.13. Let $p \in \mathbb{M}_3^*$ and $\delta \in \omega_2 \setminus A_p$. Then there is a condition $q \leq p$ such that $\delta \in A_q$.

Proof. We let $\mathcal{M}_q = \mathcal{M}_p$, $A_q = A_p \cup \{\delta\}$ and $F_q = F_p$. On $A_p \times F_p$ we let f_q be equal to f_p . We need to define $f_q(\delta, \xi)$, for $\xi \in F_p$. Consider one such ξ . If δ does not belong to $\overline{M \cap \omega_2}$, for any $M \in \mathcal{M}_q$ with $\delta_M \leq \xi$, we can define $f_q(\delta, \xi)$ arbitrarily. Otherwise, we need to ensure that if $\alpha \in A_p$ and $\alpha <_{p,\xi} \delta$ then

$$f_{p,\xi}(\alpha, \xi) + d_{p,\xi}(\alpha, \delta) \leq f_q(\delta, \xi).$$

Similarly, if $\beta \in A_p$ and $\delta <_{p,\xi} \beta$ we have to ensure that

$$f_q(\delta, \xi) + d_{p,\xi}(\delta, \beta) \leq f_p(\beta, \xi).$$

By the additivity of $d_{p,\xi}$ we know that if $\alpha <_{p,\xi} \delta <_{p,\xi} \beta$ then $d_{p,\xi}(\alpha, \beta) = d_{p,\xi}(\alpha, \delta) + d_{p,\xi}(\delta, \beta)$. Since p is a condition we know that if $\alpha, \beta \in A_p$ then $f_p(\beta, \xi) \geq f_p(\alpha, \xi) + d_{p,\xi}(\alpha, \beta)$. Let α^* be the largest element of $A_p \cap (\delta)_{p,\xi}$. We can then simply define $f_q(\delta, \xi)$ by

$$f_q(\delta, \xi) = f_p(\alpha^*, \xi) + d_{p,\xi}(\alpha^*, \delta).$$

It is straightforward to check that the q thus defined is a condition. \square

Lemma 8.3.14. *Let $p \in \mathbb{M}_3^*$ and $\xi \in \omega_1 \setminus F_p$. Then there is a condition $q \leq p$ such that $\xi \in F_q$.*

Proof. We let $\mathcal{M}_q = \mathcal{M}_p$, $A_q = A_p$ and $F_q = F_p \cup \{\xi\}$. Then we need to extend f_p to $A_q \times \{\xi\}$. Notice that we now have the following commitments. Suppose $\alpha, \beta \in A_p$ and $\alpha < \beta$, then we need to ensure that $f_q(\alpha, \xi) < f_q(\beta, \xi)$ in order for q to be an extension of p . If in addition $\alpha <_{p, \xi} \beta$ then we need to ensure that

$$f_q(\alpha, \xi) + d_{q, \xi}(\alpha, \beta) \leq f_q(\beta, \xi)$$

in order for q to satisfy (2) of Definition 8.3.12. We define $f_q(\beta, \xi)$ by induction on $\beta \in A_q$ as follows. We let $f_q(0, \xi) = 0$. For $\beta > 0$ we let $f_q(\beta, \xi)$ be the maximum of the following set:

$$\{f_q(\alpha, \xi) + 1 : \alpha \in (A_q \cap \beta) \setminus (\beta)_{q, \xi}\} \cup \{f_q(\alpha, \xi) + d_{q, \xi}(\alpha, \beta) : \alpha \in A_q \cap (\beta)_{q, \xi}\}.$$

It is easy to see that $f_q(\beta, \xi) < g(\xi)$, for all $\beta \in A_q$, and that q is a condition extending p . \square

In order to prove strong properness of \mathbb{M}_3^* we need to restrict to a relative club subset of \mathcal{E}^2 of elementary submodels of $H(\aleph_2)$ which are the restriction to $H(\aleph_2)$ of an elementary submodel of $H(2^{\aleph_1^+})$.

Definition 8.3.15. *Let \mathcal{D}^2 be the set of all $M \in \mathcal{E}^2$ such that $M = M^* \cap H(\aleph_2)$, for some $M^* \prec H(2^{\aleph_1^+})$. We let $\mathcal{D}_0^2 = \mathcal{D}^2 \cap \mathcal{E}_0^2$ and $\mathcal{D}_1^2 = \mathcal{D}^2 \cap \mathcal{E}_1^2$.*

We split the proof that \mathbb{M}_3^* is \mathcal{D}^2 -strongly proper in two lemmas.

Lemma 8.3.16. *Let $p \in \mathbb{M}_3^*$ and $M \in \mathcal{M}_p \cap \mathcal{D}_0^2$. Then p is an (M, \mathbb{M}_3^*) -strongly generic condition.*

Proof. Given $r \leq p$ we need to find a condition $r|M \in M$ such that every $q \leq r|M$ which is in M is compatible with r . By Lemma 8.3.14 we may assume that $\text{sup}(P) \in A_r$, for every $P \in \mathcal{M}_r$. The idea is to choose $r|M$ which has the same type as r over some suitably chosen parameters in M . Let $D = \{\delta_P : P \in \mathcal{M}_r\} \cap M$. Since $M \in \mathcal{D}_0^2$ there is $M^* \prec H(2^{\aleph_1^+})$ such that $M = M^* \cap H(\aleph_2)$. By elementary of M^* , we can find in M an \in -chain $\mathcal{M}_{r^*} \in \mathbb{M}^*$ extending $\mathcal{M}_r \cap M$, a finite set $A_{r^*} \subseteq \omega_2$ and an order preserving bijection $\pi : A_r \rightarrow A_{r^*}$ such:

1. π is the identity function on $A_r \cap M$,
2. if $\alpha, \beta \in A_r$ then, for every $\xi \in D$,

$$d_{r^*, \xi}(\pi(\alpha), \pi(\beta)) = d_{r, \xi}(\alpha, \beta).$$

By Lemma 8.3.13 we can extend $f_r \upharpoonright (A_r \cap M) \times (F_r \cap M)$ to a function $f_{r^*} : A_{r^*} \times (F_r \cap M) \rightarrow \omega_1$ such that $(f_{r^*}, \mathcal{M}_{r^*})$ is a condition in \mathbb{M}_3^* . Finally, we set $r \upharpoonright M = r^*$.

Suppose now $q \leq r \upharpoonright M$ and $q \in M$. We need to find a common extension s of q and r . We define \mathcal{M}_s to be the closure under intersection of $\mathcal{M}_r \cup \mathcal{M}_q$. Indeed Lemma 8.2.17 shows that $\mathcal{M}_s \in \mathbb{M}^*$. We first compute the distance function $d_{s,\xi}$ in terms of $d_{r,\xi}$ and $d_{q,\xi}$, for $\xi < \omega_1$. First notice that the new models which are obtained by closing $\mathcal{M}_q \cup \mathcal{M}_r$ under intersection do not create new links and therefore do not influence the computation of the distance function.

Now, consider an ordinal $\xi < \omega_1$. If $\xi \geq \delta_M$ then all ordinals in $\overline{M \cap \omega_2}$ are pairwise r, ξ -linked. The countable models of $\mathcal{M}_q \setminus \mathcal{M}_r$ are all included in M so they do not add any new s, ξ -links. It follows that in this case $d_{s,\xi} = d_{r,\xi}$. Consider now an ordinal $\xi < \delta_M$. By Fact 8.3.8 if $\beta \in M$ then $(\beta)_{s,\xi} = (\beta)_{q,\xi}$. If $\beta \notin M$ then, by Fact 8.3.9 there is a $\eta \in A_r \cap M$ such that $(\beta)_{s,\xi} \cap M = (\eta)_{q,\xi}$. Let $\xi^* = \max(D \cap (\xi + 1))$. Then, again by Fact 8.3.9, η and β are r, ξ^* -connected and

$$d_{s,\xi}(\alpha, \beta) = d_{q,\xi}(\alpha, \eta) + d_{r,\xi^*}(\eta, \beta).$$

Let $A_s = A_q \cup A_r$ and $F_s = F_q \cup F_r$. Our next goal is to define an extension, call it f_s , of $f_q \cup f_r$ on $A_s \times F_s$. It remains to define f_s on

$$((A_q \setminus A_r) \times (F_r \setminus F_q)) \cup ((A_r \setminus A_q) \times (F_q \setminus F_r)).$$

Case 1: Consider first $\xi \in F_r \setminus F_q$ and let us define f_s on $(A_q \setminus A_r) \times \{\xi\}$. We already know that $d_{s,\xi} = d_{r,\xi}$, so we need to ensure that if $\alpha, \beta \in A_s$ and $\alpha <_{s,\xi} \beta$ then

$$f_s(\alpha, \xi) + d_{r,\xi}(\alpha, \beta) \leq f_s(\beta, \xi).$$

Notice that all the ordinals of A_q are r, ξ -linked as witnessed by M so then we will also have that for every $\alpha, \beta \in A_q$, if $\alpha < \beta$ then $f_s(\alpha, \xi) < f_s(\beta, \xi)$. In order to define $f_s(\alpha, \xi)$, for $\alpha \in A_q$, let α^* be the maximal element of $(\alpha)_{r,\xi} \cap A_r$ and let $f_s(\alpha, \xi) = f_r(\alpha^*, \xi) + d_{r,\xi}(\alpha^*, \alpha)$. It is straightforward to check that (2) of Definition 8.3.12 is satisfied in this case.

Case 2: Consider now some $\xi \in F_q \setminus F_r$. What we have to arrange is that $f_s(\alpha, \xi) < f_s(\beta, \xi)$, for every $\alpha, \beta \in A_r$ with $\alpha < \beta$. Moreover, for every $\alpha, \beta \in A_s$ with $\alpha <_{s,\xi} \beta$ we have to arrange that

$$f_s(\alpha, \xi) + d_{s,\xi}(\alpha, \beta) \leq f_s(\beta, \xi).$$

We define f_s on $(A_r \setminus A_q) \times \{\xi\}$ by setting

$$f_s(\beta, \xi) = f_q(\pi(\beta), \xi).$$

First, we show that the function $\alpha \mapsto f_s(\alpha, \xi)$ is order preserving on A_r . To see this observe that, since $q \leq r^* = r|_M$ and $\xi \notin F_{r^*}$, the function $\alpha \mapsto f_q(\alpha, \xi)$ is strictly order preserving on A_{r^*} . Moreover, π is order preserving and the identity on $A_r \cap M = A_r \cap A_q$.

Assume now $\alpha, \beta \in A_s$ and $\alpha <_{s, \xi} \beta$. If $\alpha, \beta \in A_q$ then, since q is a condition, $f_s(\beta, \xi) \geq f_s(\alpha, \xi) + d_{q, \xi}(\alpha, \beta)$. On the other hand, we know that $d_{s, \xi}(\alpha, \beta) = d_{q, \xi}(\alpha, \beta)$, so we have the required inequality in this case. By Fact 8.3.8 the case $\alpha \in A_r \setminus A_q$ and $\beta \in A_q$ cannot happen. Suppose $\alpha \in A_q$ and $\beta \in A_r \setminus A_q$. Let $\xi^* = \max(D \cap (\xi + 1))$. By Fact 8.3.9 there is $\eta \in A_r \cap M$ such that

$$d_{s, \xi}(\alpha, \beta) = d_{s, \xi}(\alpha, \eta) + d_{r, \xi^*}(\eta, \beta).$$

By property (2) of π we have that $d_{r^*, \xi^*}(\eta, \pi(\beta)) = d_{r, \xi^*}(\eta, \beta)$. Since q extends r^* it follows that $d_{q, \xi^*}(\eta, \pi(\beta)) \geq d_{r^*, \xi^*}(\eta, \pi(\beta))$. Moreover, q is a condition and so:

$$f_q(\pi(\beta), \xi) \geq f_q(\alpha, \xi) + d_{q, \xi}(\alpha, \pi(\beta)) \geq f_q(\alpha, \xi) + d_{q, \xi^*}(\alpha, \pi(\beta)).$$

Therefore,

$$f_q(\pi(\beta), \xi) \geq f_q(\alpha, \xi) + d_{s, \xi}(\alpha, \beta).$$

The final case is when $\alpha, \beta \in A_r \setminus A_q$ and $\alpha <_{s, \xi} \beta$. Note that in this case, α and β are already r, ξ -connected, in fact, they are r, ξ^* -connected, where as before $\xi^* = \max(D \cap (\xi + 1))$. By property (2) of π we have that $\pi(\alpha)$ and $\pi(\beta)$ are r^*, ξ^* -connected and

$$d_{r^*, \xi^*}(\pi(\alpha), \pi(\beta)) = d_{r, \xi}(\alpha, \beta).$$

Since $\xi^* \leq \xi$ and q extends r^* we have that

$$d_{q, \xi}(\pi(\alpha), \pi(\beta)) \geq d_{r^*, \xi^*}(\pi(\alpha), \pi(\beta)).$$

Since q is a condition we have

$$f_q(\pi(\beta), \xi) \geq f_q(\pi(\alpha), \xi) + d_{q, \xi}(\pi(\alpha), \pi(\beta)).$$

Since $d_{s, \xi}(\pi(\alpha), \pi(\beta)) = d_{q, \xi}(\pi(\alpha), \pi(\beta))$ we have $f_s(\beta, \xi) \geq f_s(\alpha, \xi) + d_{s, \xi}(\alpha, \beta)$, as required.

It follows that s is a condition which extends q and r . This completes the proof of Lemma 8.3.16. \square

Lemma 8.3.17. *Let $p \in \mathbb{M}_3^*$ and $M \in \pi_1(\mathcal{M}_p)$. Then p is (M, \mathbb{M}_3^*) -strongly generic.*

Proof. Let $r \leq p$. We need to find a condition $r|M \in M$ such that any $q \leq r|M$ in M is compatible with r . We simply set

$$r|M = (f_r \upharpoonright (A_r \times F_r) \cap M, \mathcal{M}_p \cap M).$$

We need to show that if $q \leq r|M$ is in M , then there is a condition $s \leq q, r$. Thanks to Lemma 8.2.17 we just need to define f_s , since we already know that $\mathcal{M}_r \cup \mathcal{M}_q$ is an \in -chain and belongs to \mathbb{M}^* . Since $\omega_1 \subseteq M$ we have that $F_r \subseteq M$ so we only need to define an extension f_s on $A_r \setminus A_q \times F_q \setminus F_r$. We know that $M \cap \omega_2$ is an initial segment of ω_2 so all the elements of $A_r \setminus A_q = A_r \setminus M$ are above all the ordinals of A_q . Given an ordinal $\xi \in F_q \setminus F_r$ we define $f_s(\beta, \xi)$, for $\beta \in A_r \setminus A_q$ by induction. We set:

$$f_s(\beta, \xi) = \max(\{f_s(\alpha, \xi) + 1 : \alpha \in A_r \cap \beta\} \cup \{f_s(\alpha, \xi) + d_{s, \xi}(\alpha, \beta) : \alpha <_{s, \xi} \beta\}).$$

It is easy to check that (f_s, \mathcal{M}_s) is a condition which extends both q and r . \square

Corollary 8.3.18. *The forcing \mathbb{M}_3^* is \mathcal{D}^2 -strongly proper. Hence it preserves ω_1 and ω_2 . \square*

We have shown that for every $\alpha < \omega_2$ and $\xi < \omega_1$ the set

$$D_{\alpha, \xi} = \{p \in \mathbb{M}_3^* : \alpha \in A_p, \xi \in F_p\}$$

is dense in \mathbb{M}_3 . If G is a V -generic filter in \mathbb{M}_3^* we let

$$f_G = \bigcup \{f_p : p \in G\}.$$

It follows that $f_G : \omega_2 \times \omega_1 \rightarrow \omega_1$. For $\alpha < \omega_2$ we define $f_\alpha : \omega_1 \rightarrow \omega_1$ by letting $f_\alpha(\xi) = f_G(\alpha, \xi)$, for all ξ . It follows that the sequence $(f_\alpha : \alpha < \omega_2)$ is an increasing ω_2 -chain in $(\omega_1^{\omega_1}, <_{\text{fin}})$. We have thus completed the proof of the following.

Theorem 8.3.19. *There is a \mathcal{D}^2 -strongly proper forcing which adds an ω_2 chain in $(\omega_1^{\omega_1}, <_{\text{fin}})$. \square*

8.4 Thin very tall superatomic Boolean algebras

A Boolean algebra \mathcal{B} is called *superatomic* (sBa) iff every homomorphic image of \mathcal{B} is atomic. In particular, \mathcal{B} is an sBa iff its Stone space $S(\mathcal{B})$ is scattered. A very useful tool for studying scattered spaces is the Cantor-Bendixson derivative $A^{(\alpha)}$ of a set $A \subseteq S(\mathcal{B})$, defined by induction on α as follows. Let $A^{(0)} = A$, $A^{(\alpha+1)}$ is the set of limit points of $A^{(\alpha)}$, and $A^{(\lambda)} = \bigcap \{A^{(\alpha)} : \alpha < \lambda\}$, if λ is a limit ordinal. Then $S(\mathcal{B})$ is scattered iff for $S(\mathcal{B})^{(\alpha)} = \emptyset$, for some α .

