Title
Something Valid This Way Comes: A Study of Neologicism and Proof-Theoretic Validity

Author
Stafford, William B

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Something Valid This Way Comes: A Study of Neologicism and Proof-Theoretic Validity

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Philosophy

by

Will Stafford

Dissertation Committee:
Dean's Professor Kai Wehmeier, Chair
Associate Professor Sean Walsh
Assistant Professor Toby Meadows

2021

DEDICATION

To

All the teachers who believed in me along the way.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Prof Sean Walsh and Prof Kai Wehmeier for all their help with this project. They have been exemplary advisers. This work would have been weaker, or even nonexistent, without them. As well as being great teacher, they have also been models of collegiality and hard work, which I hope have rubbed off somewhat on me.

I would also like to thank Prof Toby Meadow's for being on the defence committee and Prof Pen Maddy and Prof Jeremy Heis for being on the advancement committee and all three of them for their feedback and insight over the years.

As well as the above, this work and I have benefited from discussions with Walter Dean, Andy Arana, Tim Button, Benedict Easterly, Marianna Antonutti Marfori, Peter Schroeder-Heister, Jeff Barrett, Simon Huttegger, JB Manchak, Brian Skyrms, and Jim Weatherall.

LPS provides a vibrant intellectual atmosphere in which this work was completed. I would like to particularly thank those who have been part of the logic wing for the ongoing conversations about this work, including Anna Bellomo, Charles Leitz, Chris Mitsch, David Mwakima, Fabio Lampert, Greg Lauro, Guillaume Masses, Jason Chen, Jeffery Schatz, Kino Zhao, Lee Killam, Nathan Gabriel, Sam Eklund, Stella Moon, Tianyu Wu, and Tom Colclough. But I also benefited from conversations ranging from interpretations of quantum mechanics to why octopuses are odd with colleagues across the department including Adam Chin, Alysha Kassam, Aydin Mohseni, Ben Feintzeig, Bruce Michael Rushing, Daniel Alexander Herrmann, David Freeborn, Elliott Chen, Ethan Galebach, Emma Cushman, Gerard Joseph Rothfus, Hannah Rubin, Helen Meskhidze, Jessica Marie Gonzalez, Jingyi Wu, Josef Kay, Kaetlin Diane Taylor, Kevin Kadowaki, Margaret Farrell, Michael Schneider, Marian Gilton, Saira Khan, Sarita Rosenstock, Timothy Schmitz, Toni Queck, and Travis Lacroix (who also provided this latex template).

I have benefited greatly from a number of research fellowships, which permitted extensive periods of focus on the research contained in this dissertation. I was supported by a Social Sciences Merit Fellowship in the Fall of 2015 and Spring of 2017, and by Associate Dean's fellowships in the Winter of 2018, the Spring of 2019 and the Winter of 2020. Additionally, I was supported by the Werner Fellowship in Winter of 2021 and the Otto W. Shaler Scholarship in the Summer of 2020.

A special thanks goes to Stella and Douglas for being friends old enough to see how far I have come and to Elliott, Kino, Hannah, Mike, Greg, Chris, and Schatz for helping me to keep going.

I would like to thank my family, Maire, Bruce, Jo, Ja, Beth, Nick, Oscar, and Mags, for their ongoing support over these many years. Finally I would like to acknowledge Anscombe, Blue, Muffin, Mister, Solo, Dot, Lady, Sophia, and Tarski for all their assistance making sure I took regular breaks.

VITA

Will Stafford

Ph.D in Philosophy **2021**
University of California, Irvine *Irvine, California*

M.A. in Philosophy **2018**
University of California, Irvine *Irvine, California*

M.Phil. in Philosophy **2015**
University of Warwick *Coventry, UK*

B.A. in Philosophy (Hons.) **2013**
University of Stirling *Stirling, UK*

PUBLICATIONS

The Potential in Frege's Theorem **FORTHCOMING**
The Review of Symbolic Logic

Proof-Theoretic Validity and Inquisitive Logic **FORTHCOMING**
The Journal of Philosophical Logic

SELECTED PRESENTATIONS

Genus as a Measure of Complexity of Proof **Nov 2020**
Midwest PhilMath Workshop

An Inquisitive Proof-Theoretic Semantics **Mar 2019**
Third Tübingen Conference on Proof-Theoretic Semantics

ABSTRACT OF THE DISSERTATION

Something Valid This Way Comes:
A Study of Neologicism and Proof-Theoretic Validity

By

Will Stafford

Doctor of Philosophy in Philosophy

University of California, Irvine, 2021

Dean's Professor Kai Wehmeier, Chair

This dissertation consists of three chapters:

Chapter 1 Is a logicist bound to the claim that as a matter of analytic truth there is an actual infinity of objects? If Hume's Principle is analytic then in the standard setting the answer appears to be yes. Hodes's work pointed to a way out by offering a modal picture in which only a potential infinity was posited. However, this project was abandoned due to apparent failures of cross-world predication. We re-explore this idea and discover that in the setting of the potential infinite one can interpret first-order Peano arithmetic, but not second-order Peano arithmetic. We conclude that in order for the logicist to weaken the metaphysically loaded claim of necessary actual infinities, they must also weaken the mathematics they recover.

Chapter 2 There have been several recent results bringing into focus the super-intuitionistic nature of most notions of proof-theoretic validity. But there has been very little work evaluating the consequences of these results. In this chapter, we explore the question of whether these results undermine the claim that proof-theoretic validity shows us which inferences follow from the meaning of the connectives when defined by their introduction rules. It is

argued that the super-intuitionistic inferences are valid due to the correspondence between the treatment of the atomic formulas and more complex formulas.

Chapter 3 Prawitz (1971) conjectured that proof-theoretic validity offers a semantics for intuitionistic logic. This conjecture has recently been proven false by Piecha and Schroeder-Heister (2019). This article resolves one of the questions left open by this recent result by showing the extensional alignment of proof-theoretic validity and general inquisitive logic. General inquisitive logic is a generalisation of inquisitive semantics, a uniform semantics for questions and assertions. The chapter further defines a notion of quasi-proof-theoretic validity by restricting proof-theoretic validity to allow double negation elimination for atomic formulas and proves the extensional alignment of quasi-proof-theoretic validity and inquisitive logic.

Introduction and Background

The problem now becomes that of finding the proof of the proposition, and following it back to the primitive truths. If, in carrying out this process, we come only on general logical laws and on definitions, then the truth is an analytic one.
— Frege, *Die Grundlagen der Arithmetik 1884*

This dissertation explores two attempts to explain the oft-claimed special character of mathematical and logical knowledge. These projects are logicism in the philosophy of mathematics and proof-theoretic validity in the philosophy of logic. There is a naturalness in treating these two projects together. For, despite their differences, they both share a wish to recognise the special character of mathematical or logical knowledge by showing how it follows from our understanding of the concepts involved. We might gloss this as the knowledge being analytic. Both projects do this by appealing to expansive and not uncontroversial notions of definition.

This dissertation consists of three chapters exploring the limits of these two projects. In both cases we find the formal results offer a mixed blessing to the philosophical projects. In the first chapter, we explore logicism and a response to the ontologically inflationary nature of their distinctive definitions by making a connection to modern work on the potentially infinite. The next two chapters address proof-theoretic validity. The second chapter examines the consequences of recent results which show that proof-theoretic validity isn't a semantics for intuitionistic logic. The third chapter shows that it is, in fact, a semantics for generalised inquisitive logic. The rest of this introduction provides a background to this work and a more detailed summary of the results.

0.1 Logicism

We turn first to logicism. Logicism is the view that mathematics can be reduced to logic and definitions. Logicism has its origins in the work of Frege, who over the course of his career developed a background logic, motivated the reduction of arithmetic to logic, and then tried to offer a formal proof of this result. Frege's logicism was restricted to arithmetic and analysis (by contrast, he thought geometry was not so reducible).

At the core of Frege's idea is a peculiar type of definition called abstraction principles. Frege motivates them with the example of parallel lines (Beaney 1997, p. 111). If you were to say what makes two lines parallel, you might say their direction. But we can instead take things the other way around. Lines form equivalence classes of parallels. Every line is parallel to itself (reflexive). If one line is parallel to another then the other is parallel to the first (symmetric). And if one line is parallel to a second and the second is parallel to a third then the first is also parallel to the third (transitive). Which is all to say that we can group all and only the lines that are parallel to one another. And what do each of the lines in a group share with one another? The simple answer is their direction. An abstraction principle gives the identity conditions of a property like direction based on an equivalence relation like 'parallel'. For this case, the abstraction principle would be:

> The direction of line a is the direction of line b iff a is parallel to b.

Frege's insight is that we can use an abstraction principle like the above to introduce or define a new kind of object. In this case, we could use the relation parallel to define directions. This is how abstraction principles are used as definitions.

In §64 of the *Grundlargen*, Frege considers whether cardinality can be defined by the following abstraction principle called Hume's Principle:

> The number of A is the number of B iff there is a bijection from A to B.

A bijection is a function which is one-to-one (injective) and onto (surjective). It takes every item in the domain to exactly one item in the range and every item in the range is the result of applying the function to some item in the domain. Hume's Principle can be motivated by considering how we might check without counting that there are the same number of knives and forks. We can do this by lining the knives and forks up one next to the other. If every fork is next to a knife, and every knife is next to a fork, then we know we have the same number of each. And in this case, we have described a bijection between the two.

Frege does not, in fact, use Hume's Principle as a definition, instead deciding it needs proof (Frege 1884, §64). His reason has become known as the Caesar problem because he motivates it by asking how we know that Julius Caesar is not identical to the number 4 (Frege 1884, §56). Hume's Principle only gives us identity conditions for numbers when we consider them as the number of a particular concept. It does not tell us how to deal with identity statements of the form 'The number of countries in the United Kingdom is Julius Caesar.' Modern logicists of the Fregean sort do not take this worry as seriously as Frege did. A common response is that identity conditions like the above are ruled out by sortal considerations (Hale and Wright 2001b). Julius Caesar is just a different sort of thing from a number and so identity conditions concerning both are necessarily false.

For Frege however, the concern led to the dropping of Hume's Principle in favour of another abstraction principle called Basic Law V which was supposedly a law of logic. Basic Law V regards the extension of concepts (Frege 1893). The equivalence relation it uses is 'having the same objects fall under'. So, for example, 'US presidents as of 2020' and 'male US presidents as of 2020' are different concepts for easy counterfactual reasons (Hilary could have been president) but they have the same objects falling under them. Basic Law V uses that to define the extensions of concepts.

The extension of A is the extension of B iff A has the same objects falling under it as B.

Then Hume's Principle can be proven from Basic Law V. Using Hume's Principle one can define zero, successor, plus, times, and natural number and show that the standard axioms of arithmetic (Peano arithmetic) hold of those defined terms. But there is an unresolvable problem with the project as laid out so far. This is because Basic Law V is inconsistent. It was Russell who discovered this and sent the proof to Frege (Russell 1967). The inconsistency is as follows. We have the concept $R = \{C \mid C \text{ is not in the extension of } C\}$, where 'the extension of' is defined by Basic Law V. Then we can ask if R is in the extension of R. If it were then it follows that it is not and if it was not then it follows that it is.

This inconsistency was thought to be fatal to logicism in the Fregean spirit. However, it was discovered by Wright (1984) that the proof in Frege's work appeared to only use Basic Law V to derive Hume's Principle and then carried on without it. It was later confirmed that there was a consistent proof of what is called Frege's Theorem (Heck 1993): the result that arithmetic is reducible to Hume's Principle and second-order logic. Or more properly that Peano arithmetic can be interpreted in Hume's Principle plus second-order logic. This led to a revival of the logicist project with a new challenge: to explain why Hume's Principle can be used given that Frege himself did not think so.

We can describe the challenge faced by the neologicist as follows: To show something is analytic you are allowed two sets of tools, the first are principles that are taken to be a part of the meaning of the terms you use. In the luckiest scenario, these are stipulative definitions. The second tool is what resources can be applied to see what follows from the first category. And the logicist needs to explain where its various principles fall.

For the logicist, there are the definitions of natural number, successor and so on that fall in the first category. And they are good candidates for such work because they have nice properties for definitions such as conservativity. And in the second category sits second-order logic, which is a controversial choice of tool to explicate following from because the logicist wants to claim it is simply logic, but others see it as set theory in disguise. However, at least second-order logic is certainly the right kind of tool for the second category. The real challenge is with Hume's Principle, which controls the behaviour of the cardinality operator in the system. Hume's Principle does not appear to be a good candidate for logic, that is to fall into the second category. If it was, then the logicist position would not be interesting. That arithmetic follows from logic given that cardinality is logical is a weaker view. So, it seems it must fall into the first category, that is it must be justified by our understanding of cardinality or must, for these purposes, stipulatively define cardinality. And it is fair to say that this is the route taken by neologicists where Hume's Principle is, for example, described as analytic of our concept of number (Wright 1984).

But there are two prominent objections to using Hume's Principle in this manner. The first is called the Bad Company Objection (Dummett 1991a, pp. 188–9). The basic idea is that Hume's Principle cannot be a method of definition because acceptable definition schemas have to satisfy the principle: for any substitution of definiens and definiendum, the resulting principle is consistent. The motivation for this is relatively easy to spell out. If a definition is something we can just stipulate then the very minimum condition one can place is that it should not result in falsehoods. Yet an inconsistent principle is as far away from this ideal as is possible. It results in all falsehoods. This relatively lenient condition is violated by abstraction principles because, as we saw, Basic Law V is inconsistent. The Bad Company Objection then argues that Hume's Principle cannot be a definition because the schema it is a case of can be used to generate inconsistent principles. This has led to a large literature attempting to restrict abstraction principles to some consistent collection which are acceptable principles for definition. Of course, these matters are not as clear cut as they are presented here, and consistency is always relative to a system. Basic Law V for example is consistent in a setting with predicative comprehension.

We will not attempt to tackle the Bad Company Objection here but rather try to respond to a distinct worry about ontological inflation (Boolos 1997). To explain this worry we need to see that Hume's Principle has hidden existential import. On any model in which Hume's Principle is true, there will be an infinity of objects in the domain. This is because in any model the fact that the cardinality operator is treated as a function gives you one object (the number of the empty concept). Then, having one object, it follows that there is the collection of only that object which Hume's Principle tells us has a number distinct from the number of the empty concept. And so, you have two distinct objects. But now the number of the collection of those two objects is, via Hume's Principle, distinct from the previous two objects and so on until you have an infinity of objects. It is argued that no definition or analytic truth can have this existential nature. We cannot define into existence an infinity of object nor is it known *a priori* that such an infinity exists. It is this objection which is addressed in Chapter 1.

It is fair to say that fewer solutions have been proposed to address the inflationary worry than the problem of Bad Company. This can be seen in Hale and Wright's response:

> To require of an acceptable abstraction that it should not be (even) weakly inflationary [that is require a countable infinity] would stop the neo-Fregean project

> dead in its tracks, before it even got moving (as it were). It will be clear that I think there is no good ground to impose such a requirement, and I shall not discuss it further. (Hale and Wright 2001a, pp. 417–8)

In Chapter 1, a potential line of response to the inflationary worry is evaluated. It builds on a suggestion considered and then abandoned by Hodes, that the logicist might benefit from turning to the modal setting.

The intellectual background to Hodes's project lies in Putnam, who put forward the idea that mathematics can be founded in modal logic and that this foundation may deserve something like the title "logicism". Putnam is motivated by scepticism of Platonism (Putnam 1967a, p. 11; Putnam 1967b, p. 17), while still wishing to retain a bivalent logic (Putnam 1967a, p. 16). Putnam distinguishes between two pictures of mathematics. The first is *the mathematical object picture*, which is the view that mathematics reduces to set theory and possibly arithmetic and that these theories make true existential claims about sets and numbers (Putnam 1967a, p. 9). The second is *the modal logical picture*, on which mathematical propositions have the form '$\Box(\varphi(\bar{P}) \rightarrow \psi(\bar{P}))$', where $\bar{P}$ are all the relations occurring in the relevant axioms and the theorem, $\varphi(\bar{P})$ is the conjunction of the relevant axioms and $\psi(\bar{P})$ is the theorem (Putnam 1967a, pp. 9–10). This view is a modal version of Russell's "if. . . thenism" (Russell 1903, §5) because of this Putnam thinks it doesn't entail the actual existence of mathematical objects.

Putnam claims that the modal logical picture and the mathematical object picture are in a sense equivalent. This is because we can translate the existential claims made by the mathematical object picture into a modal claim without changing the truth value of mathematical claims:

> 'Numbers exist'; but all this comes to, for mathematics anyway, is that (I) ω-sequences are possible (mathematically speaking); and (2) there are necessary truths of the form 'if α is an ω-sequence, then…'[.] (Putnam 1967a, pp. 11–12)

So we avoid any appeal to the actual existence of mathematical objects and instead appeal to their possibility.

Hodes offers a view that is similar to Putnam's: on both theories, the mathematical object picture turns out to be just another way of talking about an ontologically less demanding modal theory (Hodes 1984, pp. 148–9; Hodes 1990a, p. 248). However, Hodes is far more sceptical about the existence of an infinity of numbers than Putnam is. Hodes is concerned that without the mathematical object picture there is no reason to think that there are infinitely many objects (Hodes 1984, pp. 148–9; Hodes 1990a, pp. 248, 259). He rejects the idea that we can justify the existence of an infinity of things by appeal to the infinity of points in space or the like (Hodes 1984, p. 148). And given this, he is concerned that mathematics appears to make claims that are true only on infinite models. *Prima facie* the truth of such statements requires the existence of an infinity of objects and so could be false if such objects do not exist.

Hodes's response to this worry is to urge that '[a]rithmetic should be able to face boldly the dreadful chance that in the actual world there are only finitely many objects' (Hodes 1984, p. 148). This makes Hodes one of the few in the tradition following Frege to take the inflationary worry seriously. His solution is to appeal to modality and in particular, the modality that seems to be implicit in our idea of numbers; the idea of it always being possible to add one (Hodes 1990b, p. 378).

Hodes views the uses of number-words as a *façon de parler*. They are a useful way of talking about number quantifiers. So, a statement such as '$2 + 2 = 4$' is just a manner of speaking about the fact that $[\exists_2 x\ Px \wedge \exists_2 x\ Qx \wedge \forall x \neg(Px \wedge Qx)] \rightarrow \exists_4 x\ (Px \vee Qx)$ (Hodes 1984, p. 144;

Hodes 1990a, p. 247; Hodes 1990b, pp. 364–5; Hodes 1991, p. 160). The picture is completed when we add the modal component to allow for the possibility that they are arbitrarily large finite sets. Hodes (1984, p. 149) gives an example of the arithmetic statement '7+5=12' on this picture :

$$\forall X \forall Y \Box((\exists_5 x\ Xx \wedge \exists_7 x\ Yx \wedge \neg\exists x(Xx \wedge Yx)) \rightarrow \exists_{12} x(Xx \vee Yx)). \tag{1}$$

This leads him to claim in 1984 that 'Mathematics is higher-order modal logic' (Hodes 1984, p. 149). We can understand Hodes as attempting to offer a view in the spirit of Frege and Fregean neo-logicism. As it offers a picture on which one could commit to the truth of Hume's Principle as a definition of number while rejecting the idea that it implied that any actual objects were in fact numbers. However, by 1990 Hodes writes 'I tentatively conclude that an Individual-Actualist who accepts the Alternative theory does best to accept an actual infinitude' (Hodes 1990b, p. 391). This, he claimed, was due to problems with cross-world predication.

Distinct from Hodes' work there have been advances in formally modelling the potentially infinite. Linnebo has shown that one can address philosophical questions related to naïve comprehension by working in a modal setting where the set defined by comprehension maybe merely possible (2013; 2018). This allows certain paradoxes to be avoided and set theories can be interpreted into this setting. Linnebo (2018) also considers the case of arithmetic but he considers ordinals, not cardinals like those generated by Hume's Principle. Linnebo's success raises the question of whether a project like Hodes' can be reattempted in a modified modal setting.

In Chapter 1 I attempt to achieve this. There it is shown that by placing Hume's Principle in a setting modelling the potentially infinite a weak version of Frege's Theorem can be proven. In particular, the theorem can be proven for first-order arithmetic. However, it is shown that it cannot be strengthened to second-order arithmetic. This result gives a concrete answer to the role the assumption of an actual infinity of objects is having in Frege's theorem. It is not, as Wright suggested, that the project cannot even get of the ground without this assumption, but rather this assumption (and in particular the set of all numbers) is needed for the higher-order content of the theorem. This does not necessarily close the logicist off from committing to only a potential infinity. But if they choose too they will have to restrict what mathematics they take to be analytic.

0.2 Proof-Theoretic Validity

The logicist in the philosophy of mathematics essentially presupposes that logic is in good standing. It is the tool they use to discover what flows from their definitions. But why it can be assumed is left unexplained. And as we saw one of Frege's own axioms of logic was inconsistent. This may suggest that something more needs to be said about the nature of logic. The next project we turn to does just that. The idea of using proof-rules as definitions is older than the tradition of proof-theoretic semantics but within that tradition, it has become clearly articulated. The simple idea is that certain terms, such as the logical connectives, can be defined in terms of proof-rules. Once so defined logical tautologies, validities, and consequence relations will then follow from the definitions.

The most naïve approach to this allows any set of rules to be used to define a connective. But it was shown by Prior that this isn't possible (Prior 1960). He offered a connective, TONK, that had the introduction rule for disjunction and the elimination rules for conjunction. In a system with at least one tautology, this connective allows anything to be proven by its introduction and then elimination. We can think of this as starting something like a Bad Company Objection for proof-theoretic semantics. It is natural to ask how proof-rules can be used as definitions if they can be inconsistent? And as with the Bad Company Objection, this leads to a series of attempts to remove the poorly behaved definitions. Belnap (1962) gave the influential suggestion that we should only be allowed to add sets of rules that are conservative. A rule is said to be conservative if when added to a system it does not allow us to prove any statement in the old vocabulary that could not be proven before it was added. Informally we can think of a conservative set of rules as only saying things about what they themselves define. It is fair to say that spelling out which sets of rules can be taken as definitions and which cannot is the main concern of proof-theoretic semantics.

Under that banner, there are two complementary approaches. The first involves the instituting of certain conditions on acceptable sets of rules, most notably harmony, and the second, proof-theoretic validity, involves taking only a smaller set of rules as definitional and using that to generate which other rules can be used. What is true on both approaches is that definitions are given by proof-rules. What is less clear is what resources they assume we are allowed to extract. For the logicist, this was of course logic, but it would not do for that to be assumed by the proof-theoretic semanticist. Instead, they must assume some notion of consequence, or 'following from' that is more primitive, epistemologically, than the logic they are trying to demonstrate follows from the definitions.

For the first approach, it is relatively easy to spell out what ‘following from’ amounts to. Tautologies, validities and so on follow from the definitions if the relevant witness in the form of a proof can be constructed by concatenating the rules together. While finding such a witness will be as difficult as finding proofs in various systems , it seems clear that the idea of concatenation of rules is a more primitive explication of following from than a full-blown logic. As spelling out this second condition will prove to be a matter of some difficulty for proof-theoretic validity it is worth pausing to explore why the second approach is needed at all.

The first approach generally requires that the rules for a connective be broken into two groups, the introduction and elimination rules, and that these two groups of rules are in harmony with one another. In its broadest possible sense harmony is the property that Prior's counterexample TONK failed to have: it is the requirement that the elimination rule is not stronger than the introduction rule. The notion of harmony first appears in Dummett's work where he offers two explications of it. The first, in agreement with Belnap, is that conservative sets of rules are harmonious (Dummett 1991b, p. 219). The second is that the introduction and elimination rules are harmonious if local detours (where a connective is introduced then eliminated) can be removed (Dummett 1991b, pp. 247–8). However, Dummett's final position is the rather disappointing view that harmony is an intuitive property not easily formalized:

> The two complementary features of any [linguistic] practice ought to be in harmony with each other: and there is no automatic mechanism to ensure that they will be. The notion of harmony is difficult to make precise but intuitively compelling: it is obviously not possible for the two features of the use of any expression to be determined quite independently. (Dummett 1991b, p. 215)

Here we consider what a formalisation of Dummett's notions of harmony might look like. Steinberger claims that the notion of harmony requires two features (Steinberger 2011, p. 620). First, an *appropriate* balance between the grounds used to introduce the connective and the consequences of eliminating it. This can fail in two ways: *Weak E-disharmony* occurs when the elimination rule does not make full use of the power of the introduction rule. *Strong E-disharmony* occurs when the elimination rule is too strong and allow the drawing of conclusions not justified by the introduction rule. Second, no new consequences should be deducible which do not contain the new connective. In other words, it should be conservative.

We get the following suggestions on how to formalise harmony from Steinberger.

1. *Total Harmony* A connective is *totally harmonious* relative to a system if its addition to the system is a conservative extension.

2. *intrinsic harmony* A connective is *intrinsically harmonious* if there is a systematic way to remove detours in proofs. Another name for this is local peaks levelling.

3. *ideal harmony* This is intrinsic harmony plus stability. Stability comes from Dummett's work and is supposed to stop an elimination rule from being too weak but also has no formal definition. It will be discussed below.

4. *Normalizability* A system of connectives are *normalizable* just in case they have a normal form theorem.

The goal now is to assess these different notions. First, let us turn to total harmony. Steinberger has two objections to this notion of harmony. The first is that it doesn't protect against weak E-disharmony. A conservative system can be ensured by adding a connective with an elimination rule that derives much weaker conclusions that the introduction rule would permit (Steinberger 2011, p. 265). The second problem with this notion, according to Steinberger, is that it doesn't relate to the right things. Harmony should be a property of a connective but total harmony is a property of a connective and a system of other connectives and axioms. Whether or not a connective is conservative will depend on the system it is added too (Steinberger 2011, p. 265).

Let us now turn to intrinsic harmony. The first problem with intrinsic harmony is that it also doesn't protect against weak E-disharmony (Steinberger 2011, p. 269). The second objection is that the meaning of connectives which are intrinsically harmonious can change. Steinberger illustrates this idea with an example. Take quantum 'or', which we will write $*$. The connective $*$ is governed by the rules for disjunction with one modification, in both premises of $*$-elimination the hypothesis discharged must be the only assumption used. These rules are intrinsically harmonious. Now $*$ differs from $\vee$ by being weaker. Of particular interest, the distributive laws don't hold; that is $A \wedge (B * C)$ does not imply $(A \wedge B) * (A \wedge C)$. Steinberger points out that if one adds $\vee$ to a system consisting of $\wedge$ and $*$ the resulting system is not a conservative extension as one can then prove the distributive law for $*$. The addition of $\vee$ to the system essentially collapses $*$ into $\vee$, as one can now prove the unrestricted 'or' elimination rule for $*$. It is for this reason that $*$ is claimed to be at fault (Steinberger 2011, pp. 268–9).

Steinberger says that normalization is preferable to conservativeness as an explication of harmony intended to "ensure the global well-functioning of the logical fragment" (Steinberger 2011, p. 632). Dummett thought that normalization implied conservativeness. But, as Steinberger points out, there are normalization results for classical logic. And the addition of classical $\neg$ is not a conservative extension of the system without it. So classical logic is not conservative in the relevant sense (Steinberger 2011, pp. 633–4). Further, normalization, like conservativeness, is a property of a system as a whole, not of individual connectives and this makes it unacceptable for him (Steinberger 2011, p. 634).

Ideal harmony, the last of notions of harmony we will look at, is easily put aside. Steinberger does not define stability and there has been little progress towards such a definition in the literature. Instead stability is a black box in which a future explication of Dummett's notion can be placed. As it isn't clear what ideal harmony would amount to, it is not currently an adequate notion of harmony.

This situation suggests that another approach is needed if we want a formal (as opposed to informal) account of proof-theoretic semantics. This leads to the second approach, proof-theoretic validity. The hope of proof-theoretic validity is that it will offer a clear way of finding introduction and elimination rules that have the desired property such as harmony.

Proof-theoretic validity was first put forward by Prawitz building on ideas found in Gentzen. Gentzen suggested the following claim about the connection between the introduction and elimination rules:

> The introductions represent, as it were, the 'definitions' of the symbols concerned, and the eliminations are no more, in the final analysis, than the consequences of these definitions. This fact may be expressed as follows: In eliminating a symbol, we may use the formula with whose terminal symbol we are dealing only 'in the sense afforded it by the introduction of that symbol'. (Gentzen 1935, p. 80)

Using our analysis from earlier we can take Gentzen to be claiming that the introduction rules fall in the first category of definitions. How we then find what follows from them is alluded to but left undefined. What Prawitz did was to spell out precisely what follows from the definitions.

The key idea behind proof-theoretic validity is that a rule should be permitted if whenever there is a proof containing it we can find another proof with the same conclusion without the rule. The actual definition is more complicated because a condition needs to be added for proofs that contain assumptions. The definition as we work with it in this dissertation is:

Definition 0.1. *(Prawitz 1973, p. 236; Schroeder-Heister 2006, pp. 543–4) An argument $\mathcal{D}$ is an S-*valid derivation *for a set of rules S describing the behaviour of the atomic formulas if one of the following conditions holds:*

Atomic case *If $\mathcal{D}$ is a closed argument ending in an atomic formula then it is S-valid if it contains only rules in S.*

Closed introduction case *If $\mathcal{D}$ is a closed argument ending in an introduction rule then it is S-valid if the arguments for the premises of the introduction rule are S-valid.*

Closed non-introductory case *If $\mathcal{D}$ is a closed argument which does not end in an introduction rule then it is S-valid if there is a S-valid argument with the same conclusion which does end in an introduction rule.*

Open case *If $\mathcal{D}$ is an open argument of A with open assumptions $A_0, \ldots, A_n$ it is S-valid if for all S' which are acceptable extensions of S and all closed S'-valid arguments $\mathcal{D}_0, \ldots, \mathcal{D}_n$ of $A_0, \ldots, A_n$, the following argument is S'-valid:*

$$\begin{array}{ccc} \mathcal{D}_0 & \ldots & \mathcal{D}_n \\ A_0 & \ldots & A_n \\ & \mathcal{D} & \\ & A & \end{array}$$

We will discuss in detail the motivations for the definition later in Chapter 2. But it is worth pausing to spell out what role I take this definition to play. Proof-theoretic validity is supposed to spell out what tools we are allowed to use to figure out what follows from the definitions we are given, in this case, introduction rules.

It was mentioned that this should be epistemically better justified than the logic it is supposed to justify. And we saw in the case of harmony the relatively uncontroversial choice was concatenation. At first glance proof-theoretic validity is a complex and confusing notion. What reason could we have to think that we can help ourselves to it? I take the answer to be that while appearing complicated, proof-theoretic validity is justified by the theory of meaning that the proof-theoretic semanticist is committed to. So the atomic case is justified because non-logical terms are taken to have inference rules as there meanings as well. The closed introduction case is justified because introduction rules give the definitions of the connectives. And the open case is justified because the meaning of an assertion is given by a proof and so the meaning of an assumption is given by hypothesising that there is a proof.

I think a good argument can be given that, for an inferentialist about meaning, these conditions are justified by the theory of meaning they subscribe to. And as such they are more primitive epistemically than the logic justified. However, the condition for closed proofs that do not end in an introduction rule is harder to justify. In most presentations of proof-theoretic validity, this condition would take the form:

Closed non-introductory case* If $\mathcal{D}$ is a closed argument which does not end in an introduction rule then it is S-valid if there is a S-valid argument $\mathcal{D}'$ with the same conclusion which does end in an introduction rule **and a set of transformations $\mathcal{T}$ such that $\mathcal{D}'$ is the result of applying the transformations to $\mathcal{D}$.**

Here we see the notion of a transformation added. You can think of a transformation on a proof as a function that takes a proof as input and returns another proof. Perhaps the case of closed non-introductory proofs can be justified by appeal to transformation. If so, the use of transformations must be shown to be an acceptable means of discovering what follows from the meaning of the introduction rules. For this discussion, we will talk about formal arguments, or just arguments, when we want to refer to a potential formal proof which may or may not be valid.[1] We will use the term valid proof, or just proof, for those arguments that satisfy the conditions of proof-theoretic validity.

A quick note on what transformations are. It is agreed that transformations must preserve the conclusion of the argument and that they must not add assumptions. But after that point, there is little agreement on the requirements. Both Dummett and Prawitz claim that transformations must be effective (Prawitz 2006, p. 514; Dummett 1991b, p. 264). However, that is the only additional condition Dummett places, while Prawitz has several others. Firstly, he requires that transformations commute with substitution. This means that we must not end up with a different proof if we first substitute one formula for another and then apply a justification vs. applying the justification then the substitution. This condition is also required by Schroeder-Heister. Prawitz has one further requirement on transformations and that is that there must never be two transformations that could be applied at one time.

[1]Formally an argument is a tree with points labelled by formulas and a discharge function.

This condition appears to be placed to make proofs of normalisation easier and does not appear to have a robust philosophical explanation (Schroeder-Heister 2006).

Prawitz initially offers the following justification of the transformations used in normalisation: the applications of the transformations preserve identity of proof. If this were true it would be a great explanation of why the reductions were meaning preserving. After all, if one proof is valid due to the meanings of the terms involved and a apparently distinct proof, despite apparent differences, is the same proof in another guise, then it is still valid due to the meanings of the terms involved. But why think this is true? In the confines of the justification involved in normalisation, it is easy to think of the transformations as just removing unnecessary steps unrelated to the reasoning displayed by the proof. You can think of two proofs of the same conclusion from the same premises as being distinct if they show different methods of reasoning from the premises to the conclusion. Given this picture, it is tempting to think that normalisation preserves identity because the inclusion of a detour does not so much show a different way of reasoning as an error in a chain of reasoning which is really identical to the normalised proof. We should be suspicious of this claim, however, because non-normal proofs can be substantially shorter, and we may see this neither as an error nor as a replication of the reasoning in the related normal proof.

The second argument in favour of identity can be given via the Curry-Howard isomorphism, as normalisation corresponds to beta reduction $\lambda x.M(N) = M[x/N]$ (Prawitz 1971). In beta reduction the two functions are equal and equality in the lambda calculus is more fine-grained than mere extensional identity. It might be argued that given the Curry-Howard isomorphism the question of identity in proofs can be reduced to the question of identity of algorithms. And further it might be argued that these questions can be resolved by identity in the lambda calculus because it is more fine-grained than extensional alignment.

However, regardless of these arguments, this picture falls apart in the context of arbitrary transformations, which would only preserve identity if proofs were identified by their sets of assumptions and conclusion. Anyone who wished to support a claim like this would need to explain why proofs should be identified merely with the argument they make valid. When we were considering the plausibility of normalisation-preserving proof the criticism wasn't that the condition was too strict but rather that it identified arguably distinct proofs. This suggests that such a claim would be hard to defend.

Prawitz offers an alternative suggestion in later work. He proposes that we change what it is to be a proof to incorporate the need for transformations (Prawitz 2014, p. 273). On this picture what we would normally think of as a proof is incomplete. If you take an argument (that is a potential proof), then this is considered an argument skeleton and we cannot under Prawitz's proposal assess whether or not it is valid. To assess validity, we need to also have a set of transformations. Then if it is valid when only these transformations are used, we say the argument skeleton plus the set of transformations is valid. The trick here is to try and remove the question of why we can use the transformations. This is done by moving the transformations from a part of the definition of validity to a part of the proof itself. When it comes to the proof-rules as opposed to proofs themselves Prawitz says they are valid when there is a set of transformations[2] such that the one-step proof of the rule plus the set of transformations is valid.

It is worth noting that this definition of argument is odd. It certainly wouldn't coincide with what most people do when they write proofs, even with what most logicians do when they write formal proofs. But we might put two responses forward on Prawitz's behalf. The first is that most people couldn't formalise their proofs. Yet to the extent we think they are engaged in precise reasoning; we might say that they are still demonstrating the reasoning found in the formal proof. Similarly, while most people couldn't give the set of transformations needed to reduce their proofs perhaps they still demonstrate reasoning that contains these transformations. Secondly, Prawitz thinks intuitionistic logic is the proof-

[2]Prawitz's exact wording makes it sound like only one justification is allowed. But that cannot be correct as even the intuitionistic rules will require permutation and conversion for validity.

theoretically valid logic. But most work is done in classical logic. It is clear then that it is not an issue to depart from standard practice. So perhaps here too the suggestion is a revision rather than a part of current practice.

Still, the revision will not work, and the reason is simple. The elimination rules were supposed to follow from the meaning of the introduction rules. But now we have this extra part, the justification. In what sense do the transformations follow from the introduction rules? Prawitz offers no story here. And one does not seem possible. The elimination rules are valid because via proof-theoretic validity we can show that they are eliminable in favour of the introduction rules alone. But the transformations aren't eliminable. They simply don't follow from the meaning of the introduction rules. So, this move does not resolve the problem.

Dummett's answer is simple. He claims to do away with transformations and instead he asks only that another closed proof can be found, with the same conclusion, which is valid (Dummett 1991b, Ch. 11). This leads to the definition we gave initially. However, the talk of getting rid of transformations is in a sense however only terminological. This is because we can allow for Dummett's definition by simply being as liberal as we like in what counts as a transformation and then asking what is valid given all these transformations. And Dummett's does place one requirement on how we find the second proof: that the second proof is found in an effective manner.

Dummett's reason that we can use arbitrary transformations is that Dummett takes the finding of a second proof as evidence that a rule is eliminable. We can think of this in terms of meaning or content. A rule really can't have more content than a second rule if we can remove any occurrence of it in favour of the second rule, even if we remove it by finding another proof. Of course, it would now need to be shown that the whole procedure preserved content. And so, we are left with the same issue we were involved with earlier. Why does the move from one proof to another preserve content?

All that is to say, some issues remain to be resolved in explaining why proof-theoretic validity can be used as a tool to find what follows from the introduction rules that define the connectives. But recent results on what tautologies follow from proof-theoretic validity have raised a more serious challenge to the idea that it captures a notion of following from that should be unproblematic for the inferentialist. When proposing proof-theoretic validity Prawitz made the following conjecture (Prawitz 1971; Prawitz 2014, p. 270):

Conjecture 0.1 (Prawitz)**.** *Proof-theoretic validity aligns extensionally with the validities of intuitionistic logic.*

This conjecture was also thought to be exceedingly plausible by Dummett (Dummett 1991b, p. 270). However, a collection of formal results show that proof-theoretic validity actually aligns extensionally with some superintuitionisitic logic. The consequences of these results are taken up in Chapter 2. There it is shown that Harrop's rule, one of the rules that is proof-theoretically valid, isn't harmonious under the most popular views of harmony. This would seem to have the worrying consequence that proof-theoretic validity and harmony, despite intending to be complimentary explications of the same idea, are in conflict with one another.

Chapter 2 aims to show that the superintuitionistic validities are the result of the treatment of the atomic formulas and how this treatment corresponds to disjunction free sentences. This shows that the conflict about the meaning of the logical connectives as treated by proof-theoretic validity and harmony is only apparent. The extra information being extracted comes from the atomic formulas not the connectives.

Another question left open by the failure of Prawitz's conjecture is what logics do in fact align with the common definition of proof-theoretic validity. This question is answered in Chapter 3. There I prove a surprising connection between proof-theoretic validity and inquisitive logic. Inquisitive logic is the logic of inquisitive semantics which provides a uniform semantics for questions and answers (Ciardelli and Roelofsen 2011). I show that one variant of proof-theoretic validity is extensionally equivalent to generalised inquisitive logic and I offer a modified version of proof-theoretic validity which is extensionally equivalent to inquisitive logic. These two chapters taken together give a clearer idea of what proof-theoretic validity commits us too and what the consequences of their brand of 'following from' are. They also help to spell out the consequences of the recently discovered connection between proof-theoretic validity and superintuitionistic logics.

Chapter 1

The Potential in Frege's Theorem

Is a logicist bound to the claim that as a matter of analytic truth there is an actual infinity of objects? If Hume's Principle is analytic then in the standard setting the answer appears to be yes. Hodes's work pointed to a way out by offering a modal picture in which only a potential infinity was posited. However, this project was abandoned due to apparent failures of cross-world predication. We re-explore this idea and discover that in the setting of the potential infinite one can interpret first-order Peano arithmetic, but not second-order Peano arithmetic. We conclude that in order for the logicist to weaken the metaphysically loaded claim of necessary actual infinities, they must also weaken the mathematics they recover.

1.1 Introduction

1.1.1 Potentially Infinite Models

In the non-modal setting, Frege (1893; J. Burgess 1984; Boolos 1986; Heck 1993) essentially proved that second-order Peano arithmetic, PA^2, is interpretable in the theory HP^2, which consists of the Second-order Comprehension Schema and Hume's Principle:

$$\forall X, Y(\#X = \#Y \Leftrightarrow \exists \textit{ bijection } f : X \to Y). \tag{HP}$$

Hume's Principle characterises the cardinality operator #, read 'the number of' or 'octothorpe', as a type-lowering function that takes equinumerous second-order objects to the same first-order object. This definition can be motivated in the finite case by examples such as checking one has the same number of knives and forks by setting them out in pairs. Formally, Frege's result is:

Theorem 1.1 (Frege's Theorem)**.** *There is a translation from the language of* PA^2 *to the language of* HP^2 *that interprets* PA^2 *in* HP^2*.*

The formal definition of the theories mentioned here can be found in Appendix 1.8. Frege's Theorem has traditionally been regarded as philosophically important because it is supposed to show that we can derive all arithmetical theorems from an epistemically innocent system. This requires that Hume's Principle is analytic. However, on the usual semantics, Hume's Principle is only true on domains with at least a countable infinity of objects. This commits logicists like Frege to the analytic existence of an actual infinity of objects (Boolos 1998, pp. 199, 213, 233; Hale and Wright 2001a, pp. 20, 292, 309; Cook 2007, p. 7).

A commitment to a *potential* infinity, in contrast, isn't a commitment to how many things there actually are, just how many are possible. This is a much safer area in which to make analytic claims. Here we show that some but not all of the mathematics of the actual infinite is recoverable in the setting of the potential infinite. And so, to avoid problematic ontological commitments the logicist must also weaken the mathematics they recover.

To do this we must decide how to represent Hume's Principle. Below we will define 'the number of' operator # in a semantic manner. However, we are convinced that this is simply a convenience and we can think of our models as defining # as satisfying Hume's Principle with the additional criteria that this function is rigid across worlds. An axiomatization would consist of the following modification of Hume's Principle:

$$\Box\forall X, Y(\#X = \#Y \Leftrightarrow \exists\ bijection\ f : X \rightarrow Y),$$

plus a principle to rigidify the # operator. This would require working in a hybrid modal logic where worlds could be saved and recalled such as Williamson (2013, p. 370).[1] However, we leave the details of this approach for future work. As the modification is so minimal, the move to the potentially infinite doesn't undermine the justifications offered for Hume's Principle. The syntactic priority thesis can still be argued for as we can identify the behaviour of terms in a modal setting as well as in a non modal setting. Similarly if we think that abstraction principles offer implicit definitions then this justification works as well in the modal setting.

The rigidity of the octothorpe is important for the success of the project here. However, by assuming that it is rigid we are presuming that 'the number of' operator is rigid. Whether this is the case in natural language is an empirical question (e.g. Stanley 1997). We do not address this issue here, but two things are worth noting. First the question of the rigidity of 'the number of' is not the same question as e.g. whether the number of planets varies between worlds. This is because we do not apply the operator to predicates but rather to sets which do not vary their membership across worlds. The second is that this setting does rule out the possibility of multiple different number structures in the different worlds, e.g. the numbers being von Neumann ordinals in one world and Zermelo ordinals in another. This means that a certain kind of referential indeterminacy which has a prominent place in philosophy of mathematics cannot be addressed in this setting as we have presumed against it (Benacerraf 1965; Button and Walsh 2018, ch. 2).

[1]For those familiar with hybrid systems the axioms needed is $\uparrow \Box\forall\ X, y\ \downarrow [\#X = y \rightarrow \Box\#X = y]$. However, this will not play a role in what follows.

To set up our result, we define a set of second-order Kripke models, which we will call *potentially infinite models.* This idea comes from Hodes (1990b, p. 379), although he does not place exactly these constraints on the accessibility relation. We want the models to be nearly linear sequences of worlds (if there are two worlds neither of which accesses the other, there is a third world they both access), where later worlds are possible from the perspective of earlier worlds but not the other way around. Each of these worlds should contain only a finite number of objects as we are assuming actual infinities are impossible, and the number of objects should increase from one world to the next. Each world will have its own second-order domain, which as the worlds are finite, will be the full powerset. The octothorpe will implement Hume's Principle by taking sets of the same cardinality to a unique object and this object will not change from one world to the next. We define the models formally as follows:

Definition 1.1. *A* potentially infinite (PI) model *is a quadruple* $\mathcal{M} = \langle W, R, D, I \rangle$ *in the modal signature with second-order quantification and with* $\#$ *and* $\mathbf{a}$ *as the only non-logical symbols, such that the following conditions are met:*

1.1.1. W *is countably infinite and* R *is a directed partial order,*[2]

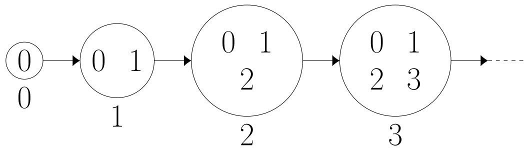
1.1.2. the first-order domain of w*, written* $D(w)$*, is non-empty and finite for all* $w \in W$*,*

1.1.3. for each $n \geq 1$*, the range of the second-order* n*-ary relational quantifiers at* w *is* $\mathscr{P}(D(w)^n)$ *consisting of all subsets of the* n*-th Cartesian power* $(D(w))^n$ *of* $D(w)$*,*

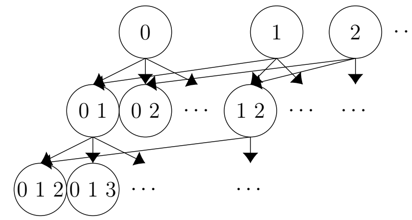
1.1.4. if $w, s \in W$ *such that* $R(w, s)$ *and* $w \neq s$*, then* $D(w) \subsetneq D(s)$*,*

1.1.5. the function $\mathbf{a} : \omega \to D$ *(where* D *is* $\bigcup_{w \in W} D(w)$*) assigns to each number* n *a distinct element* $\mathbf{a}_n$ *in one of the first-order domains, and for all* $w \in W$*, the cardinality of* X *is* n *if and only if* $\#X = \mathbf{a}_n$ *at* w*. More formally, for* $\#$ *and all* w *the interpretation function is defined as follows:* $I(\#, w) = \{\langle X, \mathbf{a}_{|X|} \rangle \mid \exists s \in W \; X \in \mathscr{P}(D(s))\}$.