When this notion is transferred to the Boolean algebra \mathcal{B} , we arrive at a sequence of ideals I_α , which we refer to as the Cantor-Bendixson ideals, defined by induction on α as follows. Let $I_0 = \{0\}$. Given I_α let $I_{\alpha+1}$ be generated by I_α together with all $b \in \mathcal{B}$ such that b/I_α is an atom in \mathcal{B}/I_α . If α is a limit ordinal, let $I_\alpha = \bigcup \{I_\xi : \xi < \alpha\}$. Then \mathcal{B} is an sBa iff some $I_\alpha = \mathcal{B}$, for some α .

The height of an sBa \mathcal{B} , $\text{ht}(\mathcal{B})$, is the least ordinal α such that $I_\alpha = \mathcal{B}$. For $\alpha < \text{ht}(\mathcal{B})$ let $\text{wd}_\alpha(\mathcal{B})$ be the cardinality of the set of atoms in \mathcal{B}/I_α . The *cardinal sequence* of \mathcal{B} is the sequence $(\text{wd}_\alpha(\mathcal{B}) : \alpha < \text{ht}(\mathcal{B}))$. We say that \mathcal{B} is κ -*thin-very tall* if $\text{ht}(\mathcal{B}) = \kappa^{++}$ and $\text{wd}_\alpha(\mathcal{B}) = \kappa$, for all $\alpha < \kappa^{++}$. If $\kappa = \omega$ we simply say that \mathcal{B} is *thin very tall*.

Baumgartner and Shelah [10] constructed a forcing notion which adds a thin very tall sBa. This is achieved in two steps. First they adjoin by a σ -closed \aleph_2 -cc forcing a function $f : [\omega_2]^2 \rightarrow [\omega_2]^{\leq \omega}$ with some special properties. Such a function is called a Δ -function. In the second step they use a Δ -function to define a ccc forcing notion which adds a thin very tall sBa. The purpose of this section is to show how this can be achieved directly by using generalizes side conditions. The following concept from [10] was made explicit by Bagaria in [7].

Definition 8.4.1. *Given a cardinal sequence $\theta = \langle \kappa_\alpha : \alpha < \lambda \rangle$, where each κ_α is an infinite cardinal, we say that a structure (T, \leq, i) is a θ -poset if $<$ is a partial ordering on T and the following hold:*

1. $T = \bigcup \{T_\alpha : \alpha < \lambda\}$, where each T_α is of the form $\{\alpha\} \times Y_\alpha$, and Y_α is a set of cardinality κ_α .
2. If $s \in T_\alpha$, $t \in T_\beta$ and $s < t$, then $\alpha < \beta$.
3. For every $\alpha < \beta < \lambda$, if $t \in T_\beta$ then the set $\{s \in T_\alpha : s < t\}$ is infinite.
4. i is a function from $[T]^2$ to $[T]^{<\omega}$ with the following properties:
 - (a) If $u \in i\{s, t\}$, then $u \leq s, t$
 - (b) If $u \leq s, t$, then there exists $v \in i\{s, t\}$ such that $u \leq v$.

We let $\Omega(\lambda)$ denote the sequence of length λ with all entries equal to ω . The following is implicitly due to Baumgartner (see [7] for a proof).

Fact 8.4.2. *Let $\theta = \langle \kappa_\alpha : \alpha < \lambda \rangle$ be a sequence of cardinals. If there exists a θ -poset, then there exists an sBa whose cardinal sequence is θ . \square*

We now define a forcing notion which adds an $\Omega(\omega_2)$ -poset. If $x \in \omega_2 \times \omega$ is of the form (α, n) then we denote α by α_x and n by n_x .

Definition 8.4.3. *Let \mathbb{M}_4 be the forcing notion whose elements are tuples $p = (x_p, \leq_p, i_p, \mathcal{M}_p)$, where x_p is a finite subset of $\omega_2 \times \omega$, \leq_p is a partial ordering on x_p , $i_p : [x_p]^2 \rightarrow [x_p]^{<\omega}$, $\mathcal{M}_p \in \mathbb{M}$ and the following hold:*

1. *if $s, t \in x_p$ and $s <_p t$ then $\alpha_s < \alpha_t$,*
2. *if $s \leq_p t$ then $i_p\{s, t\} = \{s\}$,*
3. *if $u \in i_p\{s, t\}$ then $u \leq_p s, t$,*
4. *for every $u \leq_p s, t$ there is $v \in i_p\{s, t\}$ such that $u \leq_p v$,*
5. *for every $s, t \in x_p$ and $M \in \mathcal{M}_p$ if $s, t \in M$ then $i_p\{s, t\} \in M$.*

We let $q \leq p$ if and only if $x_q \supseteq x_p$, $\leq_q \upharpoonright x_p = \leq_p$, $i_q \upharpoonright [x_p]^2 = i_p$ and $\mathcal{M}_p \subseteq \mathcal{M}_q$.

We first observe that a version of Lemma 8.1.8 holds for \mathbb{M}_4 .

Lemma 8.4.4. *Let $M \in \mathcal{E}^2$ and let $p \in \mathbb{M}_4 \cap M$. Then there is a new condition, which we will call p^M , that is the smallest element of \mathbb{M}_4 extending p such that $M \in \mathcal{M}_{p^M}$.*

Proof. If $M \in \mathcal{E}_1^2$ then simply let $p^M = (x_p, \leq_p, i_p, \mathcal{M}_p \cup \{M\})$. If $M \in \mathcal{E}_0^2$, then, as in Lemma 8.1.8, we let \mathcal{M}_{p^M} be the closure of $\mathcal{M}_p \cup \{M\}$ under intersection and let $p^M = (x_p, \leq_p, i_p, \mathcal{M}_{p^M})$. We need to check that condition (5) of Definition 8.4.3 is satisfied. Since $p \in M$ we have that $x_p \subseteq M$. In the case $M \in \mathcal{E}_1^2$ the only new model in \mathcal{M}_{p^M} is M so condition (5) holds for p^M since it holds for p . In the case $M \in \mathcal{E}_0^2$ there are also models of the form $N \cap M$, where $N \in \pi_1(\mathcal{M}_p)$. However, condition (5) holds for both N and M and so it holds for their intersection. \square

Next, we show that \mathbb{M}_4 is \mathcal{E}^2 -proper. We split this in two parts.

Lemma 8.4.5. *\mathbb{M}_4 is \mathcal{E}_0^2 -proper.*

Proof. Let θ be a sufficiently large regular cardinal and let M^* be a countable elementary submodel of $H(\theta)$ containing all the relevant objects. Then $M = M^* \cap H(\omega_2)$ belongs to \mathcal{E}_0^2 . Suppose $p \in \mathbb{M}_4 \cap M$. Let p^M be the condition defined in Lemma 8.4.4, i.e. $p^M = (x_p, \leq_p, i_p, \mathcal{M}_{p^M})$, where \mathcal{M}_{p^M} is the closure of $\mathcal{M}_p \cup \{M\}$ under intersection. We show that p^M is (M^*, \mathbb{M}_4) -generic. Let $D \in M^*$ be a dense subset of \mathbb{M}_4 and $r \leq p^M$. We need to find a condition $q \in D \cap M^*$ which is compatible with r . Note that we may assume that $r \in D$. We define a condition $r|M$ as follows. First let $x_{r|M} = x_r \cap M$ and then let $\leq_{r|M} = \leq_r \upharpoonright x_{r|M}$ and $i_{r|M} = i_r \upharpoonright [x_{r|M}]^2$. Condition (5) of Definition 8.4.3 guarantees that if $s, t \in x_{r|M}$ then $i_r\{s, t\} \subseteq M$. Finally, let $\mathcal{M}_{r|M} = \mathcal{M}_r \cap M$. It follows that $r|M = (x_{r|M}, i_{r|M}, i_{r|M}, \mathcal{M}_{r|M})$ belongs to $\mathbb{M}_4 \cap M$. By elementarity of M^* in $H(\theta)$ there is a condition $q \in D \cap M$ extending $r|M$ such that $(x_q \setminus x_{r|M}) \cap N = \emptyset$, for all $N \in \pi_0(\mathcal{M}_{r|M})$. We claim that q and r are compatible. To this end, we define a condition s as follows. We set $x_s = x_q \cup x_r$ and we let \leq_s be the transitive closure of $\leq_q \cup \leq_r$, i.e. if $u \in x_q \setminus x_r$, $v \in x_r \setminus x_q$ and $t \in x_{r|M}$ are such that $u \leq_q t$ and $t \leq_r v$, then we let $u \leq_s v$. Similarly, if $v \leq_r t$ and $r \leq_q u$ we let $v \leq_s u$. We let \mathcal{M}_s be the closure under intersection of $\mathcal{M}_q \cup \mathcal{M}_r$. It remains to define i_s . For $z \in x_s$ let $A_z = \{t \in x_{r|M} : t \leq_r z\}$ and $B_z = \{t \in x_{r|M} : t \leq_q z\}$. We let

$$i_s\{u, v\} = \begin{cases} i_q\{u, v\} & \text{if } u, v \in x_q \\ i_r\{u, v\} & \text{if } u, v \in x_r \\ \bigcup_{t \in A_v} i_q\{u, t\} \cup \bigcup_{t \in B_u} i_r\{t, v\} & \text{if } u \in x_q \setminus x_r \text{ and } v \in x_r \setminus x_q \end{cases}$$

We now need to check that, for every $P \in \mathcal{M}_s$, if $u, v \in P$, then $i_s\{u, v\} \in P$, for $u \in x_q \setminus x_r$ and $v \in x_r \setminus x_q$. First of all notice that we only need to show the above property for P in $\mathcal{M}_q \cup \mathcal{M}_r$, because the other models in \mathcal{M}_s are obtained by intersection. If $P \in \mathcal{E}_1^2$ this is straightforward, so suppose $P \in \mathcal{E}_0^2$. If $P \in \pi_0(\mathcal{M}_q)$ then $P \subseteq M$ so P does not contain any elements of $x_r \setminus x_q$. Now assume $P \in \pi_0(\mathcal{M}_r)$. Notice that, by choice of q , we have $P \cap \omega_1 = \delta_P \geq \delta_M = M \cap \omega_1$, because $u \in P \cap M$ and if $\delta_P < \delta_M$, then $P \cap M \in \pi_0(\mathcal{M}_{r|M})$, contradicting the fact that $(x_q \setminus x_{r|M}) \cap N = \emptyset$, for all $N \in \pi_0(\mathcal{M}_{r|M})$.

We now have two cases to check.

Case 1: there is a $t \in A_v$. Since $u \in P \cap M$, by Lemma 8.3.2 we have that $P \cap \alpha_u \subseteq M \cap \alpha_u$ or $M \cap \alpha_u \subseteq P \cap \alpha_u$. But $\delta_P \geq \delta_M$ and so $M \cap \alpha_u \subseteq P \cap \alpha_u$ holds. Now notice that $i_q\{u, t\} \subseteq M$, for all $t \in A_v$, because $u, t, q \in M$. Moreover if $w \in i_q\{u, t\}$, for some such t , then $\alpha_w < \alpha_u$. Hence $\alpha_w \in P$.

Case 2: there is a $t \in B_u$. As argued in Case 1 we have $M \cap \alpha_u \subseteq P \cap \alpha_u$. Now, since $t \in x_{r|M}$, it follows that $t \in P \cap \alpha_u$. Notice that $t, v \in x_r$ and they are both in $P \in \mathcal{M}_r$. So $i_r\{t, v\} \subseteq P$, because r is a condition. \square

Lemma 8.4.6. \mathbb{M}_4 is \mathcal{E}_1^2 -proper.

Proof. Let θ be a sufficiently large regular cardinal and M^* an elementary submodel of $H(\theta)$ containing all the relevant objects such that $M = M^* \cap H(\omega_2)$ belongs to \mathcal{E}_1^2 . Fix $p \in M \cap \mathbb{M}_4$. Let p^M be as in Lemma 8.4.4. We claim that p^M is (M^*, \mathbb{M}_4) -generic. In order to verify this consider a dense subset D of \mathbb{M}_4 which belongs to M^* and a condition $r \leq p^M$. We need to find a condition $q \in D \cap M^*$ which is compatible with r . By extending r if necessary we may assume it belongs to D . Let $r|M = (x_{r|M}, i_{r|M}, i_{r|M}, \mathcal{M}_{r|M})$ be as in Lemma 8.4.5. By elementarity of M^* in $H(\theta)$, we can find $q \leq r|M$, in $D \cap M$, such that $(x_q \setminus x_{r|M}) \cap N = \emptyset$, for all $N \in \mathcal{M}_{r|M}$. We claim that q and r are compatible. To see this, we define a condition s as in Lemma 8.4.5 and show that $s \leq q, r$. The only nontrivial thing to check is (5) from Definition 8.4.3. So, suppose $u \in x_q \setminus x_r$, $v \in x_r \setminus x_q$ and $P \in \mathcal{M}_s$. If $P \in \mathcal{M}_q$ then $P \subseteq M$ and so $v \notin P$. If $P \in \mathcal{M}_r$ and $P \cap M$ is \in^* below M then, since M is transitive, we have $P \cap M \in \mathcal{M}_{r|M}$. By the choice of q we have that $x_q \setminus x_r \cap (P \cap M) = \emptyset$, so $u \notin P$. The only remaining case is when P is \in^* above M and is uncountable. Then P is transitive and $M \subseteq P$. Since for every $t \in i_s\{u, v\}$ we have that $\alpha_t \leq \min(\alpha_u, \alpha_v)$, it follows that $i_s\{u, v\} \in M \subseteq P$, as desired. \square

Corollary 8.4.7. The forcing \mathbb{M}_4 is \mathcal{E}^2 -proper. Hence it preserves ω_1 and ω_2 . \square

It is easy to see that the set

$$D_{\alpha, n} = \{p \in \mathbb{M}_4 : (\alpha, n) \in x_p\}$$

is dense in \mathbb{M}_4 , for every $\alpha \in \omega_2$ and $n \in \omega$. Moreover, given $t \in \omega_2 \times \omega$, $\eta < \alpha_t$ and $n < \omega$, one verifies easily that the set

$$E_{t, \eta, n} = \{p : t \in x_p \text{ and } |\{i : (\eta, i) \in x_p \text{ and } (\eta, i) \leq_p t\}| \geq n\}$$

is dense. Then if G is V -generic filter on \mathbb{M}_4 let

$$\leq_G = \bigcup \{\leq_p : p \in G\} \quad \text{and} \quad i_G = \bigcup \{i_p : p \in G\}.$$

It follows that $(\omega_2 \times \omega, \leq_G, i_G)$ is an $\Omega(\omega_2)$ -poset in $V[G]$. We have therefore proved the following.

Theorem 8.4.8. There is an \mathcal{E}^2 -proper forcing notion which adds an $\Omega(\omega_2)$ -poset. \square

8.5 Souslin trees

The next application of the forcing \mathbb{M} will be to force an ω_2 Souslin tree. We recall that such a tree is a tree of height ω_2 whose chains and antichains have size at most \aleph_1 .

Definition 8.5.1. *Let \mathbb{M}_5 be the forcing notion whose elements are tuples $p = (X_p, \leq_p, \mathcal{M}_p)$, where $X_p = L_p \times B_p$ is a finite set of couples in $\omega_2 \times \omega_1$, the relation \leq_p is a tree order on X_p and $\mathcal{M}_p \in \mathbb{M}$, such that:*

1. *letting $\alpha_z \in L_p$ be the first component of the couple $z \in X_p$, if $x, y \in X_p$ and $x \leq_p y$, then $\alpha_x \leq \alpha_y$,*
2. *if $x, y \in X_p$ are \leq_p -incomparable, then, assuming $\alpha_x \leq \alpha_y$, there is $z \in X_p$ such that $z \leq_p y$ and $\alpha_z = \alpha_x$,*
3. *if $M \in \mathcal{M}_p$ and $x, y \in M \cap X_p$, then $x \wedge_p y \in M$, where $x \wedge_p y$ is the \leq_p -greatest common lower bound of both x and y , that is called the meet of x and y ,*
4. *letting $\xi_z \in B_p$ be the second component of the couple $z \in X_p$ and $\delta_M = M \cap \omega_1$, for $M \in \mathcal{M}_p$, if $x \in X_p$ is such that $\delta_M \leq \xi_x$ then $x \leq_p y$, implies $y \notin M$.*

The order on \mathbb{M}_5 is reverse inclusion, i.e. $q \leq p$ if $X_p \subseteq X_q$, $\leq_p \subseteq \leq_q$, $\mathcal{M}_q \leq \mathbb{M} \mathcal{M}_p$ and if $x, y \in X_p$ then $x \wedge_q y = x \wedge_p y$.

Condition (2) can be reformulated in the following way: if $x \in X_p$ then $x \upharpoonright \beta \in X_p$, for every $\beta \in L_p \cap \alpha_x$, where $x \upharpoonright \beta$ is the functional restriction, when thinking of x as the approximation of a function from ω_2 to ω_1 . In order to simplify the notation we introduce the following operation for a set $X \in [\omega_2 \times \omega_1]^{<\omega}$, equipped with a tree order \leq_X : we let $Lc(X) \in [\omega_2 \times \omega_1]^{<\omega}$ be a set including X , equipped with a tree order $\leq_{Lc(X)}$ extending \leq_X and preserving its meets, for which condition (2) holds.

Lemma 8.5.2. *Let $M \in \mathcal{E}^2$ and let $p \in \mathbb{M}_5 \cap M$. Then there is a new condition, which we will call p^M , that is the smallest element of \mathbb{M}_4 extending p such that $M \in \mathcal{M}_{p^M}$.*

Proof. If $M \in \mathcal{E}_1^2$ then simply let $p^M = (X_p, \leq_p, \mathcal{M}_p \cup \{M\})$. If $M \in \mathcal{E}_0^2$, then, as in Lemma 8.1.8, we let \mathcal{M}_{p^M} be the closure of $\mathcal{M}_p \cup \{M\}$ under intersection and let $p^M = (x_p, \leq_p, i_p, \mathcal{M}_{p^M})$. We need to check that conditions (3) and (4) of Definition 8.5.1 are satisfied. Since $p \in M$ we have that $X_p \subseteq M$ and so

condition (4) holds trivially, for M . In the case $M \in \mathcal{E}_1^2$ the only new model in \mathcal{M}_{p^M} is M so condition (3) holds for p^M since it holds for p . In the case $M \in \mathcal{E}_0^2$ there are also models of the form $N \cap M$, where $N \in \pi_1(\mathcal{M}_p)$. However, condition (3) and (4) hold for both N and M and so they hold for their intersection. \square

We now show that \mathbb{M}_5 is proper. We split the proof in two lemmas.

Lemma 8.5.3. *The forcing \mathbb{M}_5 is \mathcal{E}_0^2 -proper.*

Proof. Let θ be a sufficiently large regular cardinal and let M^* be a countable elementary submodel of $H(\theta)$ containing all the relevant objects. Then $M = M^* \cap H(\omega_2)$ belongs to \mathcal{E}_0^2 . Suppose $p \in \mathbb{M}_5 \cap M$. Let p^M be the condition defined in Lemma 8.5.2, i.e. $p^M = (X_p, \leq_p, \mathcal{M}_{p^M})$, where \mathcal{M}_{p^M} is the closure of $\mathcal{M}_p \cup \{M\}$ under intersection. We show that p^M is (M^*, \mathbb{M}_5) -generic. Let $D \in M^*$ be a dense subset of \mathbb{M}_5 and $r \leq p^M$. We need to find a condition $q \in D \cap M^*$ which is compatible with r . By density, we may assume that $r \in D$. We define a condition $r|M$ as follows. First let $X_{r|M} = X_r \cap M$ and then let $\leq_{r|M} = \leq_r \upharpoonright X_{r|M}$. Condition (3) of Definition 8.5.1 guarantees that if $x, y \in X_{r|M}$ then $x \wedge_{r|M} y \in M$. Finally, let $\mathcal{M}_{r|M} = \mathcal{M}_r \cap M$. It follows that $r|M = (X_{r|M}, \leq_{r|M}, \mathcal{M}_{r|M})$ belongs to $\mathbb{M}_5 \cap M$. By elementarity of M^* in $H(\theta)$ there is a condition $q \in D \cap M$ extending $r|M$ such that $(X_q \setminus X_{r|M}) \cap N = \emptyset$, for all $N \in \pi_0(\mathcal{M}_{r|M})$. We claim that q and r are compatible. We now define a condition s and we show that it extends both q and r .

First, let \mathcal{M}_s be the closure under intersection of $\mathcal{M}_q \cup \mathcal{M}_r$, that, by Lemma 8.2.17, is a condition in \mathbb{M} . Moreover let

$$(X_s, \leq_s) = (Lc(X_q \cup X_r), \leq_{Lc(X_q \cup X_r)}),$$

such that the following conditions hold:

- (A) if $z \in X_s \setminus (X_q \cup X_r)$, and $z \leq_s y$, with $y \in X_s \cap P$, for some $P \in \mathcal{M}_q \cup \mathcal{M}_r$, then $\xi_z < \delta_P$,
- (B) the order \leq_s extends the transitive closure of $\leq_q \cup \leq_r$, i.e. if $x \in X_q \setminus X_r$, $y \in X_r \setminus X_q$ and $z \in X_{r|M}$ are such that $x \leq_q z$ and $z \leq_r y$, then we let $x \leq_s y$. Similarly, if $y \leq_r z$ and $z \leq_q x$ we let $y \leq_s x$.

Condition (A) is imposed to respect condition (4) of Definition 8.5.1 and this is possible since both $X_s \setminus (X_q \cup X_r)$ and $\mathcal{M}_q \cup \mathcal{M}_r$ are finite.

The only non trivial part in verifying that s is a condition that extends q and r , is to check that condition (3) of Definition 8.5.1 holds for s , i.e. given $x, y \in X_s \cap P$, for some $P \in \mathcal{M}_s$, we need to check that $x \wedge_s y \in P$. Notice

that if $P \in \pi_1(\mathcal{M}_s)$, this is always the case, since $P \cap \omega_2$ is an initial segment of ω_2 and $\alpha_{x \wedge_s y} < \min\{\alpha_x, \alpha_y\}$. Moreover we can restrict ourself to models in $\mathcal{M}_q \cup \mathcal{M}_r$, since all the other models in \mathcal{M}_s are obtained by intersection. Finally notice that

$$\bigwedge(X_q \cup X_r) = \{x \wedge y : x, y \in X_q \cup X_r\} = \bigwedge(Lc(X_q \cup X_r))$$

and so we can assume x and y to be in $X_q \cup X_r$. We then have three cases to check.

Case 1: $x, y \in X_q$ and $P \in \pi_0(\mathcal{M}_r)$. Notice that $x \wedge_q y \in X_q \subseteq M$ and so, if $\delta_P < \delta_M$, then, by Fact 8.3.2, $P \cap M \in \mathcal{M}_{r|M}$, and so condition (3) holds, since q is a condition. Otherwise, if $\delta_M \leq \delta_P$, by Fact 8.3.3, we have that $P \cap M \cap \omega_2$ is an initial segment of $M \cap \omega_2$. Then $\alpha_{x \wedge_q y} < \min\{\alpha_x, \alpha_y\}$ implies that $x \wedge_q y \in P \cap M$.

Case 2: $x, y \in X_r$ and $P \in \pi_0(\mathcal{M}_q)$. Without loss of generality we can assume $x \notin M$, otherwise condition (3) holds because q is a condition. Notice that $\mathcal{M}_q \subseteq M$ and so $P \subseteq M$. But this contradicts the choice of x , hence this case cannot happen.

Case 3: $x \in X_q \setminus X_r$ and $y \in X_r \setminus X_q$. As *Case 2* shows, we can assume $P \in \pi_0(\mathcal{M}_r)$. Moreover we can also assume $\delta_M \leq \delta_P$, because $x \in P \cap M$ and, by choice of q we have $X_q \cap N = \emptyset$, for every $N \in \mathcal{M}_{r|M}$. But if $\delta_P < \delta_M$, then $P \cap M \in \mathcal{M}_{r|M}$, by Fact 8.3.2, which contradicts the choice of x . So Fact 8.3.3 implies that $P \cap M \cap \omega_2$ is an initial segment of $M \cap \omega_2$. The key observation here is that $x \wedge_s y \in X_{r|M}$ and so $x \wedge_s y \in M$. Then the fact that $\alpha_{x \wedge_s y} < \min\{\alpha_x, \alpha_y\}$, together with the fact that $P \cap M$ is an initial segment of M of height at least α_x , implies that $x \wedge_s y \in P \cap M$. □

Lemma 8.5.4. *The forcing \mathbb{M}_5 is \mathcal{E}_1^2 -proper.*

Proof. Let θ be a sufficiently large regular cardinal and M^* an elementary submodel of $H(\theta)$ containing all the relevant objects such that $M = M^* \cap H(\omega_2)$ belongs to \mathcal{E}_1^2 . Fix $p \in M \cap \mathbb{M}_5$. Let p^M be as in Lemma 8.5.2. We claim that p^M is (M^*, \mathbb{M}_5) -generic. In order to verify this consider a dense subset D of \mathbb{M}_5 which belongs to M^* and a condition $r \leq p^M$. We need to find a condition $q \in D \cap M^*$ which is compatible with r . By extending r if necessary we may assume it belongs to D . Let $r|M = (X_{r|M}, \leq_{r|M}, \mathcal{M}_{r|M})$ be as in Lemma 8.5.3. By elementarity of M^* in $H(\theta)$, we can find $q \leq r|M$, in $D \cap M$, such that $(X_q \setminus X_{r|M}) \cap N = \emptyset$, for all $N \in \mathcal{M}_{r|M}$. We claim that q and r are compatible. To see this, we define a condition s as in Lemma 8.5.3 and show that $s \leq q, r$.

First, let $\mathcal{M}_s = \mathcal{M}_q \cup \mathcal{M}_r$, that by Lemma 8.1.9, is a condition in \mathbb{M} . Now let $\bar{\alpha} = \max\{L_{r|M}\}$ and define

$$(X_s \cap \bar{\alpha}, \leq_s \upharpoonright X_s \cap \bar{\alpha}) = (Lc(X_q \cup X_r \cap \bar{\alpha}), \leq_{Lc(X_q \cup X_r \cap \bar{\alpha})}),$$

such that conditions (A) and (B) of Lemma 8.5.3 hold. Notice that the interval $(\max\{L_{r|M}\}, \max\{L_q\}]$ is disjoint from $L_r \setminus L_{r|M}$, since $M \cap \omega_2 \in \omega_2$. Then define

$$(X_s \setminus \bar{\alpha}, \leq_s \upharpoonright X_s \setminus \bar{\alpha}) = (Lc(X_q \cup X_r \setminus \bar{\alpha}), \leq_{Lc(X_q \cup X_r \setminus \bar{\alpha})}),$$

such that condition (A) of Lemma 8.5.3 holds. Notice that it is sufficient to have $\leq_{Lc(X_q \cup X_r \setminus \bar{\alpha})}$ extending $\leq_q \cup \leq_r$, without imposing (B), because if $x \in X_q \setminus X_r$ and $y \in X_r \setminus X_q$, then x and y are \leq_s -comparable if and only if there is a $z \in X_{r|M}$ in between.

It is easy to check that s is a condition extending both q and r . We still have to check that s is a condition. The only non trivial part is to verify that condition (3) holds. Again we can assume $x, y \in X_q \cup X_r$ and in $P \in \pi_0(\mathcal{M}_q \cup \mathcal{M}_r)$.

Case 1: $x, y \in X_q$ and $P \in \pi_0(\mathcal{M}_r)$. Since $M \in \pi_0(\mathcal{M}_r)$, we have $P \cap M \in \mathcal{M}_{r|M} \subseteq \mathcal{M}_q$. Then $x \wedge_q y \in P \cap M$, because q is a condition.

Case 2: $x, y \in X_r$ and $P \in \pi_0(\mathcal{M}_q)$. Since $P \subseteq M$ we have both x and y in $X_{r|M}$, but then the fact that q is a condition implies $x \wedge_q y \in P$.

Case 3: $x \in X_q \setminus X_r$ and $y \in X_r \setminus X_q$, cannot be the case. Indeed, either $P \in \pi_0(\mathcal{M}_q)$, but then it cannot contain y , or $P \in \pi_0(\mathcal{M}_r) \setminus M$. The latter is not possible since $x \in P \cap M$, but $P \cap M \in^* M$ implies $P \cap M \in M$, because $M \in \pi_1(r)$, but this contradicts the choice of q .

□

Corollary 8.5.5. *The forcing \mathbb{M}_5 is \mathcal{E}^2 -proper. Hence it preserves ω_1 and ω_2*

Using the operation Lc it is easy to see that the sets

$$D_{\alpha, \xi} = \{p \in \mathbb{M}_5 : (\alpha, \xi) \in X_p\}$$

are dense in \mathbb{M}_5 . Then if we let G be a V -generic filter for \mathbb{M}_5 , it is easy to see that

$$(T, \leq_T) = \left(\bigcup_{p \in G} X_p, \bigcup_{p \in G} \leq_p \right)$$

is an ω_2 -tree.

Lemma 8.5.6. *Let $r, q \in \mathbb{M}_5$ be as in Lemma 8.5.4, for some $M \in \pi_1(r)$. If $x \in X_q \setminus X_r$, $y \in X_r \setminus X_q$ and $z \in X_{r|M}$ are such that z is the $\leq_{r|M}$ -largest element below x and y , and $\alpha_z = \max\{L_{r|M}\}$, then there is a condition $s \leq r, q$ such that $z \leq_s x \leq_s y$.*

Proof. First let $\mathcal{M}_s = \mathcal{M}_q \cup \mathcal{M}_r$. Then define $(X_s \cap \alpha_z, \leq_s \upharpoonright X_s \cap \alpha_z)$ as in Lemma 8.5.4.

Now let

$$\leq_{s^*} = ((\leq_q \cup \leq_r) \upharpoonright (X_q \cup X_r) \setminus \{w : y \leq_r w\}) \cup \{(x, w) : y \leq_r w\}.$$

and define

$$(X_s \setminus \alpha_z, \leq_s \upharpoonright X_s \setminus \alpha_z) = (Lc(X_q \cup X_r \setminus \alpha_z), \leq_{Lc(X_q \cup X_r \setminus \alpha_z)}),$$

such that condition (A) of Lemma 8.5.3 holds and $\leq_{Lc(X_q \cup X_r \setminus \alpha_z)}$ extends \leq_{s^*} . It is easy to check that $z \leq_s x \leq_s y$ and s is a condition extending both q and r . The proof that $s \in \mathbb{M}_5$ is like that of Lemma 8.5.4, except the fact that \leq_{s^*} has different meets than $\leq_q \cup \leq_r$. \square

Proposition 8.5.7. *Every antichain in T has size at most \aleph_1 .*

Proof. Let \dot{A} be a \mathbb{M}_5 -name for an antichain in T and let $p \in \mathbb{M}_5$ force \dot{A} to be maximal. Given $M \prec H(\theta)$ an approachable model of size \aleph_1 , containing p and all the relevant parameters, let p^M be the least condition extending p and containing M , as in Lemma 8.5.2. We claim that $\dot{A} \subseteq M$. We prove it by showing that any couple (α, ξ) , forced by some condition to be in T , is \leq_T -compatible with an element of $\dot{A} \cap M$. Pick then $r \leq p^M$ such that $y = (\alpha, \xi) \in X_r$. Since r is a condition extending p , there is $z \leq_r y$ with $\alpha_z = \max\{L_p\}$. Then, either p^M forces that $\dot{A} \subseteq M$ or there is a condition $q \leq p$ in M forcing $x = (\alpha', \xi')$ to be in $\dot{A} \cap M$, with $\alpha' > \max\{L_p\}$. Since, as we notice in Lemma 8.5.4, we have

$$(\max\{L_p\}, \max\{L_q\}] \cap (L_r \setminus L_p) = \emptyset,$$

we can choose x to be such that $z \leq_q x$. Thanks to Lemma 8.5.6 there is $s \in \mathbb{M}_5$ extending q and r such that

$$z \leq_s x \leq_s y,$$

thus contradicting the fact that \dot{A} is an antichain. \square

Chapter 9

Souslin trees and side conditions

In this chapter, using the techniques introduced by Neeman in [131], we give a consistency proof of the Forcing Axiom for the class of proper forcings that preserve a Souslin tree T ($\text{PFA}(T)$). The novelty of this proof is that $\text{PFA}(T)$ is forced with finite conditions. Indeed, the known proof of this result uses a countable support iteration and a result by Miyamoto, who showed in [125] that the property “is proper and preserves every ω_1 -Souslin tree” is preserved by this kind of iteration.

The main preservation theorem presented here, Theorem 9.3.13, can be seen as a general preservation schema for properties, like being a Souslin tree, that have formulations similar to Lemma 9.1.2; i.e. in terms of the possibility to construct a generic condition for a product forcing, by means of conditions that, singularly, are generic for their respective forcings. As a matter of fact, in the proof of Theorem 9.3.13, no use is made of the fact that T is a tree.