[2] An order R is directed if for all $w, s \in W$ there exists an $t \in W$ such that $R(w, t)$ and $R(s, t)$.



(a) The minimal model



(b) The subset model

Figure 1.1: Examples of potentially infinite models

Remark 1.1. *Three brief remarks on this definition:*

First, conditions 1.1.1-4 define a PI model as a directed partial order of ever-increasing finite domains. This means that if we have several objects existing in different possible worlds we can always move to a world where they all exist.

Second, condition 1.1.5 defines the cardinality operator # using metatheoretic cardinality $|X|$. It is sufficient for Hume's Principle to hold that # picks-out cardinality, and so condition 1.1.5 ensures that all potentially infinite models are models of Hume's Principle. One reason we need $\mathscr{P}(D(w)^2)$ from 1.1.3 is because the quantifier over graphs of functions in Hume's Principle ranges over this set.

Third, condition 1.1.5 also ensures that the interpretation of the octothorpe is rigid. That is, the octothorpe is interpreted as the same relation at every world. Because of this nothing will be lost if we write $\#X = x$ and don't specify the world of evaluation. In fact, while we define $\#X$ using the $\mathbf{a}_i$'s, we could have instead simply defined it as rigid and satisfying Hume's Principle and this along with directedness would ensure the $\mathbf{a}_i$'s exist.

This definition can obscure the simplicity of the idea here, as such it helps to give several examples. The simplest potentially infinite model we can construct is the following:

Example 1.1. *The* minimal *potentially infinite model is $(\omega, \leq, D, I)$ where $D(\mathbf{n}) = \{\mathbf{0}, \ldots, \mathbf{n}\}$ and the interpretation function I interprets octothorpe as cardinality in the metalanguage.*[3]

[3] I will use bold face numbers for the numbers in the metalanguage.

That is, $I(\#, w)(X) = \mathbf{n}$ *if and only if* $|X| = \mathbf{n}$. *The minimal model is illustrated in Figure 1.1a. When working with such a model we see that a number can be missing from a world even if a set of that cardinality is present. So* $I(\#, \mathbf{1})(\{\mathbf{0}\}) = \mathbf{1}$ *and* $\mathbf{1} \in D(\mathbf{1})$, *but* $I(\#, \mathbf{1})(\{\mathbf{0}, \mathbf{1}\}) = \mathbf{2}$ *and* $\mathbf{2} \notin D(\mathbf{1})$ *even though* $\{\mathbf{0}, \mathbf{1}\} \subseteq D(\mathbf{1})$.

A less simple but similarly elementary model makes use of the non-empty finite subsets of the natural numbers. This model helps illustrate a non-linear R relation:

Example 1.2. *Let the* subset *model be* $(\mathscr{P}(\omega)^{<\omega} - \{\varnothing\}, \subseteq, D, I)$ *where* $D(X) = X$ *and again the octothorpe is cardinality. The subset model is illustrated in Figure 1.1b. Note that if we have worlds* $X_0, \ldots, X_n$ *we can always find an accessible world whose domain is* $\bigcup_{i=0}^{n} X_i$. *For example,* $\{\mathbf{0}, \mathbf{1}\}, \{\mathbf{3}\}, \{\mathbf{100}, \ldots, \mathbf{200}\}$ *are all finite subsets of the natural numbers, none of which access each other, however, their union* $\{\mathbf{0}, \mathbf{1}, \mathbf{3}, \mathbf{100}, \ldots, \mathbf{200}\}$ *is also a world, which they all access.*

It is easy to generate unintended models from these two cases. Using the minimal model, for example, we can define the **3-0** swap model:

Example 1.3. *The* **3-0** swap *model takes* ***0*** *and* ***3*** *in the domain of the minimal model and switches them around. So* $D(\mathbf{0}) = \{\mathbf{3}\}$, $D(\mathbf{1}) = \{\mathbf{3}, \mathbf{1}\}$, $D(\mathbf{2}) = \{\mathbf{3}, \mathbf{1}, \mathbf{2}\}$, $D(\mathbf{3}) = \{\mathbf{3}, \mathbf{1}, \mathbf{2}, \mathbf{0}\}$ *and then for all* $\mathbf{n} \geq \mathbf{3}$, *we have that* $D(\mathbf{n})$ *exactly as it is in the minimal model.*

These models should help illustrate the intuition behind the potentially infinite models. They will also be helpful when we need counterexamples to claims later in the chapter.

We can now define satisfaction for potentially infinite models using a standard semantics for quantified modal logic, such as in Fitting and Mendelsohn (1998). Three things to note first: (1) Our quantifiers are actualist, but free variables may be assigned to objects in any world. (2) Set variables are interpreted rigidly across worlds. That is the membership of a set doesn't change depending on the world. (3) To simplify the notation, instead of variable assignments, we work as though we had a rigid name for every object in the models. Recall that $\mathcal{M}, w \vDash \varphi$ means that given any replacement of free variables with the added constants we evaluate φ as true in $\mathcal{M}$ at world w. With this in place, the notion of potentially infinite models induces a natural validity relation, which we define as follows:

Definition 1.2. *We say that φ is true in all potentially infinite models, or $\vDash_{\mathsf{PI}} \varphi$, if for all potentially infinite models $\mathcal{M}$ and worlds $w \in W$ we have $\mathcal{M}, w \vDash \varphi$. We define $\varphi \vDash_{\mathsf{PI}} \psi$ as for all models $\mathcal{M}$ and worlds $w \in W$, if $\mathcal{M}, w \vDash \varphi$ then $\mathcal{M}, w \vDash \psi$.*

The consequence relation here is defined locally rather than globally (Fitting and Mendelsohn 1998, p. 21). This is because the deduction theorem holds for the local consequence relation but not the global one (Fitting and Mendelsohn 1998, p. 23).

1.1.2 Main Results

We will now state our two main results which together show that we can interpret the first-order theories of first-order Peano arithmetic PA^1 and first-order true arithmetic TA^1, but not the second-order theories of second-order Peano arithmetic PA^2 and second-order true arithmetic TA^2, in theories defined in terms of potentially infinite models. A deductive theory for second-order modal logic with rigid operators would be unwieldy and the complications caused by it would be likely to obscure the insights provided by the Kripke semantics. Hence, we leave development of a deductive theory for future work. We can define a theory just in terms of the potentially infinite models. This theory will be stronger than anything we could produce deductively because it does not admit nonstandard models of the natural numbers. Because of this we will call it the external theory of the potentially infinite or E_{PI}:

$$\mathsf{E}_{\mathsf{PI}} = \{\varphi \mid \vDash_{\mathsf{PI}} \varphi\}. \tag{1.1}$$

To capture something closer to what can be deduced from the models we need to use the model-theoretic validity relation defined above, relativised to a weak metatheory. The theory ACA_0 is a subsystem of PA^2 which only has comprehension for first-order formulas. More information about this theory can be found in Appendix 1.8. Since we can code finite sets of natural numbers as natural numbers in ACA_0, we can define the property of being a potentially infinite model in this theory, along with the associated validity notion $\vDash_{\mathsf{PI}}$. This gives us the internal theory of the potentially infinite or I_{PI}:

$$\mathsf{I}_{\mathsf{PI}} = \{\varphi \mid \mathsf{ACA}_0 \vdash `\vDash_{\mathsf{PI}} \varphi'\}. \tag{1.2}$$

Intuitively, this theory is every formula that can be proven valid on potentially infinite models, given the weakest metatheory that can formalise the models. A full definition is given in Appendix 1.9.[4] The definition of interpretation is traditionally restricted to theories in the same logic, whereas in this setting E_{PI} and I_{PI} are theories in second-order modal logic but PA^1, PA^2, TA^1, and TA^2 aren't modal theories. So, to state and prove our main results we need a more general notion of *generalised translation* and *interpretation* which captures those interpretations which involve not just different theories but different logics. This is defined in section 1.5. Our first main result is:

Theorem 1.2. *(i) There is a generalised translation from the language of* PA^1 *to the second-order modal language with octothorpe that interprets* TA^1 *in* E_{PI}.

(ii) There is a generalised translation from the language of PA^1 *to the second-order modal language with octothorpe that interprets* PA^1 *in* I_{PI}. *Further, this is a* PA^1*-verifiable generalised interpretation.*

[4] We picked the weakest theory because we are interested in what is deducible from PI models and if we strengthen the metatheory I_{PI} will be strengthened in ways that reflect what the metatheory thinks about finite sets (which can code consistency statements).

This result is proven in Section 1.5. The translation used is based on one offered by Linnebo (2013) in the setting of modal set theory. The key difference, compared with the standard notion of translation, is that "for all" is translated as "necessarily for all" and, similarly, "there is" is translated as "possibly there is."

The first theorem shows that the PI models capture a significant amount of mathematics. However, we cannot strengthen the result to second-order theories of arithmetic as our second main theorem shows:

Theorem 1.3. *(i) There is no generalised translation from the language of* PA^2 *to the second-order modal language with octothorpe that interprets* TA^2 *in* E_{PI}*.*

(ii) There is no generalised translation from the language of PA^2 *to the second-order modal language with octothorpe that* PA^2*-verifiably interprets* PA^2 *in* I_{PI}*.*

For both E_{PI} and I_{PI}, the results follow from the fact that PI models are Π^1_1 definable. And this follows because all of the worlds are finite. Because of this, PI models are representable in reasonably weak theories of second-order arithmetic. But then limitive results about what theories can represent about themselves will stop theories that can represent E_{PI} and I_{PI} being interpretable into E_{PI} and I_{PI}.

These results are important because they show that less mathematics is analytic on the philosophical perspective which motivates the potentially infinite models than on the traditional perspective. The external theory cannot recover TA^2 but only TA^1. And the internal theory cannot recover PA^2 but only PA^1. Further, PA^2 has traditionally been the target of Fregean interpretation results as it allows for the recovery of analysis and much of mathematics.[5] Analysis can be coded in second-order Peano arithmetic, as real numbers can be coded as sets of rationals, which in turn can be coded as naturals. This means that Frege's theorem already accounts for a larger expanse of mathematics than it might first appear. If we try to

[5] Demopoulos (1994, 238 n26) points out that Frege often uses arithmetic when he means something broader including analysis.

avoid the claim that it is analytic that there are actually infinitely many objects, however, it then seems we will not have managed to recover as much mathematics. If we are looking to show that mathematics is analytic, we have moved further from our goal.

However, we have still captured a substantial chunk of our most frequently used mathematics. Feferman (2005, p. 613) has argued that all scientifically applicable analysis can be developed in PA^1 or a conservative extension of it.[6] If this is correct then we can still recover the mathematics for which an explication of its truth is most philosophically fruitful, namely the mathematics which we rely on when we act in the world. One might wonder why a logicist would care about whether or not the mathematics recovered is used. But it seems we should keep an open mind to different parts of mathematics being justified in different ways. Maybe something as fundamental as first-order arithmetic turns out to be analytic, but it seems unlikely that the same is true of the higher reaches of set theory. With this in mind, it should not be damaging that not all mathematics turns out to be analytic.

1.1.3 A Diversity of Modal Logicisms

The idea of using the potentially infinite as a foundation of logicism has a pedigree in the work of Putnam and Hodes, and more recent work on modal foundations of mathematics and on variants of Frege's theorem in different logics. Putnam suggested that by accepting a modal picture of mathematics we could avoid being Platonists about the numbers or committing to how many objects there actually are. This is stated most clearly when he writes:

[6] For example, "By the fact of the proof-theoretical reduction of W to [PA^1], the only ontology it commits one to is that which justifies acceptance of [PA^1]." (Feferman 2005, p. 613) Feferman works in a system W which contains types for the naturals, the cross product and partial functions. The full classical analysis of continuous functions can be carried out in W. (Feferman 2005, p. 611)

> 'Numbers exist'; but all this comes to, for mathematics anyway, is that (I) ω-sequences are possible (mathematically speaking); and (2) there are necessary truths of the form 'if α is an ω-sequence, then...'[.] (Putnam 1967a, pp. 11–12)

Hodes took on this idea, but he was sceptical of the existence of actual infinities. He thought that '[a]rithmetic should be able to face boldly the dreadful chance that in the actual world there are only finitely many objects' (Hodes 1984, p. 148). His solution made use of the idea of the potentially infinite rather than the actually infinite. He appealed to modality and in particular the modality that seems to be implicit in our concept of number: the idea that it is always possible to add 1 (Hodes 1990b, p. 378).

However, by 1990, Hodes concluded that the reduction of mathematics to higher-order modal logic had failed. Hodes describes the problem as follows:

> The problem is simple: relative to [a model of Hume's Principle] for a type-0 variable v, $\Diamond(\exists v)(\underline{N}(v)\& \ldots)$ "moves us" to other worlds u and then has us seek a witnessing member of [the natural number in the model] in [the domain of u]; we may find one, but then have no way "back" to w to see what hold [sic] for it there. (Hodes 1990b, p. 388)

So we might know that there possibly exists a number with a property, but in Hodes's system, we have no way of returning to our original world to use what we have found. For example, if we find the number of a set in some world, we have no assurance that this number is available for us to talk about in the world the set came from. It is only known that it is the number of the set *in the world the number exists in.* The difficulty identified here is with cross-world predication, which occurs when we want to say something about an object in one world and how it relates to objects in another world (Kocurek 2016).

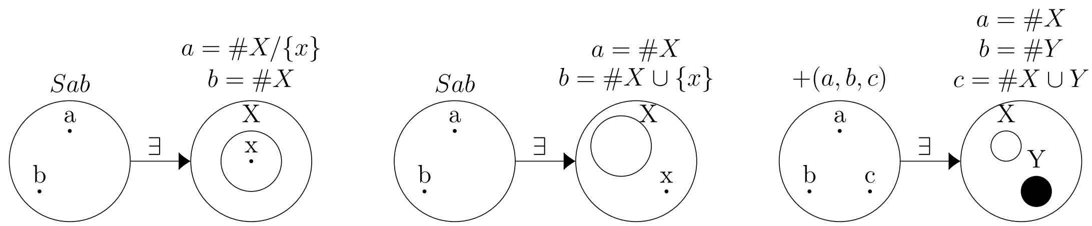
In what follows we will show that the problem is not with cross-world predication *per se*. Both by working directly with the models, but also by allowing the octothorpe to be rigid, we can mimic some of the effects of cross-world predication. Yet in this setting we recover some but not all of the arithmetic recovered by Frege's theorem. Indeed, our main results, Theorems 1.2 and 1.3, show that the situation is more complicated than Hodes suggested, and that a partial realisation of his project is possible.

There are two recent trends in the study of logicism which this project is connected to. First, Studd (2016) has suggested that the modal setting is an attractive one for the logicist because it would help to solve the bad company objections. Unlike here, Studd's is concerned with inconsistent abstraction principles and in particular set abstraction. This is interestingly connected to the naïve conception of set because one can think of the unrestricted set Comprehension Schema as similar in spirit to a modal version of Basic Law V. While work in this area goes back to Parsons (1983), it has been pursued recently by Linnebo (2013; 2018). Much of Linnebo's work has been on set theory. The concerns there are very different from ours, as it make little sense in set theory to worry about the actual infinite not existing and set theory is generally treated in first-order logic.The work in this chapter takes inspiration from the results presented in Linnebo (2013) and (2018) and makes use of a similar method of translating between the modal and non-modal setting. However, while the dynamic abstraction principles discussed by Linnebo (2018) resemble the behaviour of the *number of* operator, his preferred abstraction principle for arithmetic is ordinal abstraction (Linnebo 2018, Ch. 10.5), whereas in this chapter we work with a modal version of Hume's Principle, a cardinality principle. There is further work showing that restricting comprehension is successful in making Basic Law V consistent (Wehmeier 1999; Wehmeier 2004; Walsh 2012; Walsh 2016).

Second, there has been a lot of recent work on whether Frege's Theorem still holds when the logic is modified in certain ways. Bell (1999) and Shapiro and Linnebo (2015) have shown that Frege's Theorem is available in the intuitionistic setting. J. P. Burgess (2005) and Walsh (2016) found that a version of Frege's Theorem is possible in a certain predicative setting. Kim (2015) proves a version of Frege's Theorem in a modal setting. This employs an axiomatised version of the 'the number of F's is n' as a binary relation, instead of the traditional type-lowering 'number of' operator. Kim recovers the axioms of PA but finds that a restricted version of HP^2 holds. The modality used is S5 and meant to represent logical possibility, not potentiality. Because of this Kim's system does not have the same structure of our models, where the numbers slowly grow. Closer in spirit to the work here is that on finite models of arithmetic by Mostowski (2001). There he considers initial sequences of the natural numbers and what holds over all such models. These have a clear connection to the minimal model discussed above. Urbaniak (2016) has taken Mostowski's models and worked with them in a modal setting. They have shown that Leśniewski's typed, free logic with modal quantifiers, which proves a predicative version of HP^2, can interpret PA^2. Our setting is quite different from that of Urbaniak's paper as Leśniewski's typed, free logic differs dramatically from the one we work in here. The work in this chapter proceeds by looking at whether a version of Frege's Theorem is available in a classical second-order modal setting. Unlike these other results, we find that a modal version of Frege's Theorem for PA^2 is *not* possible, as shown by Theorem 1.3.

1.1.4 Outline of chapter

This chapter is organised as follows. Section 1.2 expands the potentially infinite models' language to include the language of arithmetic. In Section 1.3 we show that using the expanded language the potentially infinite models satisfy a weak theory of arithmetic equivalent to a modal version of Robinson's Q. In Section 1.4 we define the inductive formulas of the language and show that induction holds for them. This allows us to show Theorem 1.2, that TA^1 is interpretable in our external theory and PA^1 is interpretable in our internal theory,



(a) Diagram of when a is succeeded by b.

(b) Diagram of when a is succeeded by b for alternative definition of S.

(c) Diagram of when c is the addition of b and c.

Figure 1.2: Diagrams demonstrating the definition of successor and addition

in Section 1.5. In Section 1.6 we show that no natural interpretation of PA^2 is possible by proving Theorem 1.3.

1.2 Definitions for a Modal Grundlagen

Just as Frege in the *Grundlagen* defined the numbers and the relations on them using only the 'number of' operator, here we show how modified versions of Frege's definitions can do this in the setting of the potentially infinite.[7] Proving that these definitions satisfy the usual arithmetical axioms will occupy us in §§1.3–1.4. In this section we simply set out the definitions themselves and say a word about their motivation. While entirely rigorous, it is our hope that, as in the *Grundlagen*, the definitions will be intuitive and correspond to our understanding of cardinal numbers.

The first definition is easy and does not require any of the modal apparatus. We simply let $0 = \#\varnothing$. This follows Frege (1884, §74 p. 87) explicitly, who said that zero is "the Number which belongs to the concept 'not identical with itself'". Next we must define the successor,

[7] This has some precedent in Hodes (1990, p. 383). However, whereas we (and Frege) first define successor and then use this to build the other definitions, Hodes takes 'less than or equal to' as his primitive. In his system a number N (understood as a higher-order object) is less than or equal to another number N' just in case it is possible that there are two other second-order objects A and A' each with the same number of objects as N and N' respectively and A is a subset of A'. That this has parallels with the definition of successor offered here will be clear on inspection.

as the other definitions rely on it. The definition here is like the one offered by Frege, but it differs by allowing the sets which witness that one object is the successor of another to be merely possible. This is to ensure that if an object is ever the successor of another, then it is the successor of that object in every world where they both exist. This property will be important in the proof of induction. The definition of successor, in plain terms, is: one object is the successor of another just in case it is possible that there are two sets, which differ by one object and the successor is the number of the larger set, and the predecessor is the number of the smaller set. Figures 1.2a and 1.2b illustrate the two ways this can be done, resulting in two definitions of the successor:

Definition 1.3.

$$Sxy \equiv \Diamond \exists G, u[Gu \wedge (y = \#G) \wedge (x = \#(G - \{u\}))] \tag{1.3}$$

$$S'xy \equiv \Diamond \exists F, u[\neg Fu \wedge (x = \#F) \wedge (y = \#(F \cup \{u\}))] \tag{1.4}$$

The first of these definitions simply adds the possibility operator to the definition of successor suggested by Frege (1884, §76 p. 89). These definitions are equivalent: to see this, simply consider $F = G - \{u\}$ and $G = F \cup \{u\}$.[8] In what follows we will simply use the definition that is most convenient and will write S for both.

The definition of addition is similarly intuitive. The relation $+$ holds between three objects a, b, and c such that it is possible that there are disjoint sets X and Y of cardinality a and b respectively, and c is the cardinality of $X \cup Y$, the union of the two disjoint sets. This is illustrated by Figure 1.2c and can be written formally as:

[8] For easy of readability, we will use set theoretic notation as a convenient short hand for concepts formed using the language of the model. So $F \cup \{u\}$ is used for the concept given by $Xx \leftrightarrow (Fx \vee x = u)$.

Definition 1.4.

$$+(a, b, c) \equiv \Diamond \exists X, Y(a = \#X \wedge b = \#Y \wedge c = \#X \cup Y \wedge (X \cap Y) = \varnothing) \tag{1.5}$$

For c to be the result of multiplying a and b we need a set B of cardinality b and for each element x of B a set A_x of cardinality a. The A_x's must all be disjoint. And c must be the cardinality of the union of all the A_x's. To define the A_x's we define a binary relation P that holds between x in B and all y in A_x. So A_x is $\{y \mid Pxy\}$.

Definition 1.5.

$$\begin{aligned}\times(a, b, c) \equiv \Diamond \exists X, P[\#X = b \wedge \forall x \in X(\#\{y \mid Pxy\} = a)\\ \wedge \forall x, y \in X(x \neq y \rightarrow \{z \mid Pxz\} \cap \{z \mid Pyz\} = \varnothing) \wedge \# \bigcup_{x \in X} \{y \mid Pxy\} = c]\end{aligned} \tag{1.6}$$

The definition of the natural numbers is more complicated and require us to define the notion that one number follows another in the ordering of the natural numbers. We will make use of Frege's definition from the 1879 *Begriffsschrift* (1967, §III pp. 55 ff; 1884, §79 p. 92 ff). Russell and Whitehead (1910, p. 316) called this relation the *ancestral relation* because a good example of what it does is define the relation 'ancestor of' from the relation 'parent of'. The *strong ancestral* of φ holds between two objects a and b just in case b is contained in every set such that the set is closed under φ and the set contains everything a bears φ to. So, we can define someone's ancestors as everyone who is in every set that contains their parents and the parents of everyone in the set. It is not guaranteed that a bears this relation to itself, and so we also define the reflexive *weak ancestral.*

Definition 1.6 (The strong ancestral)**.**

$$\varphi^{+}(a, b) \equiv \forall X[(\forall x, y(Xx \wedge \varphi(x, y) \rightarrow Xy) \wedge \forall x(\varphi(a, x) \rightarrow Xx)) \rightarrow Xb].$$

Definition 1.7 (The weak ancestral).

$$\varphi^{+=}(a,b) \equiv \varphi^{+}(a,b) \vee a = b.$$

Using this definition, we define a natural number as an object that is some finite number of successor steps from 0, assuming 0 exists.