In Section 9.1 we review some basic results connecting the property of being Souslin and properness. In Section 9.2 we show, as a warm up, that given a poset \mathbb{P} we can define an operation $\mathbb{M}(\mathbb{P})$, that consists in a scaffolding of countable models around \mathbb{P} , such that if \mathbb{P} preserves a Souslin tree T , then $\mathbb{M}(\mathbb{P})$ preserves T . Then in Section 9.3 we use the method of generalized side conditions, with models of two types, to construct a model where $\text{PFA}(T)$ holds and T remains Souslin.

9.1 Souslin trees and properness

We will use the following reformulation of the definition of Souslin tree.

Lemma 9.1.1. *A tree T is Souslin iff for every countable $M \prec H(\theta)$, with θ sufficiently large such that $T \in M$, and for every $t \in T_{\delta_M}$, where $\delta_M = M \cap \omega_1$,*

t is an (M, T) -generic condition,

i.e. for every maximal antichain $A \subseteq T$ in M , there is a $\xi < M \cap \omega_1$ such that $t \restriction \xi \in A$.

Proof. On the one hand, let T be a Souslin tree, $M \prec H(\theta)$ as above, $t \in T_{\delta_M}$ and $A \in M$ a maximal antichain of T . Since T is Souslin, A is countable. Then there is a $\alpha < \delta_M$ such that for all $\beta \geq \alpha$, the set $A \cap T_\beta$ is empty. Hence there is an element $h \in A$ compatible with $t \restriction \alpha$. Then $t \restriction ht(h) = s \in A$.

On the other hand if $A \in M$ is an uncountable maximal antichain of T , then $A \setminus M$ is not empty. For $x \in A \setminus M$, let $t = x \restriction \delta$. If there is a $\xi < \delta$ such that $t \restriction \xi \in A$, then x and $t \restriction \xi$ would be compatible and both in A : a contradiction. \square

The following lemma connects preservation of Souslin trees and properness.

Lemma 9.1.2. *(Miyamoto, Proposition 1.1 in [125]) Fix a Souslin tree T , a proper poset \mathbb{P} and some regular cardinal θ , large enough. Then the following are equivalent:*

1. $\Vdash_{\mathbb{P}}$ “ T is Souslin ”,
2. *given $M \prec H(\theta)$ countable, containing \mathbb{P} and T , if $p \in \mathbb{P}$ is a (M, \mathbb{P}) -generic condition and $t \in T_{\delta_M}$, with $\delta_M = M \cap \omega_1$, then (p, t) is an $(M, \mathbb{P} \times T)$ -generic condition,*
3. *given $M \prec H(\theta)$ countable, containing \mathbb{P} and T and given $q \in \mathbb{P} \cap M$, there is a condition $p \leq q$ such that for every condition $t \in T_{\delta_M}$, with $\delta_M = M \cap \omega_1$, we have that (p, t) is an $(M, \mathbb{P} \times T)$ -generic condition.*

\square

9.2 Preservation of T and countable models

We define the scaffolding operator from an idea of Veličković.

Definition 9.2.1. *Given a proper poset \mathbb{P} and a sufficiently large cardinal θ such that $\mathbb{P} \in H(\theta)$, let $\mathbb{M}(\mathbb{P})$ be the poset consisting of conditions $p = (\mathcal{M}_p, w_p)$ such that*

1. \mathcal{M}_p is a finite \in -chain of countable elementary substructures of $H(\theta)$,
2. $w_p \in \mathbb{P}$,
3. w_p is an (M, \mathbb{P}) -generic condition for every M in \mathcal{M}_p .

Moreover, we let $q \leq p$ iff $\mathcal{M}_p \subseteq \mathcal{M}_q$ and $w_q \leq_{\mathbb{P}} w_p$.

Remark 9.2.2. Notice that $\mathbb{M}(\mathbb{P})$ does not make reference to the cardinal θ . However this notation causes no confusion as long as θ depends on \mathbb{P} and its choice is a standard negligible part of all arguments involving properness. Then, without any specification, θ will always denote a cardinal that makes possible the definition of $\mathbb{M}(\mathbb{P})$.

Remark 9.2.3. By abuse of notation we will identify an \in -chain \mathcal{M}_p and the set of models that compose it.

Our aim now is to show that properness is preserved by the scaffolding operator.

Lemma 9.2.4. *Let \mathbb{P} be a proper poset, $M \prec H(\theta)$ and $p \in \mathbb{M}(\mathbb{P}) \cap M$. Then there is a condition $p^M = (\mathcal{M}_{p^M}, w_{p^M}) \in \mathbb{M}(\mathbb{P})$ that is the largest condition extending p and such that $M \in \mathcal{M}_{p^M}$.*

Proof. First of all notice that since $p \in M$, we have $\mathcal{M}_p \subseteq M$. In particular the largest model in \mathcal{M}_p belongs to M . So $\mathcal{M}_p \cup \{M\}$ is a finite \in -chain of elementary substructures of $H(\theta)$. Moreover $w_p \in M \cap \mathbb{P}$ and, by properness, there is a $w_q \leq w_p$ that is (M, \mathbb{P}) -generic. Now, since $w_q \leq w_p$ and w_p is (N, \mathbb{P}) -generic, for every $N \in \mathcal{M}_p$, so is w_q . Then we have that w_q is a generic condition for every model in $\mathcal{M}_p \cup \{M\}$. Finally set $\mathcal{M}_{p^M} = \mathcal{M}_p \cup \{M\}$ and $w_{p^M} = w_q$ to see that the conclusion of the lemma holds. \square

Theorem 9.2.5. *Let \mathbb{P} be a proper poset. Then $\mathbb{M}(\mathbb{P})$ is proper.*

Proof. Let M^* be a countable elementary submodel of $H(\theta^*)$, for some $\theta^* > \theta$, where θ is the corresponding cardinal in the definition of $\mathbb{M}(\mathbb{P})$. If p is a condition in $\mathbb{M}(\mathbb{P}) \cap M^*$ we need to find a condition $q \leq p$ that is $(M^*, \mathbb{M}(\mathbb{P}))$ -generic. Fix then a dense $D \subseteq \mathbb{M}(\mathbb{P})$ in M^* and let $M = M^* \cap H(\theta)$. We claim that $p^M = (\mathcal{M}_p \cup \{M\}, w_p^M)$ is an $(M, \mathbb{M}(\mathbb{P}))$ -generic condition.

Thanks to Lemma 9.2.4 we have that p^M is a condition. We now prove its genericity. Let $r \leq p^M$ and without loss of generality assume it to be in D . Define

$$E = \{w_s \in \mathbb{P} : \exists \mathcal{M}_s \text{ such that } (\mathcal{M}_s, w_s) \in D \wedge \mathcal{M}_r \cap M \subseteq \mathcal{M}_s\}$$

and notice that $E \in M^*$ and $w_r \in E$.

The set E may not be dense in \mathbb{P} , but

$$E_0 = \{w_t \in \mathbb{P} : \exists w_s \in E \text{ such that } w_t \leq w_s \text{ or } \forall w_s \in E (w_t \perp w_s)\}$$

is a dense subset of \mathbb{P} that belongs to M^* .

Then thanks to the (M^*, \mathbb{P}) -genericity of w_p^M and the fact that $w_r \leq w_p^M$, we have that there is a condition $w_t \in M^* \cap E_0$ that is compatible with w_r . Since $w_r \in E$ there is a condition $w_s \in E$ such that $w_t \leq w_s$. By elementarity can find w_s in M^* . Moreover, by definition of E , there is an \mathcal{M}_s such that $(\mathcal{M}_s, w_s) \in D$ and such that $\mathcal{M}_r \cap M \subseteq \mathcal{M}_s$. Again by elementarity we can find \mathcal{M}_s in M . Hence $(\mathcal{M}_s, w_s) \in D \cap M^*$.

Finally notice that w_s is compatible with w_r , because w_t is so and $w_t \leq w_s$; let w_a be the witness of it, i.e. $w_a \leq w_s, w_r$. Besides $\mathcal{M}_s \subseteq M$ and it extends $\mathcal{M}_r \cap M$, so we have that $\mathcal{M}_a = \mathcal{M}_s \cup \{M\} \cup \mathcal{M}_r \setminus M$ is a finite \in -chain of elementary submodel of $H(\theta)$. Then, in order to show that (\mathbb{M}_a, w_a) is a condition in $\mathbb{M}(\mathbb{P})$ we need to show that w_a is (N, \mathbb{P}) -generic, for every $N \in \mathcal{M}_a$. But this is true because on one hand $s \in \mathbb{M}(\mathbb{P})$ and so w_s is (N, \mathbb{P}) -generic for every $N \in \mathcal{M}_s$ and on the other hand $r \in \mathbb{M}(\mathbb{P})$ and so w_r is (N, \mathbb{P}) -generic for every $N \in \mathcal{M}_r$. Since w_a extends both w_s and w_r , we have that w_a is generic for all the models in \mathcal{M}_a . Hence a extends both s and r , in $\mathbb{M}(\mathbb{P})$, and witnesses their compatibility. \square

We now want to show that the scaffolding operation does not effect the preservation of a Souslin tree T . In order to show this fact we will use the characterization of Lemma 9.1.2.

Lemma 9.2.6. *Let T be a Souslin tree and let \mathbb{P} be a proper forcing, such that $\Vdash_{\mathbb{P}}$ “ T is Souslin”. Moreover let M^* be a countable elementary submodel of $H(\theta^*)$, for some $\theta^* > \theta$, where θ is the corresponding cardinal in the definition of $\mathbb{M}(\mathbb{P})$. If $p \in \mathbb{M}(\mathbb{P})$, $M = M^* \cap H(\theta) \in \mathcal{M}_p$ and $t \in T_{\delta_M}$, with $\delta_M = M \cap \omega_1$, then (p, t) is an $(M^*, \mathbb{M}(\mathbb{P}) \times T)$ -generic condition.*

Proof. Fix a set $D \subseteq \mathbb{M}(\mathbb{P}) \times T$ dense in M^* and fix a condition $(r, t') \leq (p, t)$, that without loss of generality we can assume to be in D . Then define

$$E = \{(w_q, h) \in \mathbb{P} \times T \mid \exists \mathcal{M}_q \text{ such that } (q, h) \in D \text{ and } \mathcal{M}_r \cap M \subseteq \mathcal{M}_q\}$$

and notice that $E \in M$ and $(w_r, t') \in E$. Again the set E may not be dense, but the set $\bar{E} = E^{\leq} \cup E^{\perp}$, where

$$E^{\leq} = \{(w_s, u) \in \mathbb{P} \times T \mid \exists (w_q, h) \in E \text{ such that } (w_s, u) \leq (w_q, h)\} \text{ and}$$

$$E^\perp = \{(w_s, u) \in \mathbb{P} \times T \mid \forall (w_q, h) \in E(w_s, u) \perp (w_q, h)\},$$

is a dense subset of $\mathbb{P} \times T$ that belongs to M^* .

Now, since $M \in \mathcal{M}_r$, the condition w_r is (M, \mathbb{P}) -generic, by definition of $\mathbb{M}(\mathbb{P})$. Moreover since $\Vdash_{\mathbb{P}}$ “ T is Souslin ” we have that (w_r, t') is $(M^*, \mathbb{P} \times T)$ -generic. Then there is a $(w_s, u) \in \bar{E} \cap M^*$, that is compatible with (w_r, t') . This latter fact then implies that $(w_s, u) \in E^\leq \cap M^*$ and so there is a condition $(w_q, h) \in E$ such that $(w_s, u) \leq (w_q, h)$. By elementarity we can find $(w_q, h) \in M^*$ and again, by elementarity we can assume $q = (\mathcal{M}_q, w_q)$ to be in M^* and so $(q, h) \in D \cap M^*$. Finally letting $\mathcal{M}_e = \mathcal{M}_q \cup \{M\} \cup \mathcal{M}_r \setminus M$, and w_e be the witness of the compatibility between w_q and w_r , we have that $e = (\mathcal{M}_e, w_e) \in \mathbb{M}(\mathbb{P})$ and that (e, t') extends both (r, t') and (q, h) . \square

Corollary 9.2.7. *Let T be a Souslin tree and let \mathbb{P} be a proper forcing. Then $\Vdash_{\mathbb{P}}$ “ T is Souslin ” implies $\Vdash_{\mathbb{M}(\mathbb{P})}$ “ T is Souslin ”.* \square

9.3 PFA(T) with finite conditions

We now show that it is possible to force an analog of the Proper Forcing Axiom for proper poset that preserve a given Souslin tree T . We will follow Neeman’s presentation of the consistency of PFA with finite conditions, from [131], arguing that a slightly modification of his method is enough for our purposes. Then we will argue that in the model we build T remains Souslin

Recall Neeman’s definition of the forcing \mathbb{A} (Definition 6.1 from [131]). Fix a supercompact cardinal θ and a Laver function $F : \theta \rightarrow H(\theta)$ as a book-keeping for choosing the proper posets that preserve T . Moreover define Z as the set of ordinals α , such that $(H(\alpha), F \upharpoonright \alpha)$ is elementary in $(H(\theta), F)$. Then let $\mathcal{Z}^\theta = \mathcal{Z}_0^\theta \cup \mathcal{Z}_1^\theta$, where \mathcal{Z}_0^θ is the collection of all countable elementary substructure of $(H(\theta), F)$ and \mathcal{Z}_1^θ is the collection of all $H(\alpha)$, such that $\alpha \in Z$ has uncountable cofinality - hence $H(\alpha)$ is countably closed. Moreover, for $\alpha \in Z$, let $f(\alpha)$ be the least cardinal such that $F(\alpha) \in H(f(\alpha))$. Notice that, by elementarity, $f(\alpha)$ is smaller than the next element of Z above α .

Definition 9.3.1. *If \mathcal{M} is a set of models in \mathcal{Z}^θ , let $\pi_0(\mathcal{M}) = \mathcal{M} \cap \mathcal{Z}_0^\theta$ and $\pi_1(\mathcal{M}) = \mathcal{M} \cap \mathcal{Z}_1^\theta$.*

With an abuse of notation we will identify an \in -chain of models with the set of models that belong to it.

Definition 9.3.2. *Let \mathbb{H}_θ^2 the poset, whose conditions \mathcal{M}_p are \in -chains of models in \mathcal{Z}^θ , closed under intersection. If $p, q \in \mathbb{H}_\theta^2$, we define $p \leq q$ iff $\mathcal{M}_q \subseteq \mathcal{M}_p$.*

See Claim 4.1 in [131] for the proof that \mathbb{H}_θ^2 is \mathcal{Z}^θ -strongly proper.

Definition 9.3.3. *Conditions in the poset $\mathbb{A}(T)$ are pairs $p = (\mathcal{M}_p, w_p)$ so that:*

1. $\mathcal{M}_p \in \mathbb{H}_\theta^2$.
2. w_p is a partial function on θ , with domain contained in the (finite) set $\{\alpha < \theta : H(\alpha) \in p \text{ and } \Vdash_{\mathbb{A}(T) \cap H(\alpha)} \text{“}F(\alpha) \text{ is a proper poset, that preserves } T\text{”}\}$.
3. For $\alpha \in \text{dom}(w_p)$, $w_p(\alpha) \in H(f(\alpha))$.
4. $\Vdash_{\mathbb{A}(T) \cap H(\alpha)} w_p(\alpha) \in F(\alpha)$.
5. If $M \in \pi_0(\mathcal{M}_p)$ and $\alpha \in M$, then $(p \cap H(\alpha), w_p \upharpoonright \alpha) \Vdash_{\mathbb{A}(T) \cap H(\alpha)} \text{“}w_p(\alpha) \text{ is an } (M[\dot{G}_\alpha], F(\alpha))\text{-generic condition”}$, where \dot{G}_α is the canonical name for a generic filter on $\mathbb{A}(T) \cap H(\alpha)$.

The ordering on $\mathbb{A}(T)$ is the following: $q \leq p$ iff $\mathcal{M}_p \subseteq \mathcal{M}_q$ and for every $\alpha \in \text{dom}(w_p)$, $(\mathcal{M}_q \cap H(\alpha), w_q \upharpoonright \alpha) \Vdash_{\mathbb{A}(T) \cap H(\alpha)} \text{“}w_q(\alpha) \leq_{F(\alpha)} w_p(\alpha)\text{”}$.

Remark 9.3.4. This inductive definition makes sense, since $\mathbb{A}(T) \cap H(\alpha)$ is definable in any $M \in \mathcal{Z}_0^\theta$, with $\alpha \in M$.

Remark 9.3.5. Condition (5) holds for α and M iff it holds for α and $M \cap H(\gamma)$, whenever $\gamma \in Z \cup \{\theta\}$, is larger than α .