Definition 1.8 (Natural Number).

$$\mathbb{N}x \equiv S^{+=}0x \wedge \exists y(y = 0).$$

This definition closely parallels Frege's, though the definition of S is different. The existence claim is added because in the modal setting 0's existence cannot be assumed. For example, 0 does not exist at worlds **0**, **1**, and **2** in the **0-3** swap model, and, as **0** is not a member of infinitely many finite subsets of the natural numbers, 0 does not exist at infinitely many worlds in the subset model. In these worlds nothing is a natural number.

1.2.1 Some useful results

The following six lemmas will help explain the behaviour of $\mathbb{N}$ in the models. We admit the proofs as they do not pose any particular difficulty. For the following Lemmas, recall Definition 1.2 where $\vDash_{\mathsf{PI}} \varphi$ was defined as φ is true in all worlds in all potentially infinite models. First, note that the set defined by $\mathbb{N}$ at a world satisfies the antecedent of $S^{+}0x$. Intuitively, the idea here is that if x is in every set containing 0 and closed under S, and Sxy, or $S0y$, then y must also be in every set with these properties.

Lemma 1.1. $\vDash_{\mathsf{PI}} \exists x(x = 0) \rightarrow \forall y(S0y \rightarrow \mathbb{N}y))$

Lemma 1.2. $\vDash_{\mathsf{PI}} \forall x, y(\mathbb{N}x \wedge Sxy \rightarrow \mathbb{N}y)$

It follows immediately from this that if x exists at a world and at that world $\mathbb{N}y$ and Syx then $\mathbb{N}x$. However, that doesn't mean $\mathbb{N}$ is the set of all numbers across all worlds as $\mathbb{N}$ only holds of objects which exist at the world of evaluation. This contrasts with our other definitions where the objects need not exist at the world.

Lemma 1.3. $\models_{PI} \mathbb{N}x \to \exists y\, y = x$

This is because the quantifiers in $\mathbb{N}$ are plain rather than having modals in front of them. This is important because if we put the modals in front everything is a number!

We informally extend our definition of the interpretation function I to $I(\mathbb{N}, s) = \{x \in D(s) \mid \mathcal{M}, s \models \mathbb{N}x\}$. Note that by Lemma 1.3 we have $\{x \in D(s) \mid \mathcal{M}, s \models \mathbb{N}x\} = \{x \in D \mid \mathcal{M}, s \models \mathbb{N}x\}$, where D is the domain of the model not the world.

Recall that $\mathbf{a}_i$ is the unique element in D such that if $|X| = i$ then $I(\#, w)(X) = \mathbf{a}_i$ as defined in 1.1.5. We can now explicitly describe the interpretation of $\mathbb{N}$ at a world w in terms of the $\mathbf{a}_i$'s, that is, the set $I(\mathbb{N}, w)$:

Lemma 1.4. *Let w be a world and let n be the first number such that $\mathbf{a}_n \notin D(w)$. Then if $n > 0$, it follows that $\{0, \mathbf{a}_1, \ldots, \mathbf{a}_{n-1}\} = I(\mathbb{N}, w)$, and further, $n = 0$ iff $I(\mathbb{N}, w) = \varnothing$.*

This result shows us how the differences between our modal setting and the traditional non-modal setting of the *Grundlagen* become most stark in the case of the interpretation of the natural numbers at a world. Two things are worth highlighting. The first is that $\mathbb{N}$ is finite at every world, since it is a subset of the domain of the world, and the domain of every world is finite. The second is that objects that are not in $\mathbb{N}$ at one world can 'become' numbers at later worlds. This doesn't happen in the minimal model, where $I(\mathbb{N}, \mathbf{n}) = D(\mathbf{n})$ at every world. But it does in the subset model. For example, $I(\mathbb{N}, \{\mathbf{2}, \mathbf{100}\}) = \varnothing$, $I(\mathbb{N}, \{\mathbf{0}, \mathbf{1}, \mathbf{3}\}) = \{\mathbf{0}, \mathbf{1}\}$ and $I(\mathbb{N}, \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{100}\}) = \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}\}$. This distinguishes $\neg\mathbb{N}(x)$ from the other relations which have a certain stability; if objects stand in these relations at one world, then they do so in all worlds in which they all exist. The formal definition of stability is given as Definition 1.8. This difference is caused by there being no possibility operator at the beginning of the definition of $\mathbb{N}$. Despite this, once something is a number it remains one:

Lemma 1.5. *$\vDash_{\mathsf{PI}} S(x,y) \to \Box S(x,y)$ holds, as does $\vDash_{\mathsf{PI}} S^{+}(x,y) \to \Box S^{+}(x,y)$, $\vDash_{\mathsf{PI}} S^{+=}(x,y) \to \Box S^{+=}(x,y)$ and $\vDash_{\mathsf{PI}} \mathbb{N}x \to \Box \mathbb{N}x$.*

It is also worth noting that even though some cardinalities may not be numbers at ever world, the cardinality of every set eventually becomes a natural number.

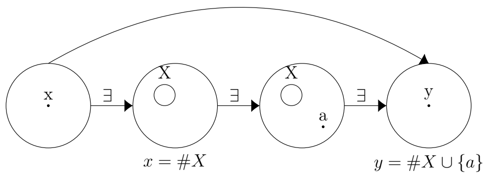
Lemma 1.6. *For all $w \in W$ and $X \subseteq D(w)$, there is a world s such that $R(w,s)$ and $\#X \in I(\mathbb{N}, s)$.*

This is because # is a function, first-order converse Barcan holds, and the accessibility relation is directed. With these preliminary results we can now show our definitions satisfy a simple theory of arithmetic.

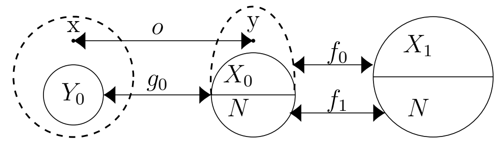
1.3 Proving Modalized Robinson's Q

In what follows we will prove that the modalized axioms of Robinson's Q are true on all PI models (cf. Definition 1.2). Robinson's Q is a weak theory of arithmetic that defines successor as an injective function that never returns 0 and gives a recursive definition of addition and multiplication. By "modalized" we mean that we write "necessarily for all" for "for all" and "possibly there is" for "there is". In other words, it is what results when we apply the Linnebo translation (mentioned in the introduction) to the axioms of Robinson's Q. The theory PA^1 is obtained by adding the mathematical induction schema to Q. We deal with PA^1 and the proof of the induction schema in Section 1.4.[9]

[9] A list of the non-modalized axioms can be found in Appendix 1.8. While what we show here is that these axioms are in the theory E_{PI}, each of the proofs that follow can be formalised in ACA_0 (cf. Appendix 1.9). That this is possible will ensures that all axioms proven here are also in the theory I_{PI} (from Section 1.1.2). This is a key point in the proof of Theorem 1.2.ii which we complete in section 1.5.



(a) Finding the successor of x



(b) Proof of the recursion clause for addition

Figure 1.3: Diagram for the behaviour of successor and addition

First we will show that our relations define the graphs of functions. The easiest case is successor.

Lemma 1.7 (S1)**.** $\models_{\mathsf{PI}} \Box\forall x, y, z \in \mathbb{N}((Sxy \wedge Sxz) \rightarrow y = z)$.

Proof. Let $s \in W$ and $x, y, z \in I(\mathbb{N}, s)$ satisfy the antecedent. As x is the predecessor in both relations it follows by directedness that there is a $w \in W$, such that $R(s, w)$ where there are $X, X' \subseteq D(w)$ and $\#X = x = \#X'$. As such there is a bijection $g : X \rightarrow X'$. There will also be $a, b \in D(w)$ such that $a \notin X$, $b \notin X'$, and $y = \#X \cup \{a\}$ and $z = \#X' \cup \{b\}$. As $a \notin X$ and $b \notin X'$ we can construct h such that for all $u \in X$, $h(u) = g(u)$ and $h(a) = b$. Clearly h is a bijection, so $y = \#X \cup \{a\} = \#X' \cup \{b\} = z$. □

Lemma 1.8 (S2)**.** $\models_{\mathsf{PI}} \Box\forall x \in \mathbb{N}\Diamond\exists y \in \mathbb{N}\ Sxy$.

Proof. As illustrated in Figure 1.3a, let $s \in W$ and $x \in I(\mathbb{N}, s)$, it follows that $x = \mathbf{a}_n$ for some n and, by Lemma 1.4, $\{0, \dots \mathbf{a}_{n-1}\} \subsetneq D(s)$. Further, $\mathbf{a}_n = \#\{0, \dots \mathbf{a}_{n-1}\}$ and $\mathbf{a}_n \notin \{0, \dots \mathbf{a}_{n-1}\}$. Thus, there must be a further world w accessible from w_1 and a $y \in D(w)$ such that $y = \#\{0, \dots \mathbf{a}_{n-1}\} \cup \{\mathbf{a}_n\}$. It follows that Sxy at w. By Lemma 1.5 $x \in I(\mathbb{N}, w)$. As $\mathbb{N}$ is closed under successor by Lemma 1.2, we have that $y \in I(\mathbb{N}, w)$. And since R is transitive, w is accessible from s. □

These two proofs offer a general outline of the reasoning for addition and multiplication. For S1 this strategy is to show that whatever x is the sets assigned to y and z will have the same cardinality. Where as for S2 one simply needs to construct a set of the correct cardinality. For this reason we do not give the proofs for the next four lemmas.

Lemma 1.9 (A1). $\vDash_{\mathsf{Pl}} \Box\forall x, y, z, z' \in \mathbb{N}(+(x, y, z) \wedge +(x, y, z') \to z = z')$.

Lemma 1.10 (A2). $\vDash_{\mathsf{Pl}} \Box\forall x, y \in \mathbb{N}\Diamond\exists z \in \mathbb{N}\ +(x, y, z)$.

Lemma 1.11 (M1). $\vDash_{\mathsf{Pl}} \Box\forall x, y, z, z' \in \mathbb{N}(\times(x, y, z) \wedge \times(x, y, z') \to z = z')$.

Lemma 1.12 (M2). $\vDash_{\mathsf{Pl}} \Box\forall x, y \in \mathbb{N}\Diamond\exists z \in \mathbb{N}\ \times(x, y, z)$.

We also need to show that 0 meets the right conditions to be a constant.

Lemma 1.13 (Z1). $\vDash_{\mathsf{Pl}} \Diamond\exists x \in \mathbb{N}(x = 0 \wedge \Box\forall y(y = 0 \to y = x))$.

Proof. By the definition of $\mathbb{N}$, it follows that $0 \in I(\mathbb{N}, s)$ for any world s with 0 in the domain. And as $0 = \#\varnothing$ there is some s with 0 in the domain. The second conjunct follows by the transitivity of identity. □

We can now move on to the recursion equations in Q. We separate these into the base steps concerning 0 and the recursive step. For the base steps, because $0 = \#\varnothing$ the proofs of the lemmas are relatively straight forward. As such we list them here without proof.

Lemma 1.14 (Q1). $\vDash_{\mathsf{Pl}} \neg\Diamond\exists x \in \mathbb{N}(Sx0)$.

Lemma 1.15 (Q3). $\vDash_{\mathsf{Pl}} \Box\forall x \in \mathbb{N}\ +(x, 0, x)$.

Lemma 1.16 (Q5). $\vDash_{\mathsf{Pl}} \Box\forall x \in \mathbb{N}\ \times(x, 0, 0)$.

What is left now is to show the recursion steps. He we only prove the case for + as one can use the same stratagy for × and the proof is simple for S.

Lemma 1.17 (Q2). $\vDash_{\mathsf{Pl}} \Box\forall x, y, z \in \mathbb{N}((Sxz \wedge Syz) \to x = y)$.

The proof simply follows from the fact that if there is a bijection between two sets X and Y then there will be a bijection between $X \cup \{a\}$ and $Y \cup \{b\}$ if a and b aren't in X or Y respectively.

Lemma 1.18 (Q4).

$$\vDash_{\mathsf{PI}} \Box\forall n, x_0, x_1, y_0, y_1, z \in \mathbb{N}(S(x_0, x_1) \wedge S(y_0, y_1) \wedge +(n, x_0, y_0) \wedge +(n, x_1, z) \to y_1 = z).$$

Proof. As illustrated in Figure 1.3b, let $s \in W$ and $n, x_0, x_1, y_0, y_1, z \in I(\mathbb{N}, s)$ satisfy the antecedent. We want to show that $y_1 = z$. By directedness, we know there is a world w containing all the objects and sets which the antecedent states possibly exist. As y_1 succeeds y_0 there is a set Y_0 and an object $a \notin Y_0$ at w such that $y_1 = \#Y_0 \cup \{a\}$ and $y_0 = \#Y_0$. We know y_0 to be the addition of n and x_0 so there are disjoint sets N and X_0 such that $n = \#N$, $x_0 = \#X_0$, and $y_0 = \#N \cup X_0$. Further there is a bijection $g_0 : Y_0 \to N \cup X_0$. Now let b be an element not in N or X_0 (we can always pick w so that such an element exists). Clearly we can define a bijection o between the singletons of a and b. Now, using g_0 and o, define the bijection $g : Y_0 \cup \{a\} \to N \cup X_0 \cup \{b\}$, as the union of g_0 and o. Now as x_1 is the successor of x_0, it follows that $x_1 = \#X_0 \cup \{b\}$. As z is the addition of n and x_1 there are disjoint sets N' and X_1 such that $n = \#N = \#N'$, $x_1 = \#X_0 \cup \{b\} = \#X_1$ and $z = \#N \cup X_1$. As such there are bijections $f_0 : X_0 \cup \{b\} \to X_1$ and $f_1 : N \to N'$. So, we can define the bijection $f : N \cup X_0 \cup \{b\} \to N' \cup X_1$ as f_0 on $X_0 \cup \{b\}$ and f_1 on N. Then as $z = \#N \cup X_1$ the composition $f \circ g$ is a bijection proving $y_1 = z$. □

Lemma 1.19 (Q6). $\vDash_{\mathsf{PI}} \Box\forall n, x_0, x_1, y_0, y_1, z \in \mathbb{N}(S(x_0, x_1) \wedge +(n, y_0, y_1) \wedge \times(n, x_0, y_0) \wedge \times(n, x_1, z) \to y_1 = z)$.

This proof is similar to the above except we end up showing that $y_1 = \#\bigcup_{x \in A_0 \cup \{u\}}\{y \mid Pxy \vee (x = u \wedge y \in N)\} = \#\bigcup_{x \in A_1}\{y \mid Txy\} = z$ where A_0, A_1, and N are of cardinality x_0, x_1, and n respectively and P is the relation given by $\times(n, x_0, y_0)$ and T by $\times(n, x_1, z)$.

These results show that we have successfully defined a modalized version of Robinson's Q in our system. The next section will recover a modalized induction schema.

1.4 Proving the Modalized Induction Schema

We have succeeded in giving a weak theory of arithmetic in a potentially infinite setting. However, we can recover more arithmetic by proving that when restricted to appropriate formulas a modalized version of the induction schema is true on all PI models. The modalized induction schema is:

$$[\varphi(0) \wedge \Box\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x,y) \rightarrow \varphi(y))] \rightarrow \Box\forall x \in \mathbb{N}\ \varphi(x) \tag{1.7}$$

Modalized induction does not hold for all formulas in our models, as will be shown in Lemma 1.22. So, we need to define a subclass of the formulas in the language of potentially infinite models for which it does hold. These we will call the inductive formulas, and in Lemma 1.21 it will be proven that induction does hold for inductive formulas.[10]

Definition 1.9. *The* inductive terms *and* formulas *are defined recursively as follows:*

1. *An inductive term is either 0 or a first-order variable.*
2. *If t_0, t_1, t_2 are inductive terms then $t_0 = t_1$, $S(t_0, t_1)$, $+(t_0, t_1, t_2)$ and $\times(t_0, t_1, t_2)$ are inductive formulas.*
3. *Applications of the propositional connectives to inductive formulas are inductive formulas.*
4. *If φ is an inductive formula then $\Box\forall x \in \mathbb{N}\ \varphi$ and $\Diamond\exists x \in \mathbb{N}\ \varphi$ are inductive formulas.*

The inductive terms and formulas are a subset of the terms and formulas respectively. Any term of the form $\#X$ is not an inductive term, and indeed no term or formula with a free

[10] This terminology is used to distinguish between these formulas and other for which induction does not hold. Hopefully no confusion will be caused by the distinct uses of the term inductive formulas elsewhere in the literature.

second-order variable is inductive. Likewise $\mathbb{N}0$, $\forall z(x = z)$ and $\exists y(S0y)$ are not inductive formulas, while $\Box\forall z \in \mathbb{N}(x = z)$ and $\Diamond\exists y \in \mathbb{N}(S0y)$ are.

A formula φ is *stable* when:

$$\vDash_{\mathsf{PI}} \varphi \rightarrow \Box\varphi. \tag{1.8}$$

Stability is taken from Linnebo's (2013, p. 211) work on set theory in a modal setting. It means once a formula has been made true it stays true. As we saw in Lemma 1.5, S, S^{+}, $S^{+=}$, and $\mathbb{N}$ are all stable and an example of an unstable formula is $\neg\mathbb{N}$. Fortunately, the inductive formulas all have the property of being stable, as we will now prove. This will allow us to prove induction for these formulas.

Lemma 1.20. *If φ is an inductive formula then $\vDash_{\mathsf{PI}} \varphi \rightarrow \Box\varphi$.*

Proof. In what follows we prove by induction on the complexity of the inductive formulas that both $\varphi \rightarrow \Box\varphi$ and $\Diamond\varphi \rightarrow \varphi$. The second condition is included to deal with the case of negation.

Base case: $x = y$ and $x = 0$: The result follows from the evaluation of $\#\varnothing$ being rigid and the identity relation being interpreted as the identity from the metalanguage. Note that for $S, +$, and $\times$ that $\Diamond\psi \rightarrow \psi$ follows simply because R is transitive and they start with a $\Diamond$. $S(x, y)$: See Lemma 1.5. $+(x, y, z)$: Assume that $\mathcal{M}, w \vDash +(a, b, c)$. It follows that there exists a world w' accessible from w and nonintersecting sets $A, B \subseteq D(w')$ satisfying $+$. Let s be a world such that $R(w, s)$. Then by directedness, there is a world s' such that $R(s, s')$ and $R(w', s')$, and $A, B \subseteq D(s')$. So $+(a, b, c)$ holds at s. $\times(x, y, z)$: The reasoning is essentially the same as that used for $+$.

Now we proceed to the induction step. We will only show the case of the quantifier as $\neg$ and $\wedge$ proceed as one would expect. $\Diamond\exists x \in \mathbb{N}\ \psi$: Assume $\mathcal{M}, s \vDash \Diamond\Diamond\exists x \in \mathbb{N}\ \psi$. It follows by transitivity that $\mathcal{M}, s \vDash \Diamond\exists x \in \mathbb{N}\ \psi$. Now we show that $(\Diamond\exists x \in \mathbb{N}\ \psi) \to (\Box\Diamond\exists x \in \mathbb{N}\ \psi)$. First take a world w such that $\Diamond\exists x \in \mathbb{N}\ \psi$ holds at w. Then take worlds s, w' such that $R(w, s)$, $R(w, w')$, $\exists x \in \mathbb{N}\ \psi$ holds at w' and we want to show $\Diamond\exists x \in \mathbb{N}\ \psi$ holds at s. At w' there is an $a \in D(w')$ such that $a \in I(\mathbb{N}, w')$ and $\psi(a)$ holds at w'. So, by Lemma 1.5, $\mathbb{N}a \to \Box\mathbb{N}a$ holds at w' and by the induction hypothesis, $\psi(a) \to \Box\psi(a)$. Let s' be such that $R(s, s')$ and $R(w', s')$, such a world exists by directedness. It follows that $\mathbb{N}a$ and $\psi(a)$ hold at s' and as s' is accessible from s we have proven $\Diamond\exists x \in \mathbb{N}\ \psi$ holds at s. □

We can now prove that the modalized induction schema holds for all inductive formulas. We do this by showing the more general result that induction holds for all stable formulas.

Lemma 1.21. *If φ is stable, then*

$$\vDash_{\mathsf{PI}} [\varphi(0) \wedge \Box\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x, y) \to \varphi(y))] \to \Box\forall x \in \mathbb{N}\ \varphi(x).$$

Proof. Let w be a world. Further, we assume the antecedent of the induction schema holds so let $\varphi(0)$ and $\Box\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x, y) \to \varphi(y))$ hold at w. Let s be a world accessible from w and let $a \in I(\mathbb{N}, s)$. We will show that $\varphi(a)$ at s. If $a = 0$ then, as φ is stable, we are done so assume not.

As $a \in I(\mathbb{N}, s)$, if we prove $\forall x, y(\varphi(x) \wedge \mathbb{N}x \wedge S(x, y) \to \varphi(y) \wedge \mathbb{N}y)$ and $\forall x(S(0, x) \to \varphi(x) \wedge \mathbb{N}x)$ hold at s then we have satisfied the antecedent of $S^{+}0a$ and so it follows that $\varphi(a) \wedge \mathbb{N}a$ at s.

At s we have $\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x, y) \to \varphi(y))$. We also have that if $x \in I(\mathbb{N}, s)$, and $S(x, y)$ hold at s then by Lemma 1.2 that $y \in I(\mathbb{N}, s)$. This proves $\forall x, y(\varphi(x) \wedge \mathbb{N}x \wedge S(x, y) \to \varphi(y) \wedge \mathbb{N}y)$ at s.

From $a \in I(\mathbb{N}, s)$ it follows that $0 \in D(s)$. Assume $x \in D(s)$ and $S0x$, as $0 \in D(s)$ it follows by Lemma 1.1 that $x \in I(\mathbb{N}, s)$. It then follows by the stability of φ that $\varphi(0)$ at s. As such we have the antecedent of $\forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x,y) \rightarrow \varphi(y))$ so we get $\varphi(x)$. And from this it follows that $\forall x(S(0,x) \rightarrow \varphi(x) \wedge \mathbb{N}x)$ holds at s. □

So we have proven the modalized induction axiom restricted to inductive formulas. But we cannot prove modalized induction for all formulas in the language of potentially infinite models, as the following counterexample shows.

Lemma 1.22. *If $\varphi(x)$ is $\forall z(z = x)$, then*

$$\nvDash_{\mathsf{PI}} [\varphi(0) \wedge \Box \forall x, y \in \mathbb{N}(\varphi(x) \wedge S(x,y) \rightarrow \varphi(y))] \rightarrow \Box \forall x \in \mathbb{N}\ \varphi(x).$$

Proof. It is sufficient to show there is a model and a world in the model where this statement is false. Take the minimal model from Example 1.1 and world $\mathbf{0}$, where $D(\mathbf{0}) = \{\mathbf{0}\}$. Clearly $\mathcal{M}, \mathbf{0} \vDash \forall z(z = 0)$. Let $w \in W$ be such that $R(\mathbf{0}, w)$ and assume that for all $x, y \in I(\mathbb{N}, w)$, that $\forall z(z = x)$ and $S(x,y)$ hold at w. As everything in the domain is equal to x it follows that $y = x$ and so $\forall z(z = y)$ at w. So $\mathcal{M}, \mathbf{0} \vDash \Box \forall x, y \in \mathbb{N}(\forall z(z = x) \wedge S(x,y) \rightarrow \forall z(z = y))$. But it does not follow that $\Box \forall x \in \mathbb{N}\ \forall z(z = x)$, because $1 \in W$ is a counterexample as $D(\mathbf{1}) = \{\mathbf{0}, \mathbf{1}\}$. □

1.5 Proof of Theorem 1.2

We now have almost all the pieces needed to prove Theorem 1.2. However, before we do that we need to discuss what a translation and interpretation are in our setting because we are moving between logics.

Intuitively, a *translation* between two languages starts with instructions on how to rewrite atomic formulas in one language into the other language. It does not make any changes to the propositional connectives but can restrict the quantifiers to objects meeting some conditions. In the current setting, however, we need a formal definition of what is to count as a translation when the underlying logics are different. This notion should, at the very least, capture the Linnebo translation. We offer the following definition as a minimal condition on any translation, though more will need to be done to ensure a widely applicable definition of translation and interpretation between logics.

Definition 1.10. *Let L_A and L_B be two logics extending first-order predicate logic, defined by the languages $\mathcal{L}_A$ and $\mathcal{L}_B$ and derivability relations $\vdash_{L_A}$ and $\vdash_{L_B}$ respectively. A generalised translation is given by a recursive map $(\cdot)^{\mathcal{G}} : \mathcal{L}_A \to \mathcal{L}_B$ which preserves free variables and a domain formula $\delta(x) \in \mathcal{L}_B$, such that the map is compositional on the propositional connectives and where for all unnested formulas*[11] *$\varphi_1, \ldots, \varphi_n, \psi$ containing free variables $x_1, \ldots, x_m$ one has the following:*

$$\varphi_1, \ldots, \varphi_n \vdash_{L_A} \psi \Rightarrow \delta(x_1), \ldots, \delta(x_m), \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}} \vdash_{L_B} \psi^{\mathcal{G}} \tag{1.9}$$

What we have done so far is an informal translation from the first-order language of arithmetic into the signature of the potentially infinite models. In Section 1.2 we showed how the atomic formulas could be translated. Further, the modalized versions of the axioms of PA^1 proven in Sections 1.3 and 1.4 are the translations of PA^1's axioms via the translation found in Section 1.2 and the Linnebo translation for the quantifiers.

While it has been set out in previous sections, for the sake of definiteness we here record the translation explicitly. We will call this translation $(\cdot)^{\mathcal{F}}$, as it is a Fregean translation. Three things are worth noting before we lay out the translation. The first is that the domain formula associated to this interpretation is $\mathbb{N}$ from Definition 1.8. The second is that the

[11] An unnested formula is one where the atomic subformulas of a formula contain at most one constant, function or relation (Hodges 1993, p. 58). We only give conditions for unnested formulas. So, for example, Sxy and $+(x, y, z)$ are unnested but $S0x$ and $+(0, 0, z)$ are nested. Every formula is equivalent to an unnested one (Hodges 1993, p. 59, Cor 2.6.2). As such the translation can be expanded to unnested formulas using this equivalence.

range of this translation is the inductive formulas from Definition 1.9. The third is that 1.11-1.13 are the same definitions given in 1.3, 1.4 and 1.5. We have not changed the definitions we are working with. Rather, we merely show how these definitions can be used to define the interpretation function $(\cdot)^{\mathcal{F}}$.

$$0^{\mathcal{F}} \equiv \#\varnothing, \tag{1.10}$$

$$Sab^{\mathcal{F}} \equiv \Diamond\exists G \exists u[Gu \wedge (b = \#G) \wedge (a = \#G \cup \{u\})], \tag{1.11}$$

$$+(a,b,c)^{\mathcal{F}} \equiv \Diamond\exists X, Y(a{=}\#X \wedge b{=}\#Y \wedge c{=}\#X \cup Y \wedge X \cap Y = \varnothing), \tag{1.12}$$

$$\times(a,b,c)^{\mathcal{F}} \equiv \Diamond\exists X, P[\#X = b \wedge \forall x \in X(\#\{y \mid Pxy\} = a) \wedge \forall x, y \in X(x \neq y \to \{z \mid Pxz\} \cap \{z \mid Pyz\} = \varnothing) \wedge \#\bigcup_{x \in X}\{y \mid Pxy\} = c], \tag{1.13}$$

$$(\psi \wedge \chi)^{\mathcal{F}} \equiv \psi^{\mathcal{F}} \wedge \chi^{\mathcal{F}}, \tag{1.14}$$

$$(\neg\psi)^{\mathcal{F}} \equiv \neg\psi^{\mathcal{F}}, \tag{1.15}$$

$$(\forall x\psi)^{\mathcal{F}} \equiv \Box\forall x(\mathbb{N}(x) \to \psi^{\mathcal{F}}), \tag{1.16}$$

$$(\forall X^n\psi)^{\mathcal{F}} \equiv \Box\forall X^n(\forall x_1, \ldots, x_n(X^n x_1 \ldots x_n \to \mathbb{N}(x_1) \wedge \cdots \wedge \mathbb{N}(x_n)) \to \psi^{\mathcal{F}}). \tag{1.17}$$

To see that this is a generalised translation all that remains to be shown is that deduction is preserved by our translation. We need this result for both E_{PI} and I_{PI}.[12]

Lemma 1.23. *Let $\varphi_0, \ldots, \varphi_n, \psi$ be unnested formulas in the language of PA^1 with free variables $v_0, \ldots, v_m$, it follows that if $\varphi_0, \ldots, \varphi_n \vdash \psi$, then $\mathbb{N}(v_0), \ldots, \mathbb{N}(v_m), \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \psi^{\mathcal{F}}$. Further, it is PA^1-provable that if $\varphi_0, \ldots, \varphi_n \vdash \psi$ then $\mathsf{ACA}_0 \vdash$ "$\mathbb{N}(v_0), \ldots, \mathbb{N}(v_m)$, $\varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \psi^{\mathcal{F}}$".*

[12] Recall that we formalised I_{PI} in ACA_0, and those interested in the nuts and bolts are directed to Appendix 1.9.

The first part of this Lemma is similar to Linnebo (2013, Thm. 5.4.). But he proves a version of this which does not restrict the quantifiers to a domain. The modification to our case is simple and so we omit the proof.

On its own a translation is not very interesting. However, a translation is an *interpretation* if the translations of the axioms of the interpreted theory can be proven in the interpreting theory.

Definition 1.11. *Let T_A and T_B be L_A and L_B theories respectively, where a theory is a set of sentences not necessarily closed under deduction. A generalised translation $(\cdot)^{\mathcal{G}} : \mathcal{L}_A \to \mathcal{L}_B$ interprets T_A in T_B, if for all $\mathcal{L}_A$ unnested sentences χ:*

$$\mathsf{T}_A \vdash_{L_A} \chi \Rightarrow \mathsf{T}_B \vdash_{L_B} \chi^{\mathcal{G}} \tag{1.18}$$

It is a recursive interpretation if the collection of $\mathcal{L}_A$ and $\mathcal{L}_B$ formulas are recursive, T_A and T_B are also recursive, as is $(\cdot)^{\mathcal{G}}$, and there are recursive maps from proofs to proofs which witness the truth of equations (1.9) and (1.18). If T extends PA^1, then say that the interpretation is T-verifiable *if the recursive functions are provably total in T and if the universal closures of the arithmetized versions of 1.9 and 1.18 are provable in T.*

So, the proofs of Sections 1.3 and 1.4 show our translation is an interpretation of PA^1 in E_{PI}. However, to show it is an interpretation in I_{PI} a certain level of caution is needed because I_{PI} does not have a background derivability relation. To resolve this, we take $\varphi_0, \ldots, \varphi_n \vdash_{L_{PI}} \varphi$ to be $\mathsf{ACA}_0 \vdash$ "$\varphi_0, \ldots, \varphi_n \vDash_{\mathsf{PI}} \varphi$", where this is as defined in Appendix 1.9. And, of course I_{PI} is just as defined in (1.2) of section 1.1, namely the set of sentences φ such that $\mathsf{ACA}_0 \vdash$ "$\vDash_{\mathsf{PI}} \varphi$". We then need to show the following:

Lemma 1.24. *For all sentences φ in the language of PA^1, if $\mathsf{PA}^1 \vdash \varphi$ then $\mathsf{ACA}_0 \vdash$ "$\vDash_{\mathsf{PI}} \varphi^{\mathcal{F}}$". Further, it is PA^1-provable that if $\mathsf{PA}^1 \vdash \varphi$ then $\mathsf{ACA}_0 \vdash$ "$\vDash_{\mathsf{PI}} \varphi^{\mathcal{F}}$".*

Proof. By Lemmas 1.8-1.12 and 1.20 and 1.21 we know that if φ is an axiom of PA^1 then $\mathsf{ACA}_0 \vdash$ "$\models_{\mathsf{PI}} \varphi^{\mathcal{F}}$". Assume $\mathsf{PA}^1 \vdash \varphi$ not an axiom, then there are n axioms of PA^1, $\varphi_0, \dots, \varphi_n$, such that $\varphi_0, \dots, \varphi_n \vdash \varphi$. Then as we can always take the universal closure of axioms and φ is a sentence it follows by Lemma 1.23 that $\mathsf{ACA}_0 \vdash$ "$\varphi_0^{\mathcal{F}}, \dots, \varphi_n^{\mathcal{F}} \models_{\mathsf{PI}} \varphi^{\mathcal{F}}$". Given that the axioms are PI valid, it follows that $\mathsf{ACA}_0 \vdash$ "$\models_{\mathsf{PI}} \varphi^{\mathcal{F}}$". □

This final piece gives us the proof of:

Theorem 1.4 (1.2.ii.)**.** *There is a generalised translation from the language of* PA^1 *to the second-order modal language with octothorpe that interprets* PA^1 *in* I_{PI}*. Further, this is a* PA^1*-verifiable generalised interpretation.*

To prove the first half of Theorem 1.2 we need to define formulas that pick out the numbers in PA^1 and E_{PI}. In PA^1 let $\tau_0(x) \equiv (x = 0)$ and $\tau_{n+1}(x) \equiv \exists y(\tau_n(y) \wedge Syx)$. In E_{PI} let $\sigma_0(x) \equiv (x = 0)$ and $\sigma_{n+1}(x) \equiv \Diamond\exists y \in \mathbb{N}(\sigma_n(y) \wedge Syx)$. Note that $(\tau_0(x))^{\mathcal{F}} \equiv (x = 0)^{\mathcal{F}} \equiv \sigma_0(x)$ and $(\tau_{n+1}(x))^{\mathcal{F}} \equiv (\exists y(\tau_n(y) \wedge Syx))^{\mathcal{F}} \equiv \Diamond\exists y \in \mathbb{N}((\tau_n(y))^{\mathcal{F}} \wedge Syx) \equiv \sigma_{n+1}(x)$. With this we can state the following preliminary Lemma; we omit the proof which is long but not illuminating:

Lemma 1.25. *For every $k \geq 0$ and every unnested formula $\theta(x_1, \dots, x_k)$ in the signature of* PA^1 *and every k-tuple of natural numbers $n_1, \dots, n_k$ one has that :*

$$\mathbb{N} \models \theta(n_1, \dots, n_k) \quad \Longrightarrow \quad \models_{\mathsf{PI}} \forall x_1, \dots, x_k \in \mathbb{N}(\bigwedge_{i=1}^{k} \sigma_{n_i}(x_i) \to \theta^{\mathcal{F}}(x_1, \dots, x_k)) \tag{1.19}$$

In the case of $k = 0$, this is to say: for every unnested sentence θ in the signature of PA^1 *one has that*

$$\mathbb{N} \models \theta \quad \Longrightarrow \quad \models_{\mathsf{PI}} \theta^{\mathcal{F}} \tag{1.20}$$

Theorem 1.2.i follows from (1.20) of Lemma 1.25. This give us our proof of:

Theorem 1.5 (1.2.i.). *There is a generalised translation from the language of* PA^1 *to the second-order modal language with octothorpe that interprets* TA^1 *in* E_{Pl}.

1.6 Proof of Theorem 1.3

It has been shown by Linnebo and Shapiro (2019, §7) that the Linnebo translation cannot interpret comprehension because modalized comprehension requires the existences of a set of all possibly existing things. However, this leaves open the question of whether there is a different translation which can interpret PA^2. Here we will demonstrate that there is no translation from TA^2 to E_{Pl} nor from PA^2 to I_{Pl} by proving Theorem 1.3, our second main theorem. The first part of Theorem 1.3 follows from relatively simple Tarskian considerations:

Theorem 1.6 (1.3.i). *There is no generalised translation from the language of* PA^2 *to the second-order modal language with octothorpe that interprets* TA^2 *in* E_{Pl}.

Proof. Assume for a contradiction that there is an interpretation $(\cdot)^{\mathcal{G}}$ that interprets TA^2 in E_{Pl}. Note that as TA^2 is complete it follows that this is a faithful interpretation; i.e. if $\vDash_{\mathsf{Pl}} \varphi^{\mathcal{G}}$ then $\mathbb{N} \vDash \varphi$. As E_{Pl} is Π^1_1-definable it follows that there is a predicate P such that for all φ in the second-order modal language with octothorpe we have $\vDash_{\mathsf{Pl}} \varphi$ if and only if $\mathbb{N} \vDash P(\text{“}\varphi\text{”})$. (Here we use quotation marks for Gödel numbering for both the language of PA^2 and the second-order modal language with octothorpe.) But then as generalised translations are recursive we can represent $(\cdot)^{\mathcal{G}}$ in $\mathbb{N}$ as g. It follows that $P(g(\text{“}\psi\text{”}))$, where ψ is in the language of PA^2, is a truth predicate for TA^2. But this contradicts Tarski's theorem. □

The proof of the second part of the theorem is trickier and requires Gödelian considerations. Recall the definition of T-verifiable generalised translation and interpretation from Definitions 1.10 and 1.11 in Section 1.5. There we proved that we have a PA^1-verifiable interpretation of PA^1 in I_{Pl} by Lemma 1.24. Given that we defined $\mathsf{I}_{\mathsf{Pl}} \vdash \varphi$ as $\mathsf{ACA}_0 \vdash \text{“}\vDash_{\mathsf{Pl}} \varphi\text{"}$, that is $\mathsf{PA}^1 \vdash \forall\varphi[\text{“}\mathsf{PA}^1 \vdash \varphi\text{”} \rightarrow \text{“}\mathsf{ACA}_0 \vdash \text{“} \vDash_{\mathsf{PI}} \varphi^{\mathcal{F}}\text{””}]$. Here we show that there is no PA^2-verifiable interpretation of PA^2 in I_{PI}. We can write this as: there is no generalised translation $(\cdot)^{\mathcal{G}}$ from the language of PA^2 to the second-order modal language with octothorpe such that $\mathsf{PA}^2 \vdash \forall\varphi[\text{“}\mathsf{PA}^2 \vdash \varphi\text{”} \rightarrow \text{“}\mathsf{ACA}_0 \vdash \text{“} \vDash_{\mathsf{PI}} \varphi^{\mathcal{G}}\text{””}]$.

Theorem 1.7 (1.3.ii)**.** *There is no generalised translation from the language of* PA^2 *to the second-order modal language with octothorpe that* PA^2*-verifiably interprets* PA^2 *in* I_{PI}*.*

Proof. The systems Π^1_k-CA_0 are subsystems of PA^2 that have comprehension for Π^1_k formulas. As proofs are finite and so can only use finitely many instances of the comprehension schema any interpretation which is PA^2-verifiable will also be Π^1_k-CA_0-verifiable for some $k \geq 1$. Let $\varphi_1, \ldots, \varphi_n$ be a finite axiomatisation of Π^1_k-CA_0 for some $k \geq 1$ (Simpson 2009, pp. 303, 311–2). We will show, from the assumption that there is a Π^1_k-CA_0-verifiable translation $(\cdot)^{\mathcal{G}}$ from the languge of PA^2 to the second-order modal language with octothorpe that interprets PA^2 in I_{PI}, that Π^1_k-CA_0 proves its own consistency. This contradicts Gödel's second incompleteness theorem and so shows that no such $(\cdot)^{\mathcal{G}}$ can exist.

Note that $\mathsf{PA}^2 \vdash \varphi_1, \ldots, \varphi_n$ as all Π^1_k-CA_0 are subsystems of PA^2. We are assuming that $(\cdot)^{\mathcal{G}}$ interprets PA^2 in I_{PI}, so it follows that $\mathsf{ACA}_0 \vdash \text{“} \vDash_{\mathsf{PI}} \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}}\text{”}$. Let $\mathcal{A}$ be a model of Π^1_k-CA_0 for some k. So, we have $\mathcal{A} \vDash \text{“} \vDash_{\mathsf{PI}} \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}}\text{”}$. If $\mathcal{M}$ is the minimal model from Example 1.1 relative to $\mathcal{A}$ then we have then we have $\mathcal{A} \vDash \text{“}\mathcal{M} \vDash \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}}\text{”}$.

Now we show that $\mathcal{A} \vDash \neg Prv_{\varphi_1,\ldots,\varphi_n}(\psi \wedge \neg\psi)$, that is the consistency of Π^1_k-CA_0. Assume for a contradiction that $\mathcal{A} \vDash \exists\pi Prf_{\varphi_1,\ldots,\varphi_n}(\pi, \psi \wedge \neg\psi)$. Then as $(\cdot)^{\mathcal{G}}$ is a Π^1_k-CA_0-verifiable interpretation it follows $\mathcal{A} \vDash Prf_{\mathsf{ACA}_0}(\pi^{\mathcal{G}}, \text{“}\varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}} \vDash_{\mathsf{PI}} \psi^{\mathcal{G}} \wedge \neg\psi^{\mathcal{G}}{''})$.

Recall that Π^1_1-CA_0 proves Σ^1_1-reflection for ACA_0 (cf. Simpson (2009) Theorem VII.6.9.(4) p. 298 and Theorem VII.7.6.(1) p. 305). As Π^1_1-$\mathsf{CA}_0 \subseteq \Pi^1_k$-$\mathsf{CA}_0$, this means that for any Π^1_1 statement ψ we know Π^1_k-CA_0 proves $Prv_{\mathsf{ACA}_0}(\psi) \rightarrow \psi$. For all ψ, we know that “ $\vDash_{\mathsf{PI}} \psi$” is Π^1_1 and similarly for the local derivability relation (see Appendix 1.9). It follows that

$\mathcal{A} \vDash$ "$\varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}} \vDash_{\mathsf{PI}} \psi^{\mathcal{G}} \wedge \neg\psi^{\mathcal{G}}$" and as $\mathcal{A} \vDash$ "$\mathcal{M} \vDash \varphi_1^{\mathcal{G}}, \ldots, \varphi_n^{\mathcal{G}}$". It follows that $\mathcal{A} \vDash$ "$\mathcal{M} \vDash \psi^{\mathcal{G}} \wedge \neg\psi^{\mathcal{G}}$". And so $\mathcal{A} \vDash$ "$\mathcal{M} \models \psi^{G}$" and $\mathcal{A} \vDash$ "$\mathcal{M} \models \neg(\psi^{G})$". $\square$

We have now shown the two main results set out in the introduction.

1.7 Conclusion

We started with the worry that Hume's Principle had only infinite models and so any claim that it was analytic would mean that the claim that there are infinitely many objects is analytic. This worry has been noted before in the literature on neo-logicism, but little has been done to address it. Hale and Wright (2001a) state that without this the neo-logicist project cannot even get off the ground:

> To require of an acceptable abstraction that it should not be (even) weakly inflationary [that is require a countable infinity] would stop the neo-Fregean project dead in its tracks, before it even got moving (as it were). It will be clear that I think there is no good ground to impose such a requirement, and I shall not discuss it further. (Hale and Wright 2001a, pp. 417–8)

In this chapter we have explored the potentially infinite as one way to address this worry. The move to the potentially infinite does not rid us of posited infinities. We still require there to be an infinity of worlds and an infinity of objects across the worlds. But these infinities are less metaphysically questionable. So, for example, while Putnam and Hodes objected to the positing of actual infinities they allowed for possible infinities. And one could always try to further avoid the commitment by adopting an instrumentalist attitude towards the metatheory.

We have shown that the theory of potentially infinite models interprets first-order Peano arithmetic or first-order true arithmetic, depending on the strength of our meta-language. But we cannot interpret the equivalent second-order arithmetic theory. The difficulty seems to be the non-existence of a set of all the numbers across all the worlds. As our models are supposed to capture the idea of the potential infinite, we do not want the set of all the numbers across all the worlds to exist. It makes sense that the potential infinite does not capture the infinite progression of the natural numbers as well as actual infinity and this might go some way to explaining why we get the weaker first-order theory.

This allows a fuller understanding of the role of the potentially infinite in the foundation of mathematics. Unlike Hodes, we see that a certain amount of mathematics can be recovered, though some other story would need to be told about more advanced mathematics. It also offers evidence that the ontological commitments that come with Hume's Principle, and which make some reject the claim that its truth is analytic, cannot be avoided by moving to the modal setting if one wants full second-order Peano arithmetic. For in weakening our ontological commitments, we also weakened the mathematical theory which we can recover.

1.8 Formal Theories

Here we will spell out the theories other than E_{PI} and I_{PI} which are used in the proofs above. Unlike E_{PI} and I_{PI} none of these are modal theories, however, most are second-order theories.

The weakest theory we consider is first-order Robinson's Q. For a more complete reference see, for example, Hájek and Pudlák (1998, p. 28).

Definition 1.12. Q *is the usual formalization of Robinson's arithmetic. It consists of the universal closure of the following axioms:*

$$s(x) \neq 0; \quad \textit{(Q1)} \qquad s(y) = s(z) \to y = z; \quad \text{(Q2)}$$

$$x + 0 = x; \quad \textit{(Q3)} \qquad x + s(y) = s(x + y); \quad \text{(Q4)}$$

$$x \times 0 = 0; \quad \textit{(Q5)} \qquad x \times s(y) = (x \times y) + y. \quad \text{(Q6)}$$

Note that in the body of the text we do not use this formulation but rather one with relations instead than functions.[13] We have offered this formulation for readability. The relation formulation gives you the obvious translation of the above, plus an additional 6 axioms ensuring that the relations $S, +, \times$ are the graphs of functions.