Definition 9.3.6. *Let β be an ordinal in $Z \cup \{\theta\}$. The poset $\mathbb{A}(T)_\beta$ consists of conditions $p \in \mathbb{A}(T)$ such that $\text{dom}(w_p) \subseteq \beta$.*

Remark 9.3.7. In order to simplify the notation, if $p \in \mathbb{A}(T)$, then we define $(p)_\alpha$ to be $(\mathcal{M}_p, w_p \upharpoonright \alpha)$, while by $p \upharpoonright H(\alpha)$ we denote $(\mathcal{M}_p \cap H(\alpha), w_p \upharpoonright \alpha)$. Notice that $(p)_\alpha \in \mathbb{A}(T)_\alpha$ and $p \upharpoonright H(\alpha) \in \mathbb{A}(T) \cap H(\alpha)$.

Following Neeman it is possible to prove the following facts. See [131] for their proofs in the case of the forcing \mathbb{A} i.e. the poset that forces PFA with finite conditions. Indeed, the only difference between \mathbb{A} and $\mathbb{A}(T)$ is that the Laver function F picks up a smaller class of proper posets; namely the class of proper poset that preserve T .

Theorem 9.3.8. *(Neeman, Lemma 6.7 in [131]) Let $\beta \in Z \cup \{\theta\}$. Then $\mathbb{A}(T)_\beta$ is \mathcal{Z}_1^θ -strongly proper. \square*

Claim 9.3.9. (Neeman, Claim 6.10 in [131]) Let $p, q \in \mathbb{A}(T)$. Let $M \in \pi_0(\mathcal{M}_p)$ and suppose that $q \in M$. Suppose that for some $\delta < \theta$, p extends $(q)_\delta$ and $\text{dom}(w_q) \setminus \delta$ is disjoint from $\text{dom}(w_p)$. Suppose further that $(\mathcal{M}_p \cap M) \setminus H(\delta) \subseteq \mathcal{M}_q$. Then there is $w_{p'}$ extending w_p so that $\text{dom}(w_{p'}) = \text{dom}(w_p) \cup (\text{dom}(w_q) \setminus \delta)$ and so that $p' = (\mathcal{M}_p, w_{p'})$ is a condition in $\mathbb{A}(T)$ extending q . \square

Theorem 9.3.10. (Neeman, Lemma 6.11 in [131]) Let $\beta \in Z \cup \{\theta\}$. Let p be a condition in $\mathbb{A}(T)_\beta$. Let $\theta^* > \theta$ and let $M^* \prec H(\theta^*)$ be countable with $F, \beta \in M^*$. Let $M = M^* \cap H(\theta)$ and suppose that $M \in \pi_0(\mathcal{M}_p)$. Then:

1. for every $D \in M^*$ which is dense in $\mathbb{A}(T)_\beta$, there is $q \in D \cap M^*$ which is compatible with p . Moreover there is $r \in \mathbb{A}(T)_\beta$ extending both p and q , so that $\mathcal{M}_r \cap M \setminus H(\beta) \subseteq \mathcal{M}_q$, and every model in $\pi_0(\mathcal{M}_r)$ above β and outside M are either models in \mathcal{M}_p or of the form $N' \cap W$, where N' is a model in $\pi_0(\mathcal{M}_p)$.
2. p is an $(M^*, \mathbb{A}(T)_\beta)$ -generic condition. \square

Theorem 9.3.11. (Neeman, Lemma 6.13 in [131]) After forcing with $\mathbb{A}(T)$, $\text{PFA}(T)$ holds. \square

In order to show that $\mathbb{A}(T)$ preserves T , we need the following claim.

Claim 9.3.12. If $\Vdash_{\mathbb{A}(T)_\alpha}$ “ T is Souslin”, then $\Vdash_{\mathbb{A}(T)_\alpha \cap H(\alpha)}$ “ T is Souslin”.

Proof. In order to show that $\mathbb{A}(T)_\alpha \cap H(\alpha)$ preserves T , we use the equivalent formulation of Claim 9.1.2. Then, fix a countable $M^* \prec H(\theta^*)$, with $\theta^* > \theta$ and $\alpha, T \in M^*$. Then, following Remark 9.3.4, both $\mathbb{A}(T)_\alpha \cap H(\alpha)$ and $\mathbb{A}(T)_\alpha$ are definable in M^* . If $p \in (\mathbb{A}(T)_\alpha \cap H(\alpha)) \cap M^*$, then we want to show that there is a condition $p' \leq p$ such that for every $t \in T_{\delta_{M^*}}$, with $\delta_{M^*} = M^* \cap \omega_1$, the condition (p', t) is $(M^*, (\mathbb{A}(T)_\alpha \cap H(\alpha)) \times T)$ -generic.

Let $M = M^* \cap H(\theta)$ and \mathcal{M}_{p^M} be the closure under intersection of $\mathcal{M}_p \cup \{M\}$. It is easy to check that it is possible to find a function w_{p^M} with the same domain of w_p such that $p^M = (\mathcal{M}_{p^M}, w_{p^M})$ is a condition in $\mathbb{A}(T)_\alpha$ and such that $p^M \upharpoonright H(\alpha) \leq p$. We claim that $p^M \upharpoonright H(\alpha)$ is the condition we need: i.e. $(p^M \upharpoonright H(\alpha), t)$ is an $(M^*, (\mathbb{A}(T)_\alpha \cap H(\alpha)) \times T)$ -generic condition, for every $t \in T_{\delta_{M^*}}$.

To this aim fix a set $D \in M^*$ dense in $(\mathbb{A}(T)_\alpha \cap H(\alpha)) \times T$, let $t \in T_{\delta_{M^*}}$ and assume $(p^M \upharpoonright H(\alpha), t) \in D$. By Theorem 9.3.10, p^M is an $(M^*, \mathbb{A}(T)_\alpha)$ -generic condition. Then, thanks to our hypothesis, (p^M, t) is an $(M, \mathbb{A}(T)_\alpha \times T)$ -generic condition.

Now define E to be the set of conditions $(q, h) \in \mathbb{A}(T)_\alpha \times T$ such that $(q \upharpoonright H(\alpha), h) \in D$ and such that $\mathcal{M}_{p^M} \cap M \subseteq \mathcal{M}_q$. Notice that $(p^M, t) \in E$ and $E \in M^*$. The set E may not be dense, but $E_0 = E_0^{\leq} \cup E_0^\perp$, where

$$E_0^{\leq} = \{(q_0, h_0) : \exists (q, h) \in E \text{ such that } (q_0, h_0) \leq (q, h)\},$$

and

$$E_0^\perp = \{(q_0, h_0) : \forall (q, h) \in E (q_0, h_0) \perp (q, h)\},$$

is a dense subset of $\mathbb{A}(T)_\alpha \times T$ belonging to M^* .

Then there is $(q_0, h_0) \in E_0 \cap M^*$ that is compatible with (p^M, t) . Since $(p^M, t) \in E$, by definition of E_0 , there is a condition $(q, h) \in E$ that is compatible with (p^M, t) . By elementarity we can assume $(q, h) \in E \cap M^*$. Now, the key observation is that by strong genericity of the pure side conditions if (r, t) witnesses that (p^M, t) and (q, h) are compatible, then $(r \upharpoonright H(\alpha), t)$ witnesses that $(p \upharpoonright H(\alpha), t)$ and $(q \upharpoonright H(\alpha), h)$ are compatible. This is sufficient for our claim, because by definition of E and since q is finite, $(q \upharpoonright H(\alpha), h) \in D \cap M^*$. \square

We can now state and prove the main preservation theorem of this chapter.

Theorem 9.3.13. *If G is a generic filter for $\mathbb{A}(T)$, then in $V[G]$ the tree T is Souslin.*

Proof. We proceed by induction on β , proving that $\mathbb{A}(T)_\beta$ preserves T . If β is the first element of Z , then $\mathbb{A}(T)_\beta = \mathbb{M}_\theta^2$.

Claim 9.3.14. *The forcing \mathbb{H}_θ^2 preserves T .*

Proof. Let $M^* \prec H(\theta^*)$ be a countable model with $\theta^* > \theta$, containing \mathbb{H}_θ^2 and T , and let $\mathcal{M}_p \in \mathbb{H}_\theta^2$ be an $(M^*, \mathbb{H}_\theta^2)$ -generic condition, with $M = M^* \cap H(\theta) \in \mathcal{M}_p$. Moreover, let $t \in T_{\delta_M}$, with $\delta_M = M \cap \omega_1$. Thanks to Lemma 9.1.2, it is sufficient to show that (\mathcal{M}_p, t) is an $(M^*, \mathbb{H}_\theta^2 \times T)$ -generic condition.

To this aim, let $D \in M^*$ be a dense subset of $\mathbb{H}_\theta^2 \times T$ and assume, by density of D , that $(\mathcal{M}_p, t) \in D$. Then define

$$E = \{h \in T : \exists \mathcal{M}_q \in \mathbb{H}_\theta^2 \text{ such that } (\mathcal{M}_q, h) \in D \wedge \mathcal{M}_p \cap M \subseteq \mathcal{M}_q\}.$$

Since $\mathbb{M}_\theta^2, D, \mathcal{M}_p \cap M \in M^*$, we have $E \in M^*$. The set E may not be dense in T but

$$\bar{E} = \{\bar{h} \in T : \exists h \in E (\bar{h} \leq h) \vee \forall h \in E (\bar{h} \perp h)\}.$$

belongs to M^* and it is dense in T .

By (M^*, T) -genericity of t , there is an $\bar{h} \in \bar{E} \cap M$ that is compatible with t . Moreover, since $(\mathcal{M}_p, t) \in D$, we have that $t \in E$. Since $t \in E$ and $\bar{h} \in \bar{E}$

are compatible, by definition of \bar{E} , there is $h \in E$, with $\bar{h} \leq h$. By elementarity pick such an h in M^* . Then, by definition of E , there is $\mathcal{M}_q \in \mathbb{H}_\theta^2$, with $\mathcal{M}_p \cap M \subseteq \mathcal{M}_q$, such that $(\mathcal{M}_q, h) \in D$. By elementarity we can find $\mathcal{M}_q \in M^*$. Then, since \mathcal{M}_p is (M, \mathbb{H}_θ^2) -strong generic and $\mathcal{M}_p \cap M \subseteq \mathcal{M}_q$, we have that \mathcal{M}_p and \mathcal{M}_q are compatible. Finally, t and \bar{h} are compatible because $t \leq \bar{h}$ and $\bar{h} \leq h$. Hence (\mathcal{M}_p, t) and (\mathcal{M}_q, h) are compatible in $\mathbb{H}_\theta^2 \times T$ and this compatibility, together with the fact that $(\mathcal{M}_q, h) \in D \cap M^*$, witnesses that (\mathcal{M}_p, t) is $(M^*, \mathbb{H}_\theta^2 \times T)$ -generic. \square

If β is the successor of α in Z , then, by inductive hypothesis $\mathbb{A}(T)_\alpha$ preserves T . In order to show that $\mathbb{A}(T)_\beta$ also preserves T , we use the characterization of Lemma 9.1.2. Then, let $M^* \prec H(\theta^*)$ be a countable model, with $\theta^* > \theta$, containing β, F and T . Notice that $\mathbb{A}(T)_\beta$ is definable in M^* , with β as a parameter. Moreover let $p \in \mathbb{A}(T)_\beta$ be an $(M^*, \mathbb{A}(T)_\beta)$ -generic condition, with $M = M^* \cap H(\theta) \in \mathcal{M}_p$, and let $t \in T_{\delta_M}$, with $\delta_M = M \cap \omega_1$. Then we want to show that (p, t) is an $(M^*, \mathbb{A}(T)_\beta \times T)$ -generic condition.

By elementarity of M^* , $\alpha \in M^*$. Now, fix a V -generic filter G over $\mathbb{A}(T)_\alpha$, with $(p)_\alpha \in G$. By Theorem 9.3.10 $(p)_\alpha$ is an $(M^*, \mathbb{A}(T)_\alpha)$ -generic condition for M^* and so $M^*[G] \cap V = M^*$.

If $H(\alpha) \notin \mathcal{M}_p$ and p cannot be extended to a condition containing $H(\alpha)$, then $\mathbb{A}(T)_\beta$, below p , is equivalent to $\mathbb{A}(T)_\alpha$. Then, forcing below p , the conclusion follows by inductive hypothesis. Then, assume $H(\alpha) \in \mathcal{M}_p$.

Let $G_\alpha = G \cap H(\alpha)$. Then, by Theorem 9.3.8, we have that G_α is a V -generic filter on $\mathbb{A}(T) \cap H(\alpha)$, because $\mathbb{A}(T)_\alpha \cap H(\alpha) = \mathbb{A}(T) \cap H(\alpha)$. Without loss of generality, we can assume $\Vdash_{\mathbb{A}(T) \cap H(\alpha)} "F(\alpha) \text{ is a proper poset that preserves } T"$, because, otherwise $\mathbb{A}(T)_\beta$ is equal to $\mathbb{A}(T)_\alpha$ and again the conclusion follows by inductive hypothesis. Let $\mathbb{Q} = F(\alpha)[G_\alpha]$. Then, by properness of \mathbb{Q} in $V[G_\alpha]$, modulo extending p , we can assume $\alpha \in \text{dom}(w_p)$.

Fix $D \subseteq \mathbb{A}(T)_\beta \times T$ dense and in M^* . Without loss of generality assume $(p, t) \in D$. Since we will work in $V[G_\alpha]$, we need to ensure that $\Vdash_{\mathbb{A}(T) \cap H(\alpha)} "T \text{ is Souslin}"$. But this is true, by inductive hypothesis, as the Claim 9.3.12 shows.

Now, in $V[G_\alpha]$, define E to be the set of couples $(u, h) \in \mathbb{Q} \times T$ for which there is a condition $(q, h) \in \mathbb{A}(T)_\beta \times T$ such that

1. $w_q(\alpha)[G_\alpha] = u$,
2. $\mathcal{M}_p \cap M \subseteq \mathcal{M}_q$,
3. $(q, h) \in D$, and
4. $q \upharpoonright H(\alpha) \in G_\alpha$.

Notice that $E \in M^*[G_\alpha]$ and that $(w_p(\alpha)[G_\alpha], t) \in E$. The set E may not be dense, but if we define $E_0 = E_0^{\leq} \cup E_0^\perp$, with

$$E_0^{\leq} = \{(u_0, h_0) \in \mathbb{Q} \times T : \exists(u, h) \in E (u_0, h_0) \leq (u, h)\}$$

and

$$E_0^\perp = \{(u_0, h_0) \in \mathbb{Q} \times T : \forall(u, h) \in E (u_0, h_0) \perp (u, h)\},$$

we have that E_0 is dense in $\mathbb{Q} \times T$. Moreover, notice that by elementarity E_0 is in $M^*[G_\alpha]$.

Now, since $M \in \pi_0(\mathcal{M}_p)$ and $\alpha \in M^* \cap H(\theta) = M$, we have that $\Vdash_{\mathbb{A}(T) \cap H(\alpha)} "w_p(\alpha) \text{ is an } (M^*[G_\alpha], F(\alpha))\text{-generic condition}"$, where \dot{G}_α is a $\mathbb{A}(T) \cap H(\alpha)$ -name for G_α . Moreover, $\Vdash_{\mathbb{A}(T) \cap H(\alpha)} "F(\alpha) \text{ is a proper poset that preserves } T"$ and, by inductive hypothesis and Lemma 9.3.12, $\Vdash_{\mathbb{A}(T) \cap H(\alpha)} "T \text{ is Souslin}"$. Then by Lemma 9.1.2 applied in $V[G_\alpha]$ we have that $(w_p(\alpha)[G_\alpha], t)$ is an $(M^*[G_\alpha], \mathbb{Q} \times T)$ -generic condition.