We also consider the extensions of Q to PA^1 by the addition of the first-order induction schema, and PA^2 by the addition of the second-order induction axiom and Comprehension Schema. PA^1 is a first-order theory, but PA^2 is a second-order theory.

Definition 1.13. PA^1 *is* Q *plus the induction schema, where φ is a first-order formula:*

$$(\varphi(0) \land \forall x(\varphi x \to \varphi(s(x)))) \to \forall x \varphi(x) \qquad \text{(Induction Schema (IS))}$$

PA^2 *is* Q *plus the induction axiom and Comprehension Schema:*

$$\forall P[(P0 \land \forall x(Px \to P(s(x)))) \to \forall x Px] \qquad \text{(Induction Axiom (IS))}$$

$$\forall \bar{y}, \bar{Y} \exists X \forall x(X(x) \leftrightarrow \varphi(x, \bar{y}, \bar{Y})) \qquad \text{(Comprehension Schema (CS))}$$

In the Comprehension Schema φ can be any formula of the language of PA^2 *in which X does not occur free.*

[13] We use a capital S for the relational successor and lower case s for the functional.

Again in the body of the text we use the natural adaptation to the setting of relations rather than functions. There are also two theories we use that are second-order and between PA^2 and PA^1 in strength. They both restrict comprehension. So, we first need to define the formulas we restrict to:

Definition 1.14. *(Simpson 2009, I.3.1, p. 6) An Arithmetical formula is a formula in the language of PA^2 which does not contain any set quantifiers, though it may contain free set variables.*

With this we can state ACA_0:

Definition 1.15. *(Simpson 2009, I.3.2, p. 7) ACA_0 is Q plus the Induction Axiom and Arithmetical Comprehension:*

$$\forall \bar{y}, \bar{Y} \exists X \forall x (X(x) \leftrightarrow \varphi(x, \bar{y}, \bar{Y})) \qquad \text{(Arithmetical Comprehension Schema (ACS))}$$

Where φ has to be an arithmetical formula and X may not occur free.

Note that as every formula of PA^1 is arithmetical, and ACA_0 contains the second-order induction axiom, every instance of the first-order induction schema is provable in ACA_0.

The next theories of arithmetic to be considered here are the Π^1_k-CA_0 which are used in the proof of Theorem 1.3. To define this theory, we first need to define Π^1_k (and Σ^1_k) formulas:

Definition 1.16. *(Simpson 2009, I.5.1, p. 16) A Π^1_1 formula is a formula in the language of PA^2 of the form $\forall X_1, \ldots, X_n \varphi$ where $X_1, \ldots, X_n$ are set variables and φ is an arithmetical formula.*

A Σ^1_1 formula is a formula in the language of PA^2 of the form $\exists X_1, \ldots, X_n \varphi$ where $X_1, \ldots, X_n$ are set variables and φ is an arithmetical formula.

A Π^1_k formula is a formula in the language of PA^2 of the form $\forall X_1, \ldots, X_n \varphi$ where $X_1, \ldots, X_n$ are set variables and φ is a Σ^1_{k-1} formula.

A Σ^1_k formula is a formula in the language of PA^2 of the form $\exists X_1, \ldots, X_n \varphi$ where $X_1, \ldots, X_n$ are set variables and φ is Π^1_{k-1} formula.

The definition of Π^1_k-CA_0 is much like the definition of ACA_0, except that the restriction on the comprehension axiom is broadened to include all Π^1_k formulas:

Definition 1.17. *(Simpson 2009, I.5.2, p. 17) Π^1_k-CA_0 is Q plus the Induction Axiom and Π^1_k Comprehension:*

$$\forall \bar{y}, \bar{Y} \exists X \forall x (X(x) \leftrightarrow \varphi(x, \bar{y}, \bar{Y})) \qquad (\Pi^1_k \text{ Comprehension Schema } (\Pi^1_k\text{CS}))$$

Where φ has to be a Π^1_k formula and X may not occur free.

We can define the intended model of these theories. Let $\mathbb{N}^1$ be $\{\omega, 0, s, +, \times\}$ where each term is interpreted as it is in the metatheory and $\mathbb{N}^2$ be $\mathbb{N}^1$ with $\mathscr{P}(\omega^n)$ as the domain of the second-order quantifiers. $\mathbb{N}^1$ is the intended model of Q and PA^1, while $\mathbb{N}^2$ is the intended model of PA^2, ACA_0, and Π^1_k-CA_0 for all k. As is well known, by Gödel's incompleteness theorems none of the theories we have seen so far are complete. We can define the complete theories of these models:

Definition 1.18. *Let TA^1 be $\{\varphi \mid \mathbb{N}^1 \vDash \varphi\}$ and TA^2 be $\{\varphi \mid \mathbb{N}^2 \vDash \varphi\}$.*

For the sake of completeness, we here define Hume's Principle (HP^2). This system is second-order also and consists of the cardinality principle displayed in Equation HP on page 24, the full Comprehension Schema, as in PA^2, and full comprehension for binary relations:

$$\forall \bar{y}, \bar{Y} \exists X \forall x, z (X(x, z) \leftrightarrow \varphi(x, z, \bar{y}, \bar{Y})) \qquad \text{(Binary Comprehension Schema (BCS))}$$

Comprehension for binary relations is required because the definition of HP^2 quantifies over bijections and when spelt out fully this turns out to be the claim that there is a second-order binary relation which is the graph of a bijection between the two sets.

1.9 Formal definition of I_{PI}

In the introduction we gave I_{PI} as the set $\{\varphi \mid \mathsf{ACA}_0 \vdash `\vDash_{\mathsf{PI}} \varphi'\}$. Here we will layout explicitly what we mean by defining the arithmetization of $\vDash_{\mathsf{PI}}$ in ACA_0.

It is importaint to note that the second-order variables in I_{PI} are taken to first-order variables in ACA_0. If all the first-order variables of I_{PI} are of the form x_i and all the second-order variables of I_{PI} are of the form Y_j then let all the first-order variables of ACA_0 be of the form x_i and Y_j, and the second-order variables of ACA_0 be of the form Z_v. In practice we will not stick to this strict distinction, but it can always be implemented by renaming the variables.

We do not restrict the domain of the first-order variables of I_{PI}; there is no need to pick out a subset of the domain of a model of ACA_0. However, the second-order variables of I_{PI} need to be restricted to codes for finite sets of numbers ordered by strict less than. This isn't difficult, we can simply borrow the coding found in the proof of incompleteness. A more complete explication can be found in Simpson (2009, Ch. 2.2). The second-order variables are required to be to some sequence $\pi(0)^{n_0} + \cdots + \pi(m)^{n_m}$ where $\pi(i)$ gives the ith prime and $n_0 < n_1 < \cdots < n_m$. Let $Seq(Y)$ be the name of the relation that ensures Y has the above properties. Further, let $nSeq(Y)$ mean that Y codes n-tuples of numbers. We will use this to code relations and relational variables. If x is the number of a sequence then let $[x]_i$ be the ith element and $ln(x)$ is the length of x.

We want to code PI models as sets of natural numbers. We know that we can always combine countably many countably infinite sets (just code n a member of the ith set as

$2^i + 3^n$). As such we will just show how to code $W, R, D, \#, \mathbf{a}$ as separate sets of natural numbers. Further, with $R, D, \#, \mathbf{a}$ we will talk about pairs (x, y), this should be understood as standing for the code $2^x + 3^y$.

(B.1) Let W be infinite $(\forall x \in W \exists y \in W(y > x))$,[14]

(B.2) let R be such that

   (a) for all $(i, j) \in R$ we have that $i, j \in W$,

   (b) $\forall x \in W \ R(x, x)$ (reflexive),

   (c) $\forall x, y, z \in W(R(x, y) \wedge R(y, z) \to R(x, z))$ (transitive),

   (d) $\forall x, y \in W(R(x, y) \wedge R(y, x) \to x = y)$ (anti-symmetric),

   (e) $\forall x, y \in W \ \exists z \in W(R(x, z) \wedge R(y, z))$ (directed),

(B.3) let D be such that

   (a) $D(w, Y)$ implies that $w \in W$ and $Seq(Y)$,

   (b) $\forall w \in W \exists Y \in Seq(D(w, Y) \wedge ln(Y) > 0)$ (every world has at least one element),

   (c) D is the graph of a function from W to Seq,

   (d) if $R(i, j)$ and $i \neq j$ and $D(i, X)$ and $D(j, Y)$ then $\exists u \forall v([X]_v \neq [Y]_u)$ (there is something in Y not in X) and $\forall v < ln(X) \exists u([X]_v = [Y]_u)$ (everything in X is in Y),

(B.4) let $\mathbf{a}$ be such that for each n there is exactly one x such that $\mathbf{a}(n, x)$ and if $\mathbf{a}(n, x)$ and $\mathbf{a}(m, x)$ then $n = m$, we then define $\#(Y, x)$ as $Seq(Y) \wedge \mathbf{a}(ln(Y), x)$.

Given a set of numbers $\mathcal{M}$ we will write $\mathcal{M} \in PIM$ to signify the set meets (B.1)–(B.4). We define sb (subset) as follows $Y \in sb(X)$ iff $Seq(Y) \wedge \forall i < ln(Y) \exists j([X]_j = [Y]_i)$. In

[14] Recall that our definition demanded that our set of worlds be countable. We cannot capture this in ACA_0 in the sense that ACA_0 has none standard models but we will have that we do not have more worlds than ACA_0 thinks there are natural numbers, which is sufficent for the role this plays in the proofs.

defining the arithmetisation note that we add free-variables for the model and the world, we will use $W_M, R_M, D_M, \#_M$, but these can be defined in terms of the model. So, if φ is a formula in the modal second-order language with octothorpe we translate it to some $\psi(w, W_M, R_M, D_M, \#_M)$ in the language of arithmetic. We define the arithmetisation as follows:

$$(x_i = x_j)^* \equiv x_i = x_j \tag{1.21}$$

$$(x_i = \#Y_j)^* \equiv \#_M(Y_j, x_i) \tag{1.22}$$

$$(Y_j x_i)^* \equiv \exists u(x_i = [Y_j]_u) \tag{1.23}$$

$$(\forall x \varphi)^* \equiv \forall x(\exists Y \in Seq(D_M(w, Y) \wedge \exists u(x = [Y]_u)) \to (\varphi)^*) \tag{1.24}$$

$$(\forall Y \varphi)^* \equiv \forall Y \in Seq(\exists X \in Seq(D_M(w, X) \wedge Y \in sb(X)) \to (\varphi)^*) \tag{1.25}$$

$$(\forall P^n \varphi)^* \equiv \forall P^n \in nSeq \tag{1.26}$$

$$(\exists X \in Seq(D_M(w, X) \wedge \forall (x_1, \ldots, x_n) \in P^n(\bigwedge_{1 \le i \le n} \exists j [X]_j = x_i)) \to (\varphi)^*)$$

$$(\Box \varphi)^* \equiv \forall s \in W_M(R_M(w, s) \to (\varphi)^*[w/s]) \tag{1.27}$$

where we commute over the logical connectives. This means that every formula arithmetised is arithmetical as defined in Appendix 1.8. For example, $\Box \forall v \Diamond \exists Z(v = \#Z)$ becomes

$$\forall s \in W_M(R_M(w, s) \to \forall v(\exists Y(D_M(s, Y) \wedge \exists u(v = [Y]_u) \to$$
$$\exists s' \in W_M(R_M(s, s') \wedge \exists Z \in Seq(\exists X(D_M(w, X) \wedge Z \in sb(X) \wedge \#_M(Z, v)))))). \tag{1.28}$$

Note '$\vDash_{\mathsf{PI}} \varphi$' means $\forall M \in PIM \forall w \in W_M(\varphi)^*$. It follows that this is then a Π^1_1 formula. Hence, if one were proceeding very formally, we would define I_{PI} as the set of all the φ such that $\mathsf{ACA}_0 \vdash \forall M \in PIM \forall w \in W_M(\varphi)^*$.

1.10 Proofs

First we prove the results discussed in Section 1.2. For the following Lemma, recall Definition 1.2 where $\vDash_{\mathsf{PI}} \varphi$ was defined as φ is true in all worlds in all potentially infinite models.

Lemma 1.26. $\vDash_{\mathsf{PI}} \exists x(x = 0) \to \forall y(S0y \to \mathbb{N}y))$

Proof. Let w be a world and assume that $0 \in D(w)$. Let $b \in D(w)$ be such that $S(0, b)$ holds at w. Assume X satisfies the antecedent of $S^{+}0b$ at w, it follows by $\forall x(S0x \to Xx)$ that Xb and so $\mathbb{N}(b)$ at w. □

Lemma 1.27. $\vDash_{\mathsf{PI}} \forall x, y(\mathbb{N}x \land Sxy \to \mathbb{N}y)$

Proof. Given a world w, let $a, b \in D(w)$ be such that $\mathbb{N}(a)$ and $S(a, b)$ at w. Assume X is such that it satisfies the antecedent of $S^{+}0b$ at w. If $a = 0$ then the result follows from Lemma 1.1, so assume not. It follows that Xa as $S^{+=}0a$ at w. But then $\forall x, y(Xx \land Sxy \to Xy)$ at w by assumption, so it follows that Xb and so $S^{+=}0b$ at w. Because $\mathbb{N}(a)$ at w, we know $\exists x(x = 0)$ at w and so $\mathbb{N}(b)$ at w. □

Lemma 1.28. $\models_{PI} \mathbb{N}x \to \exists y\, y = x$

Proof. Let $w \in W$ and assume that a is such that $\mathbb{N}(a)$ at w. Let $X = \{x \mid \mathcal{M}, w \vDash \mathbb{N}(x) \land \exists y\, y = x\}$, we will show that X satisfied the antecedent of $S^{+}0x$. We want to show that $\forall y(S0y \to \exists u(u = y) \land \mathbb{N}y)$. Assume not then $\exists y(S0y \land (\neg\exists u(u = y) \lor \neg\mathbb{N}y))$. As $\mathbb{N}(a)$ at w it follows that $\exists x\, x = 0$ at w. So, by Lemma 1.1 it follows that $\forall y(S0y \to \mathbb{N}y)$. It therefore follows that $\exists y(S0y \land \neg\exists u(u = y))$ This is clearly contradictory. We further need to show that $\forall y, z(Syz \land \exists u(y = u) \land \mathbb{N}y \to \exists v(z = v) \land \mathbb{N}z)$. Again, assume not then $\exists y, z(Syz \land \exists u(y = u) \land \mathbb{N}y \land (\neg\exists v(z = v) \lor \neg\mathbb{N}z))$. But like before, by Lemma 1.2 $\forall y, z(\mathbb{N}y \land Syz \to \mathbb{N}z)$. With this we get a contradiction as we have $\exists z Syz$ and $\neg\exists v\, v = z$. It follows that Xa at w and so $\exists y\ y = a$ at w. This result comes about because our quantifiers are actualist and everything that satisfies $\mathbb{N}$ falls in the range of a quantifier. $\square$

Lemma 1.29. *Let w be a world and let n be the first number such that $\mathbf{a}_n \notin D(w)$. Then if $n > 0$, it follows that $\{0, \mathbf{a}_1, \ldots, \mathbf{a}_{n-1}\} = I(\mathbb{N}, w)$, and further, $n = 0$ iff $I(\mathbb{N}, w) = \varnothing$.*

Proof. We start by showing $0 \notin D(w)$ iff $I(\mathbb{N}, w) = \varnothing$. Note that, by the definition of $\mathbb{N}$, 0 must exist at a world s for $\mathbb{N}x$ at s to be true of any $x \in D(s)$. So, if $0 \notin D(s)$ then $I(\mathbb{N}, s) = \varnothing$. Further, if $0 \in D(s)$ then $I(\mathbb{N}, s) \neq \varnothing$ as $0 \in I(\mathbb{N}, s)$.

Now we show that for all worlds s if $\mathbf{a}_0, \ldots, \mathbf{a}_m \in D(s)$ then $\mathbf{a}_0, \ldots, \mathbf{a}_m \in I(\mathbb{N}, s)$. We proceed by induction on m. First let $\mathbf{a}_0 \in D(s)$ then 0 exists at s, so by definition of $\mathbb{N}$ it follows that $\mathbf{a}_0 \in I(\mathbb{N}, s)$. Now assume the induction hypothesis that if $\mathbf{a}_0, \ldots, \mathbf{a}_m \in D(s)$ then $\mathbf{a}_0, \ldots, \mathbf{a}_m \in I(\mathbb{N}, s)$. Also assume $\mathbf{a}_0, \ldots, \mathbf{a}_{m+1} \in D(s)$. Note $S\mathbf{a}_m\mathbf{a}_{m+1}$ holds at w as $\mathbf{a}_m = \#\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ and $\mathbf{a}_{m+1} = \#\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \cup \{\mathbf{a}_{m+1}\}$. So, by Lemma 1.2 it follows that $\mathbf{a}_{m+1} \in I(\mathbb{N}, w)$.

The above shows that $\{0, \mathbf{a}_1, \ldots, \mathbf{a}_{n-1}\} \subseteq I(\mathbb{N}, w)$. To show they are equal we will show that nothing else is a member of $\mathbb{N}$ at w. Let $X = \{\mathbf{a}_0, \ldots \mathbf{a}_{n-1}\} \subseteq D(w)$ and $\mathbf{a}_n \notin D(w)$. We will show that X satisfies the antecedent of $S^{+=}0x$ and so any $y \in I(\mathbb{N}, w)$ is also in X.

Every $y \in D$ such that $S0y$ holds at w is equal to $\mathbf{a}_1$. This is because $0 = \#\varnothing$ and so $y = \#\varnothing \cup \{u\}$ for some u in a world w' accessible from w, and $\#\varnothing \cup \{u\}$ must be a singleton so $\#\varnothing \cup \{u\} = \mathbf{a}_1$. Now as we are in the case where $0 \in D(s)$ we have either $\mathbf{a}_1 \in D(s)$ in which case $X\mathbf{a}_1$ by assumption or $\mathbf{a}_1 \notin D(s)$ and so no $x \in D(w)$ satisfies $S0x$. In either case it follows that $\forall x(S(0, x) \to Xx))$.

Assume $x, y \in D(w)$, Xx and Sxy at w. As Xx it follows that $x = \mathbf{a}_i$ for some $0 \leq i \leq n-1$. Therefore $y = \mathbf{a}_{i+1}$, but as $y \in D(w)$ it follows that $i < n - 1$ as if $i = n - 1$ then $y = \mathbf{a}_n$

and $\mathbf{a}_n \notin D(w)$. From this we get that $0 < i+1 \leq n-1$ and so Xy. From which it follows that $(\forall x, y(Xx \wedge S(x,y) \rightarrow Xy)$. $\square$

Lemma 1.30. *$\vDash_{\mathsf{PI}} S(x,y) \rightarrow \Box S(x,y)$ holds, as does $\vDash_{\mathsf{PI}} S^+(x,y) \rightarrow \Box S^+(x,y)$, $\vDash_{\mathsf{PI}} S^{+=}(x,y) \rightarrow \Box S^{+=}(x,y)$ and $\vDash_{\mathsf{PI}} \mathbb{N}x \rightarrow \Box \mathbb{N}x$.*

Proof. For S, let s be a world and assume Sab holds at s. Let w be a world such that $R(s,w)$. Take a world t accessible from s where A and u, witnessing Sab, exist. That is $a = \#A$ and $b = \#A \cup \{u\}$. By directedness there is a world t' such that $R(t,t')$ and $R(w,t')$. Because the domains are growing A and u exist at t'. And because $\#$ is rigid $a = \#A$ and $b = \#A \cup \{u\}$ at t'. And so, it follows as $R(w,t')$ that Sab at w.

Now for S^+, take a world s and assume S^+ab holds at s. Let w be such that $R(s,w)$. We will show that S^+ab holds at w. We must show that b is a member of all sets at w which are closed under successors of a. Let $X \subseteq D(w)$ be such a set, that is $\forall x, y(Xx \wedge Sxy \rightarrow Xy)$ and $\forall x(Sax \rightarrow Xx)$ hold at w. We will show that $b \in X$ by showing that $b \in Y = X \cap D(s)$. Note that Y exists at s and is still equal to $X \cap D(s)$ as second-order variables are rigid in our system. It follows that if we can show $\forall x, y(Yx \wedge Sxy \rightarrow Yy)$ and $\forall x(Sax \rightarrow Yx)$ hold at s then Yb at s and because second-order variables are rigid Yb at w.

Let us show $\forall x, y(Yx \wedge Sxy \rightarrow Yy)$ at s. Assume $c, d \in D(s)$, $c \in Y$, and Scd at s, then $c, d \in D(w)$ because our domains are growing. It follows by the first paragraph of the proof that Scd at w. Therefore, as $\forall x, y(Xx \wedge Sxy \rightarrow Xy)$ at w and $c \in X$, because $c \in Y$, we get $d \in X$ and so $d \in X \cap D(s) = Y$ at w. But as Y is rigid, $d \in Y$ at s. From which it follows that $\forall x, y(Yx \wedge Sxy \rightarrow Yy)$ holds at s. The proof that $\forall x(Sax \rightarrow Yx)$ holds at s is similar. So as S^+ab holds at s it follows that $b \in Y$ and so $b \in X$. And as X was arbitrary, S^+ab holds at w. The proofs for $S^{+=}$ and $\mathbb{N}$ are virtually identical. $\square$

Lemma 1.31. *For all $w \in W$ and $X \subseteq D(w)$, there is a world s such that $R(w,s)$ and $\#X \in I(\mathbb{N}, s)$.*

Proof. $\#X$ is $\mathbf{a}_n$ for some n. So, given Lemma 1.4 $\mathbb{N}\mathbf{a}_n$ at a world when $\mathbf{a}_0, \ldots, \mathbf{a}_n$ exist. As every cardinality exists at some world or other, and our worlds are directed, such a world can always be found. □

Lemma 1.32. *Let $\varphi_0, \ldots, \varphi_n, \psi$ be unnested formulas in the language of PA^1 with free variables $v_0, \ldots, v_m$, it follows that if $\varphi_0, \ldots, \varphi_n \vdash \psi$, then $\mathbb{N}(v_0), \ldots, \mathbb{N}(v_m), \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \psi^{\mathcal{F}}$. Further, it is PA^1-provable that if $\varphi_0, \ldots, \varphi_n \vdash \psi$ then $\mathsf{ACA}_0 \vdash$ "$\mathbb{N}(v_0), \ldots, \mathbb{N}(v_m)$, $\varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \psi^{\mathcal{F}}$".*

Proof. The proof is an induction on the length of proof in a Hilbert-style deductive system for PA^1 and ACA_0. That is, we prove by induction on length of proof π_1 of $\varphi_0, \ldots, \varphi_n \vdash \psi$ that there is a proof π_2 of $\mathbb{N}(v_0), \ldots, \mathbb{N}(v_m), \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \psi^{\mathcal{F}}$ in ACA_0; and this induction on length of proofs is evidently PA^1-formalizable.