Hence, there is a condition $(u_0, h_0) \in E_0 \cap M^*[G_\alpha]$ that is compatible with $(w_p(\alpha)[G_\alpha], t)$. Moreover, since $(w_p(\alpha)[G_\alpha], t) \in E$ we have that $(u_0, h_0) \in E_0^{\leq}$. This means that there is $(u, h) \in E$ such that $(u_0, h_0) \leq (u, h)$. By construction (u, h) is compatible with $(w_p(\alpha)[G_\alpha], t)$ and by elementarity we can find such a condition in $M^*[G_\alpha]$. Let $u_\alpha \in \mathbb{Q}$ be a witness of the compatibility between $w_p(\alpha)[G_\alpha]$ and u . Notice that u_α is an $(N[G_\alpha], \mathbb{Q})$ -generic condition for all $N \in \pi_0(\mathcal{M}_p)$, with $\alpha \in N$, because $u_\alpha \leq w_p(\alpha)[G_\alpha]$. Since $(u, h) \in E$ there is a condition $q \in \mathbb{A}(T)_\beta$, with $\mathcal{M}_p \cap M \subseteq \mathcal{M}_q$ and $w_q(\alpha)[G_\alpha] = u$, such that $(q, h) \in D$. By elementarity let $q \in M^*[G_\alpha]$ and so $(q, h) \in M^*[G_\alpha] \cap D$. Since $M^*[G_\alpha] \subseteq M^*[G]$ and $M^*[G] \cap V = M^*$, we have $(q, h) \in D \cap M^*$. Now, by strong genericity of the pure side conditions, letting \mathcal{M}_r be the closure under intersection of $\mathcal{M}_p \cup \mathcal{M}_q$, we have that \mathcal{M}_r witnesses that \mathcal{M}_p and \mathcal{M}_q are compatible. Moreover every model in $\pi_0(\mathcal{M}_r)$ above β and outside M are either models in \mathcal{M}_p or of the form $N' \cap W$, where N' is a model in $\pi_0(\mathcal{M}_p)$ and $W \in \pi_1(\mathcal{M}_q)$. Then u_α is an $(N[G_\alpha], \mathbb{Q})$ -generic condition, for all $N \in \pi_0(\mathcal{M}_r)$, with $\alpha \in N$, because of Remark 9.3.5 together with the fact that u_α extends both $w_p(\alpha)[G_\alpha]$ and u .

Finally, back in V , let \dot{u} and \dot{u}_α be $\mathbb{A}(T)_\alpha \cap H(\alpha)$ -names for u and u_α . Moreover, let $e \in \mathbb{A}(T)_\alpha \cap H(\alpha)$ be sufficiently strong to force all the properties we showed for q, \dot{u} and \dot{u}_α . We can also assume that e extends both $q \upharpoonright H(\alpha)$ and $p \upharpoonright H(\alpha)$. Now notice that $\mathcal{M}_e \cup \mathcal{M}_r$ is already an \in -chain closed under intersection and so if $\mathcal{M}_s = \mathcal{M}_e \cup \mathcal{M}_r$ and $w_s = w_e \cup \{\alpha, \dot{u}_\alpha\}$, we have that s is a condition in $\mathbb{A}(T)_\beta$. Hence (s, t) witnesses that (p, t) and (q, h) are compatible.

If β is a limit point of Z , let again $M^* \prec H(\theta^*)$ be a countable model containing $\mathbb{A}(T)_\beta$ and F . Then if $p \in \mathbb{A}(T)_\beta$, with $M^* \cap H(\theta) = M \in \mathcal{M}_p$, and

$t \in T_{\delta_M}$, with $\delta_M = M \cap \omega_1$, then, thanks to Lemma 9.1.2, it is sufficient to show that (p, t) is an $(M^*, \mathbb{A}(T)_\beta \times T)$ -generic condition, in order to prove that $\mathbb{A}(T)_\beta$ preserves that T is Souslin.

To this aim, let $\bar{\beta} = \sup(\beta \cap M^*)$ and let $\delta < \bar{\beta}$, in $Z \cap M^*$, be such that $\text{dom}(w_p) \subseteq \delta$. Moreover fix $D \in M^*$ dense in $\mathbb{A}(T)_\beta \times T$ and assume $(p, t) \in D$.

Now, define E as the set of conditions $((q)_\delta, h) \in \mathbb{A}(T)_\delta \times T$ that extend to conditions $(q, h) \in D$, with $\mathcal{M}_p \cap M \subseteq \mathcal{M}_q$. The set E belongs to M^* , but it may not be dense in $\mathbb{A}(T)_\delta \times T$. However the set $E_0 = E_0^{\leq} \cup E_0^\perp$ is dense in $\mathbb{A}(T)_\delta \times T$ and belongs to M^* ; where

$$E_0^{\leq} = \{(q_0, h_0) \in \mathbb{A}(T)_\delta \times T : \exists ((q)_\delta, h) \in E \text{ such that } (q_0, h_0) \leq ((q)_\delta, h)\},$$

and

$$E_0^\perp = \{(q_0, h_0) \in \mathbb{A}(T)_\delta \times T : \forall ((q)_\delta, h) \in E (q_0, h_0) \perp ((q)_\delta, h)\}.$$

Then, by the inductive hypothesis, find a condition $(q_0, h_0) \in E_0 \cap M^*$ that is compatible with $((p)_\delta, t)$. Moreover, since $((p)_\delta, t) \in E$ and it is compatible with (q_0, h_0) , we have that $(q_0, h_0) \in E_0^{\leq}$. Then, by definition of E_0^{\leq} , there is a condition $((q)_\delta, h) \in E$ such that $(q_0, h_0) \leq ((q)_\delta, h)$ and, so, that is compatible with $((p)_\delta, t)$. By elementarity pick such a condition in M^* . Moreover, thanks the fact that $\mathcal{M}_p \cap M \subseteq \mathcal{M}_q$ and that $\mathcal{M}_p \cap M$ witnesses the M -strong genericity of \mathcal{M}_p , we have that the compatibility between $((p)_\delta, t) = ((p)_\beta, t)$ and $((q)_\delta, h)$ is witnessed by a condition $((\mathcal{M}_r, w_1), t)$, where \mathcal{M}_r is the closure under intersection of $\mathcal{M}_p \cup \mathcal{M}_q$. Then we have that $\mathcal{M}_r \cap M \setminus H(\beta) \subseteq \mathcal{M}_q$, and that every model in $\pi_0(\mathcal{M}_r)$ above β and outside M are either models in \mathcal{M}_p or of the form $N' \cap W$, where N' is a model in $\pi_0(\mathcal{M}_p)$ and $W \in \pi_1(\mathcal{M}_q)$.

Now, let $(q, h) \in D$ witness that $((q)_\delta, h) \in E$. By elementarity, we can find $(q, h) \in D \cap M^*$. Then, thanks to the fact that $\mathcal{M}_r \cap M \setminus H(\beta) \subseteq \mathcal{M}_q$ we can apply Claim 9.3.9 and find a function w_2 , extending w_1 , defined as $\text{dom}(w_2) = \text{dom}(w_1) \cup (\text{dom}(w_q) \setminus \delta)$, such that $((\mathcal{M}_r, w_2), t)$ extends (q, h) . Setting $w_r = w_2 \cup w_p \upharpoonright [\bar{\beta}, \beta)$, we claim that r belongs to $\mathbb{A}(T)_\beta$.

In order to show that this latter claim holds, it is sufficient to show that if $\alpha \in \text{dom}(w_p) \upharpoonright [\bar{\beta}, \beta)$, then $p \upharpoonright H(\alpha)$ forces that $w_r(\alpha) = w_p(\alpha)$ is an $(N[\dot{G}_\alpha], F(\alpha))$ -generic condition, where \dot{G}_α is the canonical name for a V -generic filter over $\mathbb{A}(T) \cap H(\alpha)$ and $N \in \pi_0(r)$, with $\alpha \in N$. Notice that $\alpha \in N$ implies $N \notin M$. Then, since p is a condition, the claim follows thanks to Remark 9.3.5 and the fact that every model in $\pi_0(\mathcal{M}_r)$ above β and outside M are either models in \mathcal{M}_p or of the form $N' \cap W$, where N' is a model in $\pi_0(\mathcal{M}_p)$.

Hence, finally we have that (r, t) belongs to $\mathbb{A}(T)_\beta \times T$ and that, by construction, it extends both (q, h) and (p, t) . □

Conclusions

The results presented in Chapter 8 show that the method of the generalized side conditions is a very fruitful method for forcing Σ_1 sentences over $H(\aleph_3)$. Indeed, to summarize the theorems of Chapter 8, it is possible to give a uniform proof for

1. how to force a club on ω_2 with finite conditions,
2. how to force the existence of a chain of length ω_2 on $(\omega_1^{\omega_2}, <_{Fin})$,
3. how to force the existence of a thin very tall superatomic Boolean algebra,
4. how to force the existence of an ω_2 -Suslin tree.

The importance of these proofs is that a single method was used - namely the method of the side conditions with models of two types - and so it was possible to unify the results obtained by different techniques, built *ad hoc* for a single purpose. This alone is a remarkable fact, but it also hints to generalizations of the method of forcing objects with finite conditions, at least in two ways:

- First of all it is reasonable to ask for an extension of the method able to force objects of size \aleph_3 and higher; in order to generalize all the results quoted above.
- In the end of [130], Itay Neeman - who first proposed to consider side conditions of models of two types closed under intersection - suggested that this study could lead to propose a higher analog of PFA, for a specific class of forcing, say $\Gamma(\omega_1, \omega_2)$, that preserve \aleph_1 and \aleph_2 . Then, this axiom, that we may call $\text{PFA}(\omega_1, \omega_2)$ would have the following form.

Definition 9.3.15. *Given a poset $\mathbb{P} \in \Gamma(\omega_1, \omega_2)$, if \mathcal{D} is a family of dense sets in \mathbb{P} and if $|\mathcal{D}| \leq \aleph_2$, then there is a \mathcal{D} -generic filter; i.e. a filter $G \subseteq \mathbb{P}$, such that $G \cap D \neq \emptyset$, for every $D \in \mathcal{D}$.*

Ideally we would like that the all poset defined in Chapter 8 and ?? fall in the class $\Gamma(\omega_1, \omega_2)$. If this was possible it could be interesting to apply this new Forcing Axiom to sentences over $H(\aleph_3)$; and in particular study the relationship between this axiom and the size of the continuum.

The possibility of extending the Forcing Axioms would be in the same argumentative line of proposing stronger axioms of ZFC that are able to crystalize, in an axiomatic form, argument patterns that are inductively found in the practice of set theory.

9.4 A different generalization

For what concerns the possibility of adding objects of size \aleph_3 , the straightforward generalization of the pure side conditions, with models with two types, would be a three types side conditions poset: i.e. a poset whose conditions are finite \in -chains of models of size \aleph_0 , \aleph_1 and \aleph_2 , elementary in $H(\aleph_3)$, closed under intersection. However, even if the method of side conditions with countable models works very well in connection with Forcing Axioms, and although the method of the generalized side conditions turned out to be very useful in pushing, from $H(\aleph_2)$ to $H(\aleph_3)$, the possibility to force the existence of new objects with simple and uniform proofs, the hope of a simple extension of \mathbb{M}_2 has been soon frustrated. Indeed the three types side condition poset, if closed under intersection, cannot be shown to be proper for all the models involved.

Definition 9.4.1. Let $\mathcal{E}^3 = \mathcal{E}_0^3 \cup \mathcal{E}_1^3 \cup \mathcal{E}_2^3$, where

1. \mathcal{E}_0^3 is the class of countable structures $Q \prec H(\aleph_3)$,
2. \mathcal{E}_1^3 is the class of \aleph_1 -internally approachable (i.e. approached by elements of \mathcal{E}_0^3) structures $P \prec H(\aleph_3)$,
3. \mathcal{E}_2^3 is the class of \aleph_2 -internally approachable (i.e. approached by elements of \mathcal{E}_1^3) structures $M \prec H(\aleph_3)$.

Lemma 9.4.2. Let \mathbb{M}^3 the poset consisting of finite \in -chains of models in \mathcal{E}^3 , closed under intersection, ordered by reverse inclusion. Then \mathbb{M}^3 cannot be \mathcal{E}^3 -proper.

Proof. Suppose that \mathbb{M}^3 is \mathcal{E}^3 -proper¹. Let $G \subseteq \mathbb{M}^3$ be a V -generic filter and let $\vec{M} = \{M_\alpha : \alpha \leq \omega_1\}$ be the set of the first ω_1 -many \aleph_2 -models in G , enumerated

¹Notice that it is straightforward to show that \mathbb{M}^3 is \mathcal{E}_2^3 -proper, by transitivity of the \aleph_2 -models.

in increasing order. Now let P be the first \aleph_1 -model after M_{ω_1} , in G , and let Q be the first \aleph_0 -model after P , in G . Then $M_{\omega_1} \in P \in Q$ and $(M_{\omega_1}, P)_G \cap \mathcal{E}_2^3 = \emptyset$ and $(P, Q)_G \cap \mathcal{E}_1^3 = \emptyset$; where $(A, C)_G$ is defined in the obvious way as the set of models $B \in G$ such that $A \in^* B \in^* C$, with \in^* the transitive closure of the \in -relation. Moreover, thanks to our assumption, we have that $Q[G] \cap V = Q$ and that $P[G] \cap V = P$.

Define, for $i \in \{1, 2, 3\}$ the functions $f_i : G \rightarrow \omega_i$ that associate to a model $M \in G$ its sup on ω_i ; i.e. $f_i(M) = \sup(M \cap \omega_i)$. Notice that f_3 is an increasing function, since G is an \in -chain and that if R, R' are two consecutive \aleph_2 -models in G , then $f_2 \upharpoonright (R, R')_G$ is increasing. Moreover, since there are no \aleph_2 -models between P and Q , we have that $f_2(P) = P \cap \omega_2 < f_2(Q)$.

Notice that $\vec{M} \cap Q[G] = \{M_\alpha : \alpha < \delta\}$, for $\delta = Q \cap \omega_1$. Then, since $Q[G] \cap V = Q$ we have that $\{M_\alpha : \alpha < \delta\} \subseteq Q$. Moreover, since $\omega_1 \subseteq P$ we have $\vec{M} \subseteq P[G]$. Hence, by properness, $\vec{M} \subseteq P[G] \cap V = P$.

Now consider $P \cap M_\delta$ and $Q \cap M_\delta$. Since $\delta \in P$, G is closed under intersection and f_3 is increasing, we have that

$$f_3(P \cap M_\delta) > \sup \bigcup_{\alpha < \delta} (M_\alpha \cap \omega_3) = \lambda.$$

On the other hand we have that $f_3(Q \cap M_\delta) \leq \lambda$. Indeed, if this was not the case then it would be possible to correctly compute δ in $Q \cap M_\delta$ as the index of the \aleph -model after $\{M_\alpha : \alpha < \delta\}$, contradicting the fact that $(Q \cap M_\delta) \cap \omega_1 = \delta$. Indeed, since both δ and $\{M_\alpha : \alpha < \delta\}$ are subsets of $Q \cap M_\delta$, we have $f_3(Q \cap M_\delta) = \lambda$. Then

$$f_3(Q \cap M_\delta) < f_3(P \cap M_\delta).$$

The key observation now, is that since $f_3(Q \cap M_\delta) = \lambda$ are \vec{M} enumerate the first ω_1 -many \aleph_2 -models in G , there are no models of size \aleph_2 in the interval $(Q \cap M_\delta, P \cap M_\delta)_G$. This latter fact then implies that

$$f_2(Q \cap M_\delta) < f_2(P \cap M_\delta).$$

Finally, since $\omega_2 \subseteq M_\delta$, the above inequality implies

$$f_2(Q) < f_2(P),$$

contradicting our choice of P and Q . □

This negative results undermines not only the possibility of a simple generalization of the method of side conditions, in the direction of a general method for forcing objects of size \aleph_3 , but also the search for a higher analog of PFA.

As a matter of fact the results contained in [131] and presented in Chapter 9 show a strong similarity between the pure side condition posets and the theory of iterated forcing. Even if the consistency proof of PFA with finite conditions, that we find in [131], is not, properly speaking, an iteration, it shares many similarities with it. Then, following this analogy, one would expect to use a three side conditions poset in order to give a consistency proof for a higher analog of PFA. In this sense Lemma 9.4.2 points in the direction of a different structure of a consistency proof for such an axiom.

Moreover, if it exists, the class of posets for which a generalization of PFA can be proved, seems to be quite different from a straightforward extension of the class of proper posets. Indeed a higher analog of the Forcing Axioms, able to have, among its consequences, all the results presented in Chapter 8, would not be an extension of PFA, because, as was noticed by Magidor and Veličković, the possibility to add a club in ω_2 with finite conditions contradicts the club guessing principle, that can be proved, in ZFC, to hold at ω_2 .

The fact that a generalization of the Forcing Axioms is not an easy task, and requires completely new ideas, should not be surprising, because, as we noticed before, the main fact that the universe of set theory cannot be characterized by a single formula suggests that if we aim to complete ZFC in the way described by Woodin's program - i.e. step by step, climbing cardinal by cardinal the cumulative hierarchy in terms of the structures $H(\theta)$'s - then we cannot hope to extend too easily a solution found for a structure, to a larger one.

Acknowledgments

There are many people that I want to thank and whose contribution has been essential in the conception, realization and writing of this thesis. First of all my two advisors: Gabriele Lolli and Boban Veličković. They guided me, helped me and taught me so many things in these years that I owe them a debt that is not only professional, but also personal: they influenced and shaped my point of view not only on the different subjects that are discussed in this work, but also on the general methodology of my research. On one hand Gabriele Lolli made me discover the beauty and the importance of a plural approach to mathematical problems, instilling into my mind the seed of a multidisciplinary point of view, while on the other hand Boban Veličković helped me to master the difficult techniques that are necessary for the growth of an autonomous speculation on the foundations of mathematics.