For the base case let the proof be of length 1. Then the conclusion is either a hypothesis (this case is trivial) or it is a logical axiom. As the translation commutes over the connectives the only case we have to consider is the logical axioms associated with the quantifiers (Troelstra and Schwichtenberg 2000, p. 51). Consider the logical axiom $\forall x\psi(x) \to \psi(t)$. It follows that t must be 0 or a variable y (since we are using the relational presentation of PA^1 the only constant in the language is 0 and there are no function symbols). So it is sufficent to show $\mathbb{N}(y) \vDash_{\mathsf{PI}} (\Box\forall x \in \mathbb{N}\psi(x)^{\mathcal{F}}) \to \psi(y)^{\mathcal{F}}$ and $\vDash_{\mathsf{PI}} (\Box\forall x \in \mathbb{N}\psi(x)^{\mathcal{F}}) \to \psi(0)^{\mathcal{F}}$. The first follows easily given the elimination of the necessity operator and the universal quantifier and also that $\mathbb{N}(y)$. For the second note that $\Box\forall x \in \mathbb{N}\psi(x)^{\mathcal{F}}$ and $\neg\psi(0)^{\mathcal{F}}$ are contradictory as, even though 0 may not exist in some worlds, $\neg\psi(0)^{\mathcal{F}}$ is stable and so were it true it would hold at all future worlds which cannot be the case.

Next consider the logical axiom $\forall x(\psi \to \chi(x)) \to (\psi \to \forall y\chi(y))$ where x does not occur free in ψ and either $y = x$ or y not free in χ. Let s be an arbitrary world such that $\Box\forall x \in \mathbb{N}(\psi^{\mathcal{F}} \to \chi(x)^{\mathcal{F}})$ and $\psi^{\mathcal{F}}$ holds at s. Take a world w accessible from s then $\forall x \in \mathbb{N}(\psi^{\mathcal{F}} \to \chi(x)^{\mathcal{F}})$

holds at w and $\psi^{\mathcal{F}}$ holds at w by the stability of inductive formulas. Let $a \in I(\mathbb{N}, w)$, then $\psi^{\mathcal{F}} \to \chi(a)^{\mathcal{F}}$ at w and so $\chi(a)^{\mathcal{F}}$. From this it follows that $\forall y \in \mathbb{N}\chi(y)$ at w. As w was arbitrary, $\Box\forall y \in \mathbb{N}\chi(y)$ at s. From this it follows that $\vDash_{\mathsf{PI}} \Box\forall x \in \mathbb{N}(\psi^{\mathcal{F}} \to \chi(x)^{\mathcal{F}}) \to (\psi^{\mathcal{F}} \to \Box\forall y \in \mathbb{N}\chi(y))$.

For the inductive step, we only need to consider the rules of modus ponens and universal instantiation, since we are working in a Hilbert-style deductive system. The modus ponens case is trivial and so we focus on universal instantiation.

Assume we have $\varphi_1, \ldots, \varphi_n \vdash \psi(x)$ where x does not occur free in φ_i. We must show $\mathbb{N}(v_0), \ldots, \mathbb{N}(v_m), \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \Box\forall x \in \mathbb{N}\, \psi(x)^{\mathcal{F}}$ where $v_0, \ldots, v_m$ are all the free variables in the formulas. By the induction hypothesis we have that $\mathbb{N}(x), \mathbb{N}(v_0), \ldots, \mathbb{N}(v_m), \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \psi^{\mathcal{F}}(x)$. Using conditional and universal introduction we get $\mathbb{N}(v_0), \ldots, \mathbb{N}(v_m), \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}} \vDash_{\mathsf{PI}} \forall x \in \mathbb{N}\psi^{\mathcal{F}}$. By the definition of $\vDash_{\mathsf{PI}}$ this means that any $\mathcal{M}, w$ such that $\mathcal{M}, w \vDash \mathbb{N}(v_0), \ldots, \mathbb{N}(v_m), \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}}$ also satisfy $\mathcal{M}, w \vDash \forall x \in \mathbb{N}\ \psi^{\mathcal{F}}$.

Let s be such a world, then $\mathcal{M}, s \vDash \mathbb{N}(v_0), \ldots, \mathbb{N}(v_m),\ \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}}$. As $\varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}}$ are all inductive formulas, $\vDash_{\mathsf{PI}} \varphi_i^{\mathcal{F}} \to \Box\varphi_i^{\mathcal{F}}$ and, by Lemma 1.5, $\vDash_{\mathsf{PI}} \mathbb{N}x \to \Box\mathbb{N}x$. It follows that $\mathcal{M}, s \vDash \Box\mathbb{N}(v_0), \ldots, \Box\mathbb{N}(v_m),\ \Box\varphi_0^{\mathcal{F}}, \ldots, \Box\varphi_n^{\mathcal{F}}$. So, let s' be such that $R(s, s')$. It follows that $\mathcal{M}, s' \vDash \mathbb{N}(v_0), \ldots, \mathbb{N}(v_m),\ \varphi_0^{\mathcal{F}}, \ldots, \varphi_n^{\mathcal{F}}$. But for any such world we also have $\mathcal{M}, s' \vDash \forall x \in \mathbb{N}\ \psi^{\mathcal{F}}$. And as s' was arbitrary it follows that $\mathcal{M}, s \vDash \Box\forall x \in \mathbb{N}\ \psi^{\mathcal{F}}$. $\Box$

Recall we defined σ_i recursively for all i as $\sigma_0(x) \equiv (x = 0)$ and $\sigma_{n+1}(x) \equiv \Diamond\exists y \in \mathbb{N}(\sigma_n(y) \wedge Syx)$.

Lemma 1.33. *(i)* $\vDash_{\mathsf{PI}} \sigma_n(\mathbf{a}_n)$, *(ii)* $\vDash_{\mathsf{PI}} \Diamond\exists!\ x\sigma_n(x)$.

Proof. (i) We proceed by induction on n. Clearly $\vDash_{\mathsf{PI}} \mathbf{a}_0 = 0$. Assume $\vDash_{\mathsf{PI}} \sigma_n(\mathbf{a}_n)$. We want to show $\vDash_{\mathsf{PI}} \Diamond\exists y \in \mathbb{N}(\sigma_n(y) \wedge Sy\mathbf{a}_{n+1})$. Let s be a world and let w be accessible from s and have $\mathbf{a}_0, \ldots, \mathbf{a}_n \in I(\mathbb{N}, w)$. We know that $\sigma_n(\mathbf{a}_n)$ holds at w. We also know that $S\mathbf{a}_n\mathbf{a}_{n+1}$ holds at w. It follows that $\exists y \in \mathbb{N}(\sigma_n(y) \wedge Sy\mathbf{a}_{n+1})$ holds at w.

(ii) It follows from (i) that $\vDash_{\mathsf{PI}} \Diamond\exists x\, \sigma_n(x)$. We show $\vDash_{\mathsf{PI}} \sigma_n(x) \to x = \mathbf{a}_n$. Clearly this holds for $n = 0$. Assume $\sigma_n(x)$ at some w for $n > 0$. It follows that there is a world s accessible from w where $S\mathbf{a}_{n-1}x$ holds. So there is a world s' accessible from s, a set $X \subseteq D(s')$ of cardinality $n-1$ and a $u \in D(s)$ not in X such that $x = \#X \cup \{u\} = \mathbf{a}_n$. □

Lemma 1.34. *For every $k \geq 0$ and every unnested formula $\theta(x_1, \ldots, x_k)$ in the signature of PA^1 and every k-tuple of natural numbers $n_1, \ldots, n_k$ one has that :*

$$\mathbb{N} \models \theta(n_1, \ldots, n_k) \quad \Longrightarrow \quad \vDash_{\mathsf{PI}} \forall x_1, \ldots, x_k \in \mathbb{N}(\bigwedge_{i=1}^{k} \sigma_{n_i}(x_i) \to \theta^{\mathcal{F}}(x_1, \ldots, x_k)) \tag{1.29}$$

In the case of $k = 0$, this is to say: for every unnested sentence θ in the signature of PA^1 one has that

$$\mathbb{N} \models \theta \quad \Longrightarrow \quad \vDash_{\mathsf{PI}} \theta^{\mathcal{F}} \tag{1.30}$$

Proof. The argument for (1.19) is by induction on the quantifier complexity of formula. In what follows, the argument is given for $k > 0$, but the argument for $k = 0$ (namely the case of (1.20)) is just a special case of this argument where fewer variables need be introduced.

Let $\mathbb{N} \models \theta(n_1, \ldots, n_k)$, where $\theta(x_1, \ldots, x_k)$ is quantifier free. It follows that $\mathsf{PA}^1 \vdash (\bigwedge_{i=1}^{k} \tau_i(x_i) \to \theta(x_1, \ldots, x_k))$ because Q is correct for quantifier-free sentences (Hájek and Pudlák 1998, Theorem I.1.8 p. 30). And as $(\cdot)^{\mathcal{F}}$ interprets PA^1 it follows that

$$\mathbb{N}(x_1), \ldots, \mathbb{N}(x_k) \vDash_{\mathsf{PI}} \bigwedge_{i=1}^{k} \sigma_{n_i}(x_i) \to \theta^{\mathcal{F}}(x_1, \ldots, x_k)$$

and so $\vDash_{\mathsf{PI}} \forall x_1, \ldots, x_k \in \mathbb{N}(\bigwedge_{i=1}^{k} \sigma_{n_i}(x_i) \to \theta^{\mathcal{F}}(x_1, \ldots, x_k))$.

Now, assume we have the result for $\theta(x_1, \ldots, x_k, y_1, \ldots, y_l)$ and we want to show it for $\forall y_1, \ldots, y_l\ \theta(x_1, \ldots, x_k, y_1, \ldots, y_l)$. We know that if $\mathbb{N} \models \theta(n_1, \ldots, n_k, m_1, \ldots, m_l)$ then

$$\models_{\mathsf{PI}} \forall x_1, \ldots, x_k, y_1, \ldots, y_l \in \mathbb{N}(\bigwedge_{i=1}^{k} \sigma_{n_i}(x_i) \wedge \bigwedge_{j=1}^{l} \sigma_{m_j}(y_j) \rightarrow \theta^{\mathcal{F}}(x_1, \ldots, x_k, y_1, \ldots, y_l)).$$

Assume for a contradiction that $\mathbb{N} \models \forall y_1, \ldots, y_l\ \ \theta(n_1, \ldots, n_k, y_1, \ldots, y_l)$ but

$$\not\models_{\mathsf{PI}} \forall x_1, \ldots, x_k \in \mathbb{N}(\bigwedge_{i=1}^{k} \sigma_{n_i}(x_i) \rightarrow \Box\forall y_1, \ldots, y_l \in \mathbb{N}\ \theta^{\mathcal{F}}(x_1, \ldots, x_k, y_1, \ldots, y_l)).$$

Then there is a world s containing $a_1, \ldots, a_k \in I(\mathbb{N}, s)$ where $\bigwedge_{i=1}^{k} \sigma_{n_i}(a_i)$ holds and a world w accessible from s where there are $b_1, \ldots, b_l \in I(\mathbb{N}, w)$ such that $\theta^{\mathcal{F}}(a_1, \ldots, a_k, b_1, \ldots, b_l)$ is not true at w. But as $b_1, \ldots, b_l \in I(\mathbb{N}, w)$ it follows that each $b_u = \mathbf{a}_{m_u}$ by Lemma 1.4. Then we have that $\sigma_{m_u}(b_u)$ at w. Further, by reductio hypothesis we have that $\mathbb{N} \models \theta(n_1, \ldots, n_k, m_1, \ldots, m_l)$. From which it follows, by induction hypothesis, that at w we have that $\bigwedge_{i=1}^{k} \sigma_{n_i}(a_i) \wedge \bigwedge_{j=1}^{l} \sigma_{m_j}(b_j) \rightarrow \theta^{\mathcal{F}}(a_1, \ldots, a_k, b_1, \ldots, b_l)$ implies $\theta^{\mathcal{F}}(a_1, \ldots, a_k, b_1, \ldots, b_l)$. This contradicts our earlier claim.

Finally, assume we have the result for $\theta(x_1, \ldots, x_k, y_1, \ldots, y_l)$ and we want to show it for $\exists y_1, \ldots, y_l\ \theta(x_1, \ldots, x_k, y_1, \ldots, y_l)$. Now, suppose that $\mathbb{N} \models \exists y_1, \ldots, y_l\ \theta(n_1, \ldots, n_k, y_1, \ldots, y_l)$. Choose an l-tuple of natural numbers $m_1, \ldots, m_l$ such that $\mathbb{N} \models \theta(n_1, \ldots, n_k, m_1, \ldots, m_l)$. The induction hypothesis implies that

$$\models_{\mathsf{PI}} \forall x_1, \ldots, x_k, y_1, \ldots, y_l \in \mathbb{N}(\bigwedge_{i=1}^{k} \sigma_{n_i}(x_i) \wedge \bigwedge_{j=1}^{l} \sigma_{m_j}(y_j) \rightarrow \theta^{\mathcal{F}}(x_1, \ldots, x_k, y_1, \ldots, y_l)).$$

Let s be a world and $a_1, \ldots, a_k \in I(\mathbb{N}, s)$ be such that $\bigwedge_{i=1}^{k} \sigma_{n_i}(a_i)$. Let w be a world accessible from s such that there are $b_1, \ldots, b_l \in I(\mathbb{N}, w)$ and $\bigwedge_{i=1}^{l} \sigma_{m_i}(b_i)$, using again Lemma 1.4 and the definition of the σs. It follows by the induction hypothesis that $\theta^{\mathcal{F}}(a_1, \ldots, a_k, b_1, \ldots, b_l)$

at w and so $\Diamond \exists y_1, \ldots, y_l \in \mathbb{N}\ \theta^{\mathcal{F}}(a_1, \ldots, a_k, y_1, \ldots, y_l)$ at s. Arrow and universal introduction will get our result. $\square$

Chapter 2

The Philosophy Behind Proof-Theoretic Validity

There have been several recent results bringing into focus the super-intuitionistic nature of most notions of proof-theoretic validity. But there has been very little work evaluating the consequences of these results. In this chapter, we explore the question of whether these results undermine the claim that proof-theoretic validity shows us which inferences follow from the meaning of the connectives when defined by their introduction rules. It is argued that the super-intuitionistic inferences are valid due to the correspondence between the treatment of the atomic formulas and more complex formulas.

2.1 Introduction

Proof-theoretic validity was first proposed by Prawitz as an explication of Gentzen's famous observation that the elimination rules of intuitionistic logic appear to follow from the introduction rules.

> The introductions represent, as it were, the 'definitions' of the symbols concerned, and the eliminations are no more, in the final analysis, than the consequences of these definitions. This fact may be expressed as follows: In eliminating a symbol, we may use the formula with whose terminal symbol we are dealing only 'in the sense afforded it by the introduction of that symbol'. (Gentzen 1935, p. 80)

Here we see Gentzen state clearly that the introduction rules for logical connectives should be treated as definitions, while the (intuitionistic) elimination rules should be treated as consequences of the introduction rules. Consequence here cannot mean logical consequence in the traditional model-theoretic sense. And Gentzen reiterates what he means in the second sentence, where he tells us that the elimination rules only allow us to conclude what the introduction rules demonstrated was a part of the meaning of the connective.

An illustration can help precisify intuitions. The rules for implication introduction and elimination are listed below:

$$\dfrac{\begin{array}{c}[A]\\ \vdots \\ B\end{array}}{A \to B} \qquad \dfrac{A \to B \quad A}{B}$$

When we examine the introduction rule for implication, we see that we can introduce an implication when we have a proof that takes A as an assumption and transforms it into B. Gentzen's proposal is that we can think of this as a method for going from a proof of A to a proof of B. When we turn to the elimination rule we see that it allows us to infer from A implying B and A holding that B holds. Because of this we can think of the elimination rule as following from the meaning of implication; 'A implies B' encodes a procedure from A to B and we have A and so via the procedure, it should follow that B.

We can connect this with the idea of a detour and normalization. A detour in a proof is when a connective is introduced and then eliminated. For $\to$ we can remove a detour as follows:

$$\dfrac{\dfrac{\begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \to B} \qquad \begin{matrix}\vdots\\ A\end{matrix}}{B} \qquad \Rightarrow \qquad \begin{matrix}\vdots\\ A\\ \vdots\\ B\end{matrix}$$

A proof normalises when all detours can be eliminated. We can prove normalisation (that all proofs normalise) in intuitionistic logic.[1] Prawitz took normalisation to capture the intuition presented by Gentzen that the elimination rules are only extracting information given by the introduction rules and he used this in developing proof-theoretic validity.

We will present the formal definition of proof-theoretic validity in Section 2.3. For now, however it is sufficient to take an inference to be proof-theoretically valid if it is one of the introduction rules for intuitionistic logic or, roughly, if given proofs of any assumptions, it is unnecessary for proving the conclusion. The second condition means that a proof can be found without that inference. An inference's unnecessariness or eliminability is supposed to demonstrate that it follows from the introduction rules because under the right circumstances any uses of it can be removed in favour of introduction rules.

It is relatively easy to show that intuitionistic logic is proof-theoretically valid. Prawitz (1973, p. 246) conjectured that the reverse was also the case; that only intuitionistic logic was proof-theoretically valid. This conjecture remained open for almost 30 years before negative results started to appear. Piecha and Schroeder-Heister (2019) show that Prawitz's conjecture is false for all the prominent definitions of proof-theoretic validity. Particularly

[1] This is a central and standard result in proof theory. See Chapter 6 of Troelstra and Schwichtenberg 2000 and Chapter 8 of Negri and von Plato 2008

central to this result is the intuitionistically invalid Harrop's rule:

$$\frac{\neg A \to (B \vee C)}{(\neg A \to B) \vee (\neg A \to C)}$$

which when added to intuitionistic logic results in Kreisel-Putnam logic. It turns out that this rule is often proof-theoretically valid.[2]

It is worth highlighting that this result comes as somewhat of a surprise. The truth of Prawitz's conjecture has been stated as seeming obvious by Prawitz:

> It seems obvious that the elimination rules of Gentzen's system are the elimination rules that correspond to his introduction rules. Or, again to put it more carefully: although there are of course weaker elimination rules and even elimination rules that are deductively equivalent with the ones formulated by Gentzen, there are no stronger rules that can be formulated in the language of predicate logic and are justifiable in terms of the introduction rules.(Prawitz 2014, p. 270)

and exceedingly plausible by Dummett:

> It is exceedingly plausible that, on a verificationist meaning-theory, the correct logic will be intuitionistic; and we have noted that the standard introduction rules for 'and', 'or', 'if', and the two quantifiers will validate every intuitionistically valid rule involving these constants, where, by the nature of the case, we need to appeal only to those introduction rules governing the logical constants involved in the general formulation of the rule in question. (Dummett 1991b, p. 270)

That Harrop's rule is proof-theoretically valid is not just unexpected, it threatens to undermine the whole project of proof-theoretic validity. Proof-theoretic validity is supposed

[2]By 'often' here I mean it holds in several prominent presentations of proof-theoretic validity (Piecha, Campos Sanz, and Schroeder-Heister 2015; Piecha and Schroeder-Heister 2019)

to return as valid those inference rules that add nothing on top of the introduction rule but merely follow from them. But Harrop's rule is a poor candidate for following from the introduction rules. In the next section, we will show this by demonstrating that Harrop's rule is not harmonious under several prominent definitions of harmony. It will do no harm for now to think of the introduction and elimination rules for a connective being *harmonious* if the introduction of the connective followed by its elimination does not allow anything new to be proven. The outcome that Harrop's rule is not harmonious arguably suggests that the proof-theoretically valid inferences are not consequences of the intuitionistic introduction rules.

I think this is a serious concern for advocates of proof-theoretic validity. But I believe a defence can be mustered by carefully examining the causes of the non-intuitionistic inferences. In the later sections of this chapter, I argue that the super-intuitionistic inferences present in one of the most prominent notions of proof-theoretic validity stem from the treatment of the atomic formulas and not the treatment of the connectives.[3] That the source of these unwanted inferences is the atomic formulas gives a line of argument in defence of proof-theoretic validity. After all it is not unsurprising that atomic formulas bring with them invalid inferences but what proof-theoretic validity is concerned with is the behaviour of the logical connectives.

The chapter is structured as follows. In the next section I will argue that Harrop's rule is not harmonious as an elimination rule for $\to$. This sets up the main complaint the chapter is directed at answering. With this in place Section 2.3 then sets out the formal definition of proof-theoretic validity that is relevant for our purposes. In section 2.4, we explore in detail a surprising complication in the definition of proof-theoretic semantics caused by the atomic formulas. We discuss there how the inference rules used to define the atomic formulas are equivalent to disjunction-free formulas. We then take a slight detour in sections 2.5 to look at

[3]As a note we will uses super-intuitionistic for logics strictly stronger than intuitionsitic logic.

$$\frac{A \quad B}{A \wedge B} \qquad \frac{A \wedge B}{A} \qquad \frac{A \wedge B}{B} \qquad \frac{A \circ B}{A \Box B}$$

Figure 2.1: The introduction rule for $\wedge$, the two elimination rules for $\wedge$, and a fictional example of a rule that is both an introduction and an elimination rule.

the initial response to proof-theoretic validity being super-intuitionistic, which was to restrict questions to the fragment closed under substitution and argue briefly that this tactic was ill-advised. Then finally in Section 2.6, I outline my argument that the super-intuitionistic inferences follow from the treatment of the atomic formulas.

2.2 Harrop's Rule is not Harmonious

In this section, I will argue that Harrop's Rule does not satisfy two prominent definitions of harmony. Recall that Harrop's rule is:

$$\frac{\neg A \rightarrow (B \vee C)}{(\neg A \rightarrow B) \vee (\neg A \rightarrow C)}$$

If a schematic inference's conclusion has a main connective then it is an *introduction rule* for that connective. If a schematic inference has a main connective in a premise then it is a *elimination rule* for that connective. This definition may be broader than is needed but the intention is to not exclude any examples. By this definition, a rule can be both an introduction and an elimination rule. See Figure 2.1 for examples.

By the above definitions, Harrop's Rule is either an introduction rule for $\vee$ or an elimination rule for $\rightarrow$. However, we will only discuss the case where it is an elimination rule for $\rightarrow$. The reason for this is simple: Harrop's rule is valid according to proof-theoretic validity which, given a set of introduction rules, labels as valid those rules that follow from the introduction rules. As such, the returned rules should be considered as elimination rules.

The term 'harmony' was introduced by Dummett in *The Logical Basis of Metaphysics*. Harmony is a restriction on which proof rules can define a connective and was intended to rule out phenomena like TONK, an example presented by Prior (1960). The connective 'TONK' is a counterexample to the claim that any set of inference rules can define a connective as it is inconsistent in any system with at least one theorem. TONK has the introduction rules for 'or' and the elimination rules for 'and'.

$$\frac{A}{A \text{ TONK } B} \qquad \frac{B}{A \text{ TONK } B} \qquad \frac{A \text{ TONK } B}{A} \qquad \frac{A \text{ TONK } B}{B}$$

The contradiction can be proven as follows:

$$\cfrac{\cfrac{A}{A \text{ TONK } \neg A}}{\neg A}$$

Dummett gives two definitions of harmony. The first defines a connective as harmonious just in case it is conservative over the base theory, an idea first found in Belnap (1962):

> The concept [conservativity] thus adapted offers at least a provisional method of saying more precisely what we understand by 'harmony': namely that there is harmony between the two aspects of the use of any given expression if the language as a whole is, in this adapted sense, a conservative extension of what remains of the language when that expression is subtracted from it.(Dummett 1991b, p. 219)

This definition make harmony relative to a logic or theory. The introduction and elimination rules for a connective are harmonious relative to a system S, on this definition, if there is no sentence of S not containing the connective defined that is provable with the new connective but not without it. Clearly, TONK does not satisfy this definition as in our earlier example $\neg A$ was a counterexample to this form of harmony. Following Steinberger (2011), let us call this *total harmony*.