Besides, I believe that my ideas and the realization of this thesis benefited by the discussion and the exchanges with other people who, in these years, contributed directly and indirectly to my growth and to the developing of this work. In particular I would like to express my gratitude to Vincenzo De Risi, José Ferreiros and Marco Panza, for the interesting conversations on Hilbert, the role of the history of mathematics and the importance of the practice of mathematics in the study of its philosophy; Uri Abraham, Justin Moore and Matteo Viale, for sharing with me their insights and discoveries on the study of the Forcing Axioms and for giving me fundamental advises in the most difficult moments of my work; Davide Crippa, Luca San Mauro and Lauri Tuomi for the interesting and fruitful collaborations and friendship.

During these four years of PhD in cotutelle there are many institutions and organizations that made possible a work between Italy and France and that allowed me to take part to many conferences and meeting, where I managed to learn, discuss and exchange knowledge and ideas. First of all I would like to thank the Scuola Normale Superiore di Pisa, that founded the first three years of my PhD and that, together with the Université Paris Diderot, allowed me to benefit of their financial supports and of their different working environments.

Moreover, I would like to thank the Università Italo-Francese, for the Bando Vinci grant that I received in 2010, the Infty network, that founded my staying in Paris during the academic year 2011-2012, the Fields Institute, that allowed me to profit of a useful month in Toronto in 2012, and the MALOA project (founded from the [European Community's] Seventh Framework Programme [FP7/2007-2013] under grant agreement n° 238381), that made possible my staying in Paris between the end of 2012 and the beginning of 2013, so that I could finish this work.

Finally I am sure that this thesis would have never been accomplished without the support, comprehension and help of friends and family that made comfortable, exciting and fundamental the last four years of my life. In particular I would like to thank Giorgio and Helene for the nice time spent in Rue du Bac and for the opportunity they gave me to better understand the importance of a nice and tasty life for a better work. The delightful and instructive experiences that I lived with them, together with Michele and Sarah, helped me to clear the meaning of friendship and of good kitchen. I would also like to thank Marta, Jérôme and Maud (thank you also for your help with the French language) for they always made me feel at home in Paris. Moreover, thank you Alberto, Angiola, Antonio, Sofia, Enrica, Francesco, Michela, Stijn, Natasha, Stefano and Luisa for making me always forget the real difference between friendship and being part of a family. Last but not least, I would like to thank Fernanda for her kindness, the naturalness that she brought in my life and because in finishing this thesis she made me understand the lightness of this job and what does it mean to devote oneself to a project.

In the end I would like to thank the people who helped me in the incomprehensible sea of bureaucracy: thank you Elisabetta Terzuoli and Regine Guittard.

Bibliography

- [1] U. Abraham, M. Rubin, and S. Shelah. On the consistency of some partition theorems for continuous colorings, and the structure of \aleph_1 -dense real order types. *Annals of Pure and Applied Logic*, 29(2):123–206, 1985.
- [2] U. Abraham and S. Shelah. Martin’s axiom does not imply that every two \aleph_1 -dense sets of reals are isomorphic. *Israel Journal of Mathematics*, 38(1-2):161–176, 1981.
- [3] U. Abraham and S. Shelah. Isomorphism types of Aronszajn trees. *Israel Journal of Mathematics*, 50(1-2):75–113, 1985.
- [4] D. Aspero, P. Larson, and J. Moore. Forcing Axioms and the Continuum Hypothesis. *Acta Mathematica*, 210:1–29, 2013.
- [5] S. Awoday and E. H. Reck. Completeness and categoricity. Part I: nineteenth-century axiomatics to twentieth-century metalogic. *History and Philosophy of Logic*, 23(1):1–30, 2002.
- [6] S. Awoday and E. H. Reck. Completeness and categoricity. Part I: nineteenth-century axiomatics to twentieth-century metalogic. *History and Philosophy of Logic*, 23(2):77–92, 2002.
- [7] J. Bagaria. Locally-generic boolean algebras and cardinal sequences. *Algebra Universalis*, 47:283–302.
- [8] J. Bagaria. Bounded forcing axioms as principles of generic absoluteness. *Archive for Mathematical Logic*, 39(6):393–401, 2000.
- [9] J. Bagaria. Natural axioms on set theory and the continuum problem. *CRM Preprint*, 591:19, 2004.
- [10] J. Baumgartner and S. Shelah. Remarks on superatomic boolean algebras. *Annals of Pure and Applied Logic*, 33(2):109–129, 1987.

- [11] J. E. Baumgartner. All \aleph_1 -dense sets of reals can be isomorphic. *Fundamenta Mathematicae*, 79:101–106.
- [12] J. E. Baumgartner. Bases for Aronszajn trees. *Tsukuba Journal of Mathematics*, 9:31–40.
- [13] M. Bekkali. Lebesgue measurability, large cardinals, forcing axioms, ρ -functions, Notes on lectures by Stevo Todorčević. In *Topics in Set Theory*. Springer-Verlag, Berlin, 1991.
- [14] P. Benacerraf. What numbers could not be. *Philosophical Review*, 74(1):47–73, 1965.
- [15] V. Benci, M. Forti, and M. Di Nasso. An aristotelian notion of size. *Annals of Pure and Applied Logic*, 143:43–53, 2006.
- [16] V. Benci and M. Di Nasso. Numerosities of labelled sets: a new way of counting. *Advances in Mathematics*, 173:50–67, 2003.
- [17] P. Bernays. On mathematical platonism. In P. Benacerraf and H. Putnam, editors, *Philosophy of mathematics: selected readings*, pages 258–271. Cambridge University Press, 1983.
- [18] G. Boolos. The iterative conception of set. *Journal of Philosophy*, 68(8):215–231, 1971.
- [19] G. Cantor. Über die Ausdehnung eines Satzes aus der Theorie der trigonometrischen Reihen. *Mathematische Annalen*, 5:123–132, 1872.
- [20] C. Chihara. A gödelian thesis regarding mathematical objects: do they exist? and can we perceive them? *The Philosophical Review*, 91(2):211–227, 1982.
- [21] J. P. Cohen. The independence of the continuum hypothesis. volume 50(6), pages 1143–1148, 1963.
- [22] D. Corfield. *Towards a philosophy of real mathematics*. Cambridge University Press, 2004.
- [23] L. Corry. Axiomatics, Empiricism, and *Anschauung* in Hilbert’s Conception of Geometry: Between Arithmetic and General Relativity. In J. Ferreiros and J. Gray, editors, *The architecture of modern mathematics. Essays in history and philosophy*, pages 133–156. Oxford University Press, 2006.

- [24] M. Detlefsen. The purity of methods. In P. Mancosu, editor, *The philosophy of mathematical practice*, pages 179–197. Oxford University Press, 2008.
- [25] W. Ewald. *From Kant to Hilbert: a source book in the foundations of mathematics. Volume II*. Clarendon Press, 1996.
- [26] I. Farah. All automorphisms of the Calkin algebra are inner. *Annals of Mathematics*, 173(2):619–661, 2010.
- [27] S. Feferman. Gödel’s program for new axioms: why, where, how and what? In P. Hájek, editor, *Gödel ‘96: Logical Foundations of Mathematics, Computer Science, and Physics. Lectures Notes in Logic 6*, pages 371–390. Cambridge University press, 1996.
- [28] S. Feferman. Does mathematics need new axioms? *American Mathematical Monthly*, 106:106–111, 1999.
- [29] J. Ferreirós. Hilbert, logicism, and mathematical existence. *Synthese*, 170(1):33–70, 2009.
- [30] J. Ferreirós. On arbitrary sets and ZFC. *Bullettin of Symbolic logic*, 17(3):361–393, 2011.
- [31] A. Fraenkel. Über die Zermelosche Begründung der Mengenlehre. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 30(2):45–46, 1921.
- [32] A. Fraenkel, Y. Bar-Hillel, and A. Levy. *Foundation of set theory*. Elsevir, 1973.
- [33] C. Francks. *The autonomy of mathematical knowledge: Hilbert’s program revisited*. Cambridge University Press, 2009.
- [34] M. Frapolli. Is cantorinan set theory an iterative conception of set? *Modern Logic*, 1(4):301–318, 1991.
- [35] G. Frege. *Grundgesetze der Arithmetik*. Verlag Hermann Pohle, 1893.
- [36] G. Frege. *Philosophical and mathematical correspondence*. Basil Blackwell, 1980.
- [37] C. Freiling. Axioms of symmetry: throwing darts at the real line. *Journal of Symbolic Logic*, 51:190–200, 1986.

- [38] H. Friedman. A consistent Fubini-Tonelli theorem for non-measurable functions. *Illinois Journal of Mathematics*, 24:390–395, 1980.
- [39] M. Friedman. Explanation and Scientific Understanding. *The Journal of Philosophy*, 71:5–19, 1974.
- [40] S. D. Friedman. Forcing with side conditions. Preprint. 2004.
- [41] S. D. Friedman and T. Arrigoni. The hyper universe program. *Bulletin of Symbolic Logic*, 19(1):77–96, 2013.
- [42] M. Giaquinto. Hilbert’s philosophy of mathematics. *The British Journal for Philosophy of Science*, 34(2):119–132, 1983.
- [43] V. Gitman, Joel David Hamkins, and T. Johnstone. What is the theory ZFC without power set? <http://arxiv.org/abs/1110.2430>.
- [44] K. Gödel. The present situation in the foundations of mathematics. In S. Feferman, J. Dawson, W. Goldfarb, C. Parsons, and R. Solovay, editors, *Collected works. Volume III Unpublished Essays and Lectures*, pages 45–53. Cambridge University press, 1933.
- [45] K. Gödel. Lecture at Zilsel’s. In S. Feferman, J. Dawson, W. Goldfarb, C. Parsons, and R. Solovay, editors, *Collected works. Volume III Unpublished Essays and Lectures*, pages 87–113. Cambridge University press, 1938.
- [46] K. Gödel. Remarks before the Princeton bicentennial conference on problems in mathematics. In S. Feferman, J. Dawson, S. Kleene, G. Moore, and J. Van Heijenoort, editors, *Collected works. Volume II 1938-1974*, pages 150–153. Cambridge University press, 1946.
- [47] K. Gödel. What is Cantor’s continuum problem? In S. Feferman, J. Dawson, S. Kleene, G. Moore, and J. Van Heijenoort, editors, *Collected works. Volume II 1938-1974*, pages 254–270. Cambridge University press, 1964.
- [48] K. Gödel. On the extension of finitary mathematics which has not yet been used. In S. Feferman, J. Dawson, S. Kleene, G. Moore, and J. Van Heijenoort, editors, *Collected works. Volume II 1938-1974*, pages 271–280. Cambridge University press, 1972.
- [49] K. Gödel. *Collected works. Volume II 1938-1974*. Oxford University Press, 1990.

- [50] M. Hallet. *Cantorian Set Theory and Limitation of Size*. Clarendon Press, 1984.
- [51] M. Hallet. Reflections on the purity of method in Hilbert’s *Grundlagen der Geometrie*. In P. Mancosu, editor, *The philosophy of mathematical practice*, pages 198–255. Oxford University Press, 2008.
- [52] M. Hallet and U. Majer. *David Hilbert’s lectures on the foundations of geometry, 1891-1902*. Springer-Verlag, 2004.
- [53] J. Hamkins. The set-theoretic multiverse. *Review of Symbolic Logic*, 5:416–449.
- [54] K. Hauser. Gödel’s program revisited. part i: the turn to phenomenology. *The Bulletin of symbolic Logic*, 124:529–590.
- [55] J. Van Heijenoort. *From Frege to Gödel: a source book in mathematical logic, 1879–1931*. Harvard University Press, 1967.
- [56] D. Hilbert. *Mathematische Notizbücher*, 3 notebooks, Niedersächsische Staats-und Universitätsbibliothek Göttingen, Handschriftenabteilung, Cod. Ms. D. Hilbert. 600:1–3.
- [57] D. Hilbert. Mathematical problems. *Bulletin of the American Mathematical Society*, 37:407–436.
- [58] D. Hilbert. “Projektive Geometrie”, lectures notes for a course held during the year 1891, in Königsberg. In M. Hallet and U. Majer, editors, *David Hilbert’s lectures on the foundations of geometry, 1891-1902*, pages 21–55. Springer-Verlag, 1891.
- [59] D. Hilbert. Über die Theorie der algebraischen Invarianten. Mathematical paper read at the international Mathematical Congress Chicago. In *Gesammelte Abhandlungen II*, pages 376–383. Springer-Verlag, 1893.
- [60] D. Hilbert. “Die Grundlagen der Geometrie”, lectures notes for a course held during the year 1894, in Königsberg. In M. Hallet and U. Majer, editors, *David Hilbert’s lectures on the foundations of geometry, 1891-1902*, pages 72–123. Springer-Verlag, 1894.
- [61] D. Hilbert. “Ferienkursus: Über den Begriff des Unendlichen”, lectures notes for a course held during the years 1898, in Göttingen. In M. Hallet and U. Majer, editors, *David Hilbert’s lectures on the foundations of geometry, 1891-1902*, pages 160–178. Springer-Verlag, 1898.

- [62] D. Hilbert. “Grundlagen der Euklidischen Geometrie”, lectures notes for a course held during the years 1898-1899, in Göttingen. In M. Hallet and U. Majer, editors, *David Hilbert’s lectures on the foundations of geometry, 1891-1902*, pages 221–286. Springer-Verlag, 1898-1899.
- [63] D. Hilbert. “Elemente der Euklidischen Geometrie”, *Ausarbeitung* form the “Grundlagen der Euklidischen Geometrie”. In M. Hallet and U. Majer, editors, *David Hilbert’s lectures on the foundations of geometry, 1891-1902*, pages 302–395. Springer-Verlag, 1899.
- [64] D. Hilbert. Grundlagen der Geometrie (*Festschrift*). In M. Hallet and U. Majer, editors, *David Hilbert’s lectures on the foundations of geometry, 1891-1902*, pages 436–525. Springer-Verlag, 1899.
- [65] D. Hilbert. *Les principes fondamentaux de la géométrie*. Gauthier-Villars, 1900.
- [66] D. Hilbert. On the concept of number (Appendix). In *The foundations of geometry*. The Open Court Publishing Company, 1900.
- [67] D. Hilbert. Über den Satz von der Gleichheit der Basiswinkel im gleichschenkligen Dreieck. volume 35, pages 50–68, 1902-1903.
- [68] D. Hilbert. *Grundlagen der Geometrie, Zweite, durch Zusätze vermehrte und mit fünf Anhängen versehene Auflage*. Teubner, 1903.
- [69] D. Hilbert. Über die Grundlagen der Geometrie. *Mathematische Annalen*, 56:381–422, 1903.
- [70] D. Hilbert. On the foundations of logic and arithmetic. In J. Van Heijenoort, editor, *From Frege to Gödel: a source book in mathematical logic, 1879-1931*, pages 129–138. Harvard University Press, 1904.
- [71] D. Hilbert. “Logische Principien des mathematischen Denkens”, lecture notes for a course held in the Sommersemester of 1905 in Göttingen. Translated as “Foundations of mathematics”. In W. Ewald, U. Majer, and W. Sieg, editors, *David Hilbert’s lectures on the foundations of logic and arithmetic, 1894-1917*, pages 131–148. Springer-Verlag, 1905.
- [72] D. Hilbert. Wesen und Ziele einer Analysis der unendlichvielen unabhängigen Variablen. In *Gesammelte Abhandlungen III*, pages 230–245. Springer-Verlag, 1909.

- [73] D. Hilbert. Axiomatic thought. In W. Ewald, editor, *From Kant to Hilbert: a source book in the foundations of mathematics. Volume II*, pages 1105–1114. Clarendon Press, 1918.
- [74] D. Hilbert. “Natur und mathematisches Erkennen”, lecture notes for a course held during the year 1919, in Göttingen. In D. Rowe, editor, *Natur und Mathematisches Erkennen: Vorlesungen, gehalten 1919-1920 in Göttingen. Nach der Ausarbeitung von Paul Bernays. I*, Birkhäuser (1992). Birkhäuser, 1919.
- [75] D. Hilbert. “Grundlagen der Mathematik”, lecture notes for a course held during the academic year 1921-1922, in Göttingen. In W. Ewald and W. Sieg, editors, *David Hilbert’s lectures on the foundations of logic and arithmetic, 1917-1933*, pages 168–201. Springer-Verlag, 1921-1922.
- [76] D. Hilbert. The new grounding of mathematics. First report. In W. Ewald, editor, *From Kant to Hilbert: a source book in the foundations of mathematics. Volume II*, pages 1117–1133. Clarendon Press, 1922.
- [77] D. Hilbert. The logical foundations of mathematics. In W. Ewald, editor, *From Kant to Hilbert: a source book in the foundations of mathematics. Volume II*, pages 1134–1147. Clarendon Press, 1923.
- [78] D. Hilbert. On the infinite. In J. Van Heijenoort, editor, *From Frege to Gödel: a source book in mathematical logic, 1879-1931*, pages 367–392. Harvard University Press, 1925.
- [79] D. Hilbert. The foundations of mathematics. In J. Van Heijenoort, editor, *From Frege to Gödel: a source book in mathematical logic, 1879-1931*, pages 464–480. Harvard University Press, 1927.
- [80] D. Hilbert. Logic and the knowledge of nature. In W. Ewald, editor, *From Kant to Hilbert: a source book in the foundations of mathematics. Volume II*, pages 1157–1165. Clarendon Press, 1930.
- [81] D. Hilbert. *The Foundations of Geometry. Translated from the 2nd German edition*. The Open Court Publishing Company, 1950.
- [82] D. Hilbert. *Grundlagen der Geometrie. Mit Supplementen von Paul Bernays. 10th Auflage*. Teubner, 1968.
- [83] D. Hilbert. *Gesammelte Abhandlungen II*. Springer, 1970.

- [84] D. Hilbert. *Foundations of geometry. (Second English edition, translated by the 10th german edition)*. Open Court Library, 1971.
- [85] J. Hintikka. Hilbert vindicated. *Synthese*, 110(1):15–36, 1997.
- [86] W. Hodges and S. Shelah. Naturality and definability i. *Journal of the London Mathematical Society*, 33(2):1–12, 1986.
- [87] I. Jané. The iterative conception of sets from a Cantorian perspective. In P. Hájek, editor, *Logic, Methodology and Philosophy of Science. Proceedings of the Twelfth International Congress*, pages 373–393. King’s College Publications, 2005.
- [88] T. Jech. *Set theory, The Third Millennium Edition, Revised and Expanded*. Springer, 2002.
- [89] A. Kanamori. *The higher infinite. Large cardinals in set theory from their beginnings*. Springer-Verlag, 1994.
- [90] P. Kitcher. Hilbert’s epistemology. *Philosophy of Science*, 43(1):99–115, 1967.
- [91] P. Kitcher. Explanation, conjunction, and unification. *The Journal of Philosophy*, 73(8):207–212, 1976.
- [92] P. Kitcher. Explanatory unification. *Philosophy of Science*, 48(4)(4):507–531, 1981.
- [93] P. Kitcher. *The Nature of Mathematical Knowledge*. Oxford University Press, 1984.
- [94] P. Kitcher. Explanatory unification and the causal structure of the world. In P. Kitcher and W. Salmon, editors, *Scientific Explanation, vol. 13 of Minnesota Studies in the Philosophy of Science*, pages 410–505. University of Minnesota Press, 1989.
- [95] F. Klain. Gutachen, betreffend den dritten Band der Theorie der Transformationsgruppen von S. Lie anlässlich der ersten Verteilung des Lobatschewsky Preises. *Mathematischen Annalen*, 50:583–600, 1897.
- [96] P. Koeller and H. Woodin. Incompatible ω -complete theories. *Journal of Symbolic Logic*, 74(4):1155–1170, 2009.
- [97] P. Koepke. Naturalness in formal mathematics. preprint.

- [98] B. König, P. Larson, J. T. Moore, and B. Veličković. Bounding the consistency strength of a five element linear basis. *Israel Journal of Mathematics*, 164(1):1–18, 2005.
- [99] P. Koszmider. On the existence of strong chains in $\mathcal{P}(\omega_1)/\text{Fin}$. *Journal of Symbolic Logic*, 63(3):1055–1062, 1998.
- [100] P. Koszmider. On strong chains of uncountable functions. *Israel Journal Mathematics*, 118:289–315, 2000.
- [101] G. Kreisel. Hilbert’s programme. *Dialectica*, 12(3-4):346–372, 2007.
- [102] K. Kunen. *Set theory. An introduction to independence proofs*. North-Holland, 1980.
- [103] M. Laczkovic. Fubini’s theorem and sierpínski’s set. preprint. 1985.
- [104] P. B. Larson. *The Stationary Tower: Notes on a Course by W. Hugh Woodin*. AMS, 2004.
- [105] E. Lask. *Fichtes Idealismus und die Geschichte*. Mohr, 1914.
- [106] J. Legris. On the epistemological justification of Hilbert’s metamathematics. *Philosophia Scientiae*, 9(2):225–238, 2005.
- [107] G. W. Leibniz. Meditationes de cognitione, veritate, et ideis. In *Philosophical essays*, pages 23–27. Hackett, 1989.
- [108] S. MacLane. *Mathematics: Form and Function*. Springer-Verlag, 1986.
- [109] P. Maddy. *Naturalism in mathematics*. Clarendon Press, 1997.
- [110] P. Maddy. Three forms of naturalism. In S. Shapiro, editor, *The Oxford handbook of Philosophy of Mathematics and logic*, pages 437–459. Oxford University Press, 2005.
- [111] P. Maddy. *Second philosophy. A naturalistic method*. Oxford University Press, 2007.
- [112] U. Majer. Husserl and Hilbert on completeness. *Synthese*, 110:37–56, 1997.
- [113] U. Majer. The relation of logic and intuition in Kant’s philosophy of science, particularly geometry. In E. Carson and R. Huber, editors, *Intuition and the Axiomatic Method*, pages 47–66. Springer, 2006.

- [114] P. Mancosu. Bolzano and Cournot on Mathematical Explanation. *Revue d'Histoire des Sciences*, 52:429–455, 1999.
- [115] P. Mancosu. Mathematical explanation: problems and prospects. *Topoi*, 20:97–117, 2001.
- [116] P. Mancosu. *The philosophy of mathematical practice*. Oxford University Press, 2008.
- [117] P. Mancosu and J. Hafner. The varieties of mathematical explanation. In P. Mancosu, K. Jorgensen, and S. Pedersen, editors, *Visualization, Explanation and Reasoning Styles in Mathematics*, pages 215–250. Springer, 2005.
- [118] P. Mancosu and J. Hafner. Beyond Unification. In P. Mancosu, editor, *The Philosophy of Mathematical Practice*, pages 151–179. Oxford University Press, 2008.
- [119] A. Marcone. Equivalenze tra teoremi: il programma di ricerca della reverse mathematics. *La Matematica nella Società e nella Cultura, Rivista dell'Unione Matematica Italiana*, 2:101–126, 2009.
- [120] C. Martinez. It is consistent that the Aronszajn lines are well quasi-ordered. Preprint. 2009.
- [121] H. Mehlberg. The present situation in the philosophy of mathematics. *Synthese*, 12:197–210, 1960.
- [122] W. J. Mitchell. Notes on a proof of Koszmider. Preprint. 2003.
- [123] W. J. Mitchell. Adding closed unbounded subsets of ω_2 with finite forcing. *Notre Dame Journal Formal Logic*, 46(3):357–371, 2005.
- [124] W. J. Mitchell. $I[\omega_2]$ can be the nonstationary ideal on $\text{Cof}(\omega_1)$. *Transactions of the American Mathematical Society*, 361(2):561–601, 2009.
- [125] T. Miyamoto. ω_1 -Souslin trees under countable support iterations. *Fundamenta Mathematicae*, 143:257–261, 1993.
- [126] D. Molinini. Toward a pluralist approach to mathematical explanation of physical phenomena. PhD Thesis.
- [127] J. T. Moore. Set mapping reflection. *Journal of Mathematical Logic*, 5(1):87–97, 2005.

- [128] J. T. Moore. A five element basis for the uncountable linear orders. *Annals of Mathematics (2)*, 163(2):669–688, 2006.
- [129] J. T. Moore. A universal Aronszajn line. *Mathematical Research Letters*, 16(1):121–131, 2009.
- [130] I. Neeman. Forcing with side conditions. slides <http://www.math.ucla.edu/~ineeman/fwsc.pdf/>. 2011.
- [131] I. Neeman. Forcing with sequences of models of two types. Preprint. 2012.
- [132] L. Nelson. Kritische Philosophie und mathematische Axiomatik. *Unterrichtsblätter für Mathematik und Naturwissenschaften*, 34:108–142, 1928.
- [133] W. Nilson and H. Meschkowski. *Georg Cantor. Briefe*. Springer Verlag, 1991.
- [134] Y. Ogawa. The pursuit of rigor: Hilbert’s axiomatic method and the objectivity of mathematics. *Annals of the Japan Association for Philosophy of Science*, 12(2):89–108, 2004.
- [135] C. Parsons. What is the iterative conception of set? In R. Butts and J. Hintikka, editors, *Logic, Foundations of Mathematics, and Computability Theory*, pages 335–367. D. Reidel, Dordrecht, 1977.
- [136] C. Parsons. Platonism and mathematical intuition in Kurt Gödel’s thought. *Bulletin of Symbolic Logic*, 1:44–74, 1995.
- [137] C. Parsons. Finitism and intuitive knowledge. In M. Schirn, editor, *The Philosophy of Mathematics Today*, pages 249–270. Clarendon Press, 1998.
- [138] A. Paseau. Proofs of the compactness theorem. *History and Philosophy of Logic*, 31:73–98, 2010.
- [139] V. Peckhaus. *Hilbertprogramm und Kritische Philosophie. Das Göttinger Modell interdisziplinärer Zusammenarbeit zwischen Mathematik und Philosophie*. Göttingen: Vandenhoeck & Ruprecht, 1990.
- [140] V. Peckhaus. The pragmatism of Hilbert’s program. *Synthese*, 137:141–156, 2003.
- [141] V. Peckhaus. Calculus Ratiocinator vs. Characteristica Universalis? The two traditions in logic, revisited. *History and Philosophy of Logic*, 25(1):3–14, 2004.

- [142] N. Phillips and N. Weaver. The calkin algebra has outer automorphisms. *Duke Mathematical Journal*, 139(1):185–202, 2007.
- [143] H. Putnam. What is mathematical truth? In *Mathematics, Matter and Method*, pages 60–78. Cambridge University Press, 1975.
- [144] W. Quine. Five milestones of empiricism. In *Theories and Things*, pages 67–72. Harvard University Press, 1981.
- [145] M. Resnik. Mathematics as a science of patterns: ontology and reference. *Noûs*, 15(4):529–550, 1981.
- [146] M. Resnik. *Mathematics as a science of patterns*. Clarendon Press, 1997.
- [147] B. Russell. *The principles of mathematics*. Cambridge University Press, 1903.
- [148] B. Russell. The regressive method of discovering the premises of mathematics. In D. Lackey, editor, *Essays in Analysis*, pages 272–283. George Allen & Unwin, 1907.
- [149] B. Russell and A. W. Whitehead. *Principia mathematica, Volume I*. Cambridge University Press, 1910.
- [150] S. Shelah. Whitehead groups may not be free, even assuming CH. I. *Israel Journal of Mathematics*, 28(3):193–203, 1977.
- [151] S. Shelah. Whitehead groups may not be free, even assuming CH. II. *Israel Journal of Mathematics*, 35(4):257–285, 1980.
- [152] S. Shelah. Logical dreams. *The Bulletin of the American Mathematical Society*, 40(2):203–228, 2003.
- [153] S. Shelah. On long increasing chains modulo flat ideals. *Mathematical Logic Quarterly*, 56(4):397–399, 2010.
- [154] R. Shore. Reverse mathematics: the playground of logic. *The Bulletin of Symbolic Logic*, 16:378–402, 2010.
- [155] T. Sider. Naturalness and arbitrariness. *Philosophical Studies*, 81:283–301, 1996.
- [156] W. Sieg. Hilbert’s programs: 1917-1922. *The Bulletin of Symbolic Logic*, 5(1):1–44, 1999.

- [157] W. Sierpiński. Sur les rapports entre l'existence des intégrales $\int_0^1 f(x, y)dx$, $\int_0^1 f(x, y)dy$ et $\int_0^1 dx \int_0^1 f(x, y)dy$. *Fundamenta Mathematicae*, 1:142–147, 1920.
- [158] J. Da Silva. Husserl's two notion of completeness. *Synthese*, 125:417–438, 2000.
- [159] S. C. Simpson. *On long increasing chains modulo flat ideals*. Cambridge University Press, 2009.
- [160] R. Solovay. A model of set-theory in which every set of reals is Lebesgue measurable. *Annals of Mathematics. Second Series*, 92(1):1–56, 1970.
- [161] R. Solovay and S. Tennenbaum. Iterated Cohen extensions and Souslin's problem. *Annals of Mathematics*, 94:201–245, 1971.
- [162] E. Steinhart. Why numbers are sets. *Synthese*, 133(3):343–361, 2002.
- [163] W. W. Tait. Gödel on intuition and on Hilbert's finitism. In S. Feferman, C. Parsons, and S. Simpson, editors, *Kurt Gödel. Essays for his centennial*, pages 249–263. Cambridge University Press, 2010.
- [164] J. Tappenden. Proof Style and Understanding in Mathematics I: Visualization, Unification and Axiom Choice. In P. Mancosu, K. Jorgensen, and S. Pedersen, editors, *Visualization, Explanation and Reasoning Styles in Mathematics*, pages 147–214. Springer-Verlag, 2005.
- [165] J. Tappenden. Mathematical concepts and definitions. In P. Mancosu, editor, *The philosophy of mathematical practice*, pages 256–275. Oxford University Press, 2008.
- [166] J. Tappenden. Mathematical concepts: fruitfulness and naturalness. In P. Mancosu, editor, *The philosophy of mathematical practice*, pages 276–301. Oxford University Press, 2008.
- [167] R. Thiele. Hilbert's Twenty-fourth problem. *America Mathematical Monthly*, 110:1–24, 2003.
- [168] S. Todorčević. A note on the proper forcing axiom. In Baumgartner, Martin, and Shelah, editors, *Axiomatic set theory*, pages 209–218. AMS, 1984.
- [169] S. Todorčević. *Partition Problems In Topology*. AMS, 1989.

- [170] S. Todorčević. A classification of transitive relations on ω_1 . *Proceeding of the London Mathematical Society*, 3(73):501–533, 1996.
- [171] S. Todorčević. A dichotomy for P-ideals of countable sets. *Fundamenta Mathematicae*, 166(3):251–267, 2000.
- [172] S. Todorčević. Biorthogonal systems and quotient spaces via Baire category methods. *Mathematische Annalen*, 335(3):687–715, 2006.
- [173] S. Todorčević. Coherent sequences. In Foreman and Kanamori, editors, *Handbook of Set Theory*, pages 215–296. North-Holland, 2010.
- [174] M. Toepell. On the origins of David Hilbert’s *Grundlagen der Geometrie*. *Archive for History of Exact Sciences*, 35(4):329–340, 1986.
- [175] M. Toepell. *Über die Entstehung von David Hilberts “Grundlagen der Geometrie”*. *Dissertation*. Göttingen: Vandenhoeck & Ruprecht, 1986.
- [176] M. Toepell. The origin and the further development of Hilbert’s *Grundlagen der Geometrie*. *Le Matematiche*, LV(1):124–153, 2000.
- [177] R. Torretti. *Philosophy of geometry from Riemann to Poincaré*. Springer, 1984.
- [178] O. Veblen. A system of axioms for geometry. *Transaction of the American Mathematical Society*, 5:439–441, 1904.
- [179] B. Veličković. Forcing axioms and stationary sets. *Advances in Mathematics*, 94(2):256–284, 1992.
- [180] B. Veličković and G. Venturi. A direct proof of the five element basis theorem. *Preprint*.
- [181] B. Veličković and G. Venturi. Proper forcing remastered. In Cummings and Schimmerling, editors, *Appalachian Set Theory*, pages 331–352. Cambridge University Press, 2012.
- [182] M. Viale. Category forcings, MM^{+++} , and generic absoluteness for the theory of strong forcing axioms. To appear.
- [183] M. Viale. A family of covering properties. *Mathematical Research Letters*, 15:221–238.
- [184] M. Viale. Martin’s maximum revisited. To appear.

- [185] M. Viale. The proper forcing axiom and the singular cardinal hypothesis. *Journal of Symbolic Logic*, 71:473–479.
- [186] H. Wang. The concept of set. In *From mathematics to philosophy*, pages 181–123. Humanities Press, 1974.
- [187] N. Weaver. Is set theory indispensable? <http://arxiv.org/abs/0905.1680>.
- [188] J. C. Webb. *Mechanicism, mentalism and metamathematics*. Kluwer, 1980.
- [189] F. Weismann. Analytic-Syntethic III. *Analysis*, 11(3):49–61, 1951.
- [190] W. H. Woodin. The Continuum Hypothesis. Part I. *Notices of the AMS*, 48(6):567–576, 2001.
- [191] E. Zermelo. *Georg Cantor. Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Springer Verlag, 1932.
- [192] E. Zermelo. *Collected works. Volume I.*. Springer-Verlag, 2010.