The question of whether total harmony is undermined in Kreisel-Putnam logic comes down to the question of whether intuitionistic logic with Harrop's rule is conservative over the $\{\vee, \wedge, \bot\}$ fragment of intuitionistic logic. As far as I know this question is open.[4] However, this definition of harmony is not the one that has received the most attention in the literature and Harrop's rule can be shown to not be harmonious on two more prominent definitions.

Dummett give a second definition but this time in terms of the removability of immediate detours (proofs where one introduces then immediately eliminates a connective):

> The analogue, within the restricted domain of logic, for an arbitrary logical constant c, is that it should not be possible, by first applying one of the introduction rules for c, and then immediately drawing a consequence from the conclusion of that introduction rule by means of an elimination rule of which it is the major premiss, to derive from the premisses of the introduction rule a consequence that we could not otherwise have drawn. (Dummett 1991b, pp. 247–8)

An example of such a detour removal was given in the introduction for $\to$. Following Steinberger (2011), we will call this *intrinsic harmony*.

Harrop's rule is not intrinsically harmonious because we cannot eliminate detours when Harrop's rule is treated as an $\to$ elimination rule. The following is what a proof that introduces $\to$ then uses Harrop's rule as $\to$ elimination looks like:

$$\dfrac{\dfrac{\begin{matrix}[\neg A]\\ \vdots \\ B \vee C\end{matrix}}{\neg A \to (B \vee C)}}{(\neg A \to B) \vee (\neg A \to C)}$$

[4] A proof-theoretic validity notion can be provided for the super-intuitionistic inquisitive logic which extends intuitionistic logic with Harrop's rule and double negation elimination for atomic formulas. In inquisitive logic, Harrop's rule is totally harmonious as an introduction rule for $\vee$ against the background system of the $\{\to, \wedge, \bot\}$ fragment of intuitionistic logic. This is shown by Ciardelli and Roelofsen (2011) who shows that every disjunction-free formula of inquisitive logic is provable in intuitionistic logic.

It is possible that in the elided part of the above proof between $\neg A$ and $B \vee C$ we reach $B \vee C$ via something other than $\vee$ introduction, in which case it cannot be transformed into a proof of the conclusion not including the 'detour'. For a rather trivial example of this take the following proof of the intuitionistically unprovable $D \to (B \vee C), \neg A \to D \vdash (\neg A \to B) \vee (\neg A \to C)$.

$$\dfrac{\dfrac{\dfrac{D \to B \vee C \qquad \dfrac{\neg A \qquad \neg A \to D}{D}}{B \vee C}}{\neg A \to (B \vee C)}}{(\neg A \to B) \vee (\neg A \to C)}$$

There is no way to transform this into a proof which removes the 'detour'. This is because any transformation would have to be to a proof in intuitionistic logic (the only other rule is Harrop's and it isn't applicable anywhere else in the proof but the 'detour', though proving this would be an endeavour). And as the conclusion is not provable from the premises in intuitionistic logic we know no such transformation exists.

What we can do is eliminate 'double detours'.

$$\dfrac{\dfrac{\dfrac{\begin{matrix}[\neg A]\\ \vdots \\ C\end{matrix}}{B \vee C}}{\neg A \to (B \vee C)}}{(\neg A \to B) \vee (\neg A \to C)} \quad \Rightarrow \quad \dfrac{\dfrac{\begin{matrix}[\neg A]\\ \vdots \\ C\end{matrix}}{\neg A \to C}}{(\neg A \to B) \vee (\neg A \to C)}$$

But no one has suggested this as a hallmark of harmonious rules. And as the above example shows there are proofs where a double detour cannot be constructed.

We will add one more definition of harmony that appears in the literature. This is called *general elimination harmony*. Given an introduction rule with premises $P_1, \ldots, P_n$ and conclusion C, the corresponding general elimination rule has $n+1$ premises consisting of C and proofs from assumptions $P_1, \ldots, P_n$ to some formula G and conclusion G (Read 2010).

$$\text{Introduction:} \quad \frac{P_1 \quad \ldots \quad P_n}{C} \quad \text{Elimination:} \quad \frac{C \qquad \begin{matrix}[P_1 \ldots P_n] \\ \vdots \\ G\end{matrix}}{G}$$

There is an appropriate modification for rules with assumptions. These elimination rules are modelled after the elimination rules for 'or' and 'exists'. The guiding idea behind this notion of harmony is that the elimination rules only allow you to derive conclusions that the premises of the introduction rules already derive. General elimination harmony has the nice feature of producing an elimination rule for any introduction rule we could give it. And all the elimination rules for intuitionistic logic are equivalent to those generated by general elimination harmony.

Now let us consider whether Harrop's rule is harmonious given general elimination harmony. For this to be the case it would mean that Harrop's rule would need to be equivalent to the general elimination rule for $\rightarrow$. That rule is: $\dfrac{A \rightarrow B \qquad \begin{matrix}\left[\frac{A}{B}\right] \\ \vdots \\ C\end{matrix}}{C}$ where $\left[\frac{A}{B}\right]$ means you can use the inference from A to B 'for free' in the proof of C. This rule is provably equivalent to the normal $\rightarrow$ elimination rule. But it follows from that that as an elimination rule for $\rightarrow$, Harrop's rule is not general elimination harmonious with the introduction rule. This is clear because Harrop's rule is admissible but not derivable in intuitionistic logic but if it was equivalent to this rule it would be due to the equivalence with the usual elimination rule for $\rightarrow$.[5]

[5]What about as an additional introduction rule for $\vee$? This gives us the general elimination rule:

$$\frac{(\neg A \rightarrow B) \vee (\neg A \rightarrow C) \qquad \begin{matrix}\neg A \rightarrow (B \vee C) \\ \vdots \\ D\end{matrix}}{D}$$

This rule is in fact derivable in intuitionistic logic because it can be proven that $\neg A \rightarrow (B \vee C)$ follows from $(\neg A \rightarrow B) \vee (\neg A \rightarrow C)$.

To conclude this section, I have argued that Harrop's rule isn't harmonious as an elimination rule for $\to$ under two prominent definitions of harmony: namely intrinsic harmony and general elimination harmony.[6] This is relevant because recent results have shown that Harrop's rule is proof-theoretically valid under several reasonable constraints. As such it appears proof-theoretic validity and harmony come apart.

Proof-theoretic validity was supposed to provide us with a method of elucidating Gentzen's claim that the elimination rules for intuitionistic logic were consequences of the introduction rules. The validity of super-intuitionistic rules demonstrates that this goal has not been achieved. Still, one might wonder if it wasn't Gentzen who was wrong in thinking that only the intuitionistic elimination rules followed from the introduction rules. The plausibility of such a claim is undermined by the failure of the resulting system to be harmonious. Harmony in its various forms is supposed to show that the elimination rules do not contain more information than the introduction rules. By failing to be harmonious, it is clear that Harrop's rule does not follow from the introduction rules in the sense intended. This raises a real and pressing concern for those who use or advocate for proof-theoretic validity. How can the notion be defended as a useful tool in proof-theoretic semantics if it labels as valid some rules which do not follow from the introduction rules? In fact, the elimination rules appear to be stronger than the introduction rules. We will try and answer this challenge in the next section by highlighting how the super-intuitionistic validities follow from the treatment of atomic formulas. Before that however we will introduce the definition of proof-theoretic validity.

[6] Whether it is totally harmonious isn't yet known.

2.3 Proof-Theoretic Validity Defined

As we discussed in the introduction, proof-theoretic validity is an attempt to formally explicate what Gentzen meant when he claimed that the elimination rules are consequences of the introduction rules. Proof-theoretic validity takes normalisation and the reductions which remove intuitionistic elimination rules as central. The goal is to give a definition of when a proof is valid based on it being reducible, in some sense, to the introduction rules.

While proof-theoretic validity is supposed to be a general notion, we will here restrict attention to the connectives 'and', 'or', and 'if…then' with negation treated as $A \to \bot$. Notably missing from this list are the quantifiers 'all' and 'some'. This restriction is made because proof-theoretic validity is going to be implausible in general if it cannot be made to work in this specific case. In particular there is no point moving on to look at the complications brought by quantifiers if the connectives alone cannot be handled. However, we should have a healthy suspicion towards any results generalising easily to other setting.

To give this definition we will need to introduce the term *argument* for a potential proof, which is something that has the correct shape of a proof: it has a single conclusion, and these follow from premises in a step-by-step way.[7] Proof-theoretic validity is then a property of arguments, and those arguments which satisfy the condition of being proof-theoretically valid will be called proofs. Prawitz's definition has four conditions depending on the state of the argument. The full definition will be provided at the end of this section (Definition 2.1). To understand the cases, we need to define what it means to say an argument is open or closed. An argument is closed if every top line of the argument is either an axiom or is an assumption that has been discharged. Arguments are open if they have an undischarged assumption.

[7]Formally an argument is a tree labelled by formulas and a discharge relation.

The simplest case in which an argument is valid is simply when it consists entirely of introduction rules. The introduction rules are supposed to be the definition of their main connectives and a proof is supposed to be valid when it follows from the introduction rules. So, there can be no problems with the use of an introduction rule. When an introduction rule is used in an argument containing other rules then the application of the introduction rule will be valid, and the validity of the whole argument will depend on the validity of the rest of the argument. We can capture this idea in the following condition:

Closed introduction case. If $\mathcal{D}$ is a closed argument ending in an introduction rule, then it is valid iff the arguments for the premises of the introduction rule are valid.

Note that we restrict the condition to closed arguments. This is because it is easier to have one condition for all open arguments, which we will return to shortly.

With this simple case out of the way we have three remaining situations to deal with. These are: closed arguments that end in rules other than the introduction rules, arguments that end in atomic formulas, and arguments that are open. We will first look at the treatment of closed proofs which do not end in introduction rules.

When a closed argument ends in a rule that isn't an introduction rule we want it to be valid when it can be shown to, in some sense, follow from the introduction rules. Prawitz's insight was that if a proof only contains intuitionistic introductions or elimination rules then any closed proof ending in an elimination rule can be transformed via normalisation into a closed proof which ends in an introduction rule. And in this way, one can slowly remove the elimination rules. This gives an initial condition for closed non-introduction proofs:

Preliminary closed non-introduction case. If $\mathcal{D}$ is a closed argument which does not end in an introduction rule, then it is valid iff *it can be transformed by the reductions used in the proof of normalisation into* a valid argument with the same conclusion which does end in an introduction rule.

Prawitz defends this definition by claiming that a proof with a detour and the proof that results from the elimination of the detour are syntactically distinct but semantically the same proof (Prawitz 1971, p. 257; Prawitz 1973, p. 234). The thought is that if two arguments are the same proof and one of them is valid (because of the use of introduction rules) then the other one must also be valid. He backs up this assertion by appealing to the Curry-Howard isomorphism and the use of identity between the algorithms on either side of beta-reduction.

This method and justification, however, only work when one is considering proofs in intuitionistic logic, not arbitrary arguments. And we must consider arbitrary arguments otherwise we beg the question in favour of intuitionistic logic. To resolve this problem, we need to generalise the concept of normalisation to arbitrary arguments. Prawitz does this by arguing that there must be a computable transformation on an argument to a valid argument with the same conclusion which ends in an introduction rule (Prawitz 1973). Schroeder-Heister (2006, pp. 552–3) does away with the requirement that the transformation is computable. However, in both cases, no reason is given to think that the transformed argument is semantically the same argument as the original. This gives us the following condition. It does not deal with transformations on proofs, but it can be shown that the definition is equivalent to one that does, given a sufficiently broad understanding of transformations. Not including reference to transformations is therefore a technical convenience. Thus, what we end up with is the following condition:

Closed non-introduction case. If $\mathcal{D}$ is a closed argument which does not end in an introduction rule, then it is valid iff *there is* a valid argument with the same conclusion which does end in an introduction rule.

With these two conditions on closed proofs in place we can move on to arguments with open assumptions.

How arguments with open assumptions are treated demonstrates a key commitment embedded in the idea that a connective can be defined by a proof rule. When a sentence with a logic connective is asserted it is taken to be a claim that there is a proof of the sentence. An assumption is then an assertion with the assertoric force cancelled. The reasoning that follows from an assumption is hypothetical on the statement being asserted. As such the reasoning that follows is valid if it would be valid were there a valid proof of the assumptions. It is this fundamental idea that motivates the treatment of open assumptions; to repeat for an argument with open assumptions to be valid the reasoning has to be valid if we had proofs of the assumptions. It will not do to check one proof, any closed valid proofs substituted for assumptions must result in a valid argument. This gives us the following, preliminary, condition:

Preliminary open case. If $\mathcal{D}$ is an open argument of A with open assumptions $A_0, \ldots, A_n$ it is valid iff for all closed valid arguments $\mathcal{D}_0, \ldots, \mathcal{D}_n$ of $A_0, \ldots, A_n$, the following argument is valid:

$$\begin{array}{ccc} \mathcal{D}_0 & \ldots & \mathcal{D}_n \\ A_0 & \ldots & A_n \\ & \mathcal{D} & \\ & A & \end{array}$$

However, our treatment of proofs with open assumptions is currently incomplete. The reason for this is that as it stands, we have not explained how to deal with a proof such as the following:

Example 2.1. $\dfrac{p \to q \qquad p}{q}$

This is because there are no closed valid proofs of the atomic formula p as things have been set up so far.[8] None of the intuitionistic introduction rules ends in an atomic formula and so no proof of p ends in one of the intuitionistic introduction rules. Note that p as an arbitrary propositional variable here stands for any atomic sentence. In model-theoretic semantics,

[8]We will use lowercase letters for atomic formulas and uppercase letters for arbitrary formulas. It is important in this setting to distinguish the two.

arbitrary propositions are dealt with by considering the different meanings they could have, which are spelt out in terms of truth values. Here we are taking terms to be defined by their proof rules and so, rather in parallel with model-theoretic semantics, we are going to consider different assignments of proof rules to the atomic sentences. So, we relativize validity to a set S of rules for the atomic formulas. This can be understood as something like a theory which constrains the use and hence interpretation of the propositional letters. Then we require what is valid to be relativized to a particular meaning for the atomic sentences.

Before we revise our definition for open arguments, we can now give the condition for atomic formulas:

Atomic case. If $\mathcal{D}$ is a closed argument ending in an atomic formula then it is S-valid iff it contains only rules in S.

So when we have an atomic formula, we can restrict our attention to arguments that do not contain the logical connectives. This gives us the base case for proofs by induction on the definition of proof-theoretic validity and we will say a lot more about what counts as an atomic rule in the next section.

We took this detour initially because we didn't know what to do when we had an open assumption of an atomic formula. We see now that we will need to substitute it for a valid proof of the formula, and we know what such a valid proof looks like by the above condition. As it is relative to a set of atomic rules, we will need to make all our other conditions relative to a set of atomic rules by replacing validity with S-validity in the definitions. But for open assumptions we are going to make one other change.

Imagine we are dealing with Example 2.1 still and are considering S-validity where S is a set of atomic rules from which p is not provable. In this case our preliminary open case is trivially satisfied because there are no valid arguments for p. But this isn't what we want to happen, our reasoning is hypothetical on there being a proof for p and we can certainly imagine ways to prove p.[9] To get around this we will consider not just our initial set of atomic rules S but also any extensions of it. This way there should be at valid arguments for any atomic formula. This gives us the condition for proofs with open assumptions:

Open case. If $\mathcal{D}$ is an open argument of A with open assumptions $A_0, \dots, A_n$ it is S-valid iff for all S' which are acceptable extensions of S and all closed S'-valid arguments $\mathcal{D}_0, \dots, \mathcal{D}_n$ of $A_0, \dots, A_n$, the following argument is S'-valid:

$$\begin{array}{ccc} \mathcal{D}_0 & \dots & \mathcal{D}_n \\ A_0 & \dots & A_n \\ & \mathcal{D} & \\ & A & \end{array}$$

We will now bring all the conditions together.

Definition 2.1. *(Prawitz 1973, p. 236; Schroeder-Heister 2006, pp. 543–4) An argument $\mathcal{D}$ is an S-*valid derivation* for a set of rules S describing the behaviour of the atomic formulas if one of the following conditions holds:*

Atomic case *If $\mathcal{D}$ is a closed argument ending in an atomic formula then it is S-valid iff it contains only rules in S.*

Closed introduction case *If $\mathcal{D}$ is a closed argument ending in an introduction rule then it is S-valid iff the arguments for the premises of the introduction rule are S-valid.*

Closed non-introductory case *If $\mathcal{D}$ is a closed argument which does not end in an introduction rule then it is S-valid iff there is a S-valid argument with the same conclusion which does end in an introduction rule.*

Open case *If $\mathcal{D}$ is an open argument of A with open assumptions $A_0, \dots, A_n$ it is S-valid iff for all S' which are acceptable extensions of S and all closed S'-valid arguments $\mathcal{D}_0, \dots, \mathcal{D}_n$ of $A_0, \dots, A_n$, the following argument is S'-valid:*

[9] There is a variant of this notion where things are still relative to a set of atomic rules S but no extensions are considered. This is now Prawitz's preferred notion (Prawitz 2014). Both this and the displayed definition are super-intuitionistic.

$$\begin{array}{ccc} \mathcal{D}_0 & \dots & \mathcal{D}_n \\ A_0 & \dots & A_n \\ & \mathcal{D} & \\ & A & \end{array}$$

A rule of inference is then *proof-theoretically valid* if it is valid on all acceptable sets of atomic rules.

This presentation of the definition of proof-theoretic validity has highlighted the important role that the treatment of atomic inferences plays in the account. As we have seen, how it handles inferences between atomics is just as crucial to it as Tarski's treatment of atomics is to the traditional model-theoretic account of logical consequence. By contrast, in some popular discussions of proof-theoretic validity, the project is portrayed as being excessively focused on analytic inferences:

> Similar distortions can be observed in the study of logical constants. Both in proof theory and in less formal investigations of the epistemology of logic, the focus has been too much on 'analytic' rules of inference, even when those get restricted to a weak fragment of the logic we actually and successfully use. To achieve a more faithful understanding of the cognitive aspects of our ordinary practice of using the logical constants, we need to stop concentrating on 'analytic' rules. (Williamson 2020, p. 120)

To be sure, proof-theoretic validity is aiming at something like a notion of analyticity, insofar as this gives one a notion of 'following from' distinct from the traditional model-theoretic notion of entailment. However, it is precisely by concentrating on non-analytic inferences between atomics that we are able to forge this definition. And, as we will see in the next section, it is the treatment of atomics which leads proof-theoretic validity into the super-intuitionistic.

2.4 Atomic Rules and Proof-Theoretic Systems

In the previous section I laid out the definition of proof-theoretic validity (or at least a variant of it), however, I did not say much about what the set of atomic rules used in the definition was to be like. The result of a growing body of formal work is that we must pay careful attention to the role of the atomic formulas in proof-theoretic validity.

To discuss this in detail we need to first discuss what an atomic rule is. The easiest rules to understand are those that look like axioms. That is atomic rules we can prove in one step. They can be written down as follows: $\overline{p}$. However, we can also have rules which look like the proof rules for the connectives but containing only atomic rules. The following example contrasts an atomic rule with $\wedge$ introduction.

Example 2.2.

$$\frac{A \quad B}{A \wedge B} \qquad \frac{p \quad q}{r}$$

Similarly, but slightly more complicatedly we could have a rule which discharges assumptions like $\rightarrow$ introduction or $\vee$ elimination.

Example 2.3.

$$\textbf{a.}\quad \frac{\begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \rightarrow B} \qquad \frac{\begin{matrix}[p]\\ \vdots \\ q\end{matrix}}{r} \qquad\qquad \textbf{b.}\quad \frac{A \vee B \quad \begin{matrix}[A]\\ \vdots \\ C\end{matrix} \quad \begin{matrix}[B]\\ \vdots \\ C\end{matrix}}{C} \qquad \frac{p \quad \begin{matrix}[q]\\ \vdots \\ s\end{matrix} \quad \begin{matrix}[r]\\ \vdots \\ s\end{matrix}}{s}$$

So, we might take the sets of atomic rules that we consider in the definition of proof-theoretic validity to be any collection of rules of the three types (atomic, premise-to-conclusion, and with assumptions) we have just discussed.

However, this would give us a logic so strange it may not deserve that name. To see this, we need to note that there is a relationship between these odd atomic rules and formulas. First note we can treat the inference from the premises to the conclusion as an implication $\to$ and we can simply conjoin the premises. So $\dfrac{p \quad q}{r}$ can be transformed into $(p \wedge q) \to r$. The following two proofs show that the formula and the rule are interchangeable in proofs. The parts that correspond to the rule have been highlighted in bold to make the proofs easier to read.

$$\dfrac{\dfrac{\dfrac{[p \wedge q]}{\mathbf{p}} \qquad \dfrac{[p \wedge q]}{\mathbf{q}}}{\mathbf{r}}}{(p \wedge q) \to r} \qquad\qquad \dfrac{\dfrac{\mathbf{p} \quad \mathbf{q}}{p \wedge q} \qquad (p \wedge q) \to r}{\mathbf{r}}$$

That is the first proof shows that the formula can be proven using the rule and the second proof shows that the rule can be replaced by the formula in a proof. It can be proven that any rule from atomic premises to atomic conclusion is equivalent to a formula in just this way (Piecha, Campos Sanz, and Schroeder-Heister 2015).

And this can also be done when we discharge assumptions. We just need to treat the proof from the assumption to its premise as an implication also. So, Example 2.3.a is interchangeable with $(p \to q) \to r$ and Example 2.3.b is interchangeable with $(p \wedge (q \to s) \wedge (r \to s)) \to s$. Using Example 2.3.b, we can prove this interchangeability as follows:[10]

$$\dfrac{\dfrac{\dfrac{[p \wedge (q \to s) \wedge (r \to s)]}{\mathbf{p}} \qquad \dfrac{[\mathbf{q}] \qquad \dfrac{[p \wedge (q \to s) \wedge (r \to s)]}{q \to s}}{\mathbf{s}} \qquad \dfrac{[\mathbf{r}] \qquad \dfrac{[p \wedge (q \to s) \wedge (r \to s)]}{r \to s}}{\mathbf{s}}}{\mathbf{s}}}{(p \wedge (q \to s) \wedge (r \to s)) \to s}$$

$$\dfrac{\dfrac{\mathbf{p} \qquad \dfrac{\begin{matrix}[\mathbf{q}] \\ \vdots \\ \mathbf{s}\end{matrix}}{q \to s} \qquad \dfrac{\begin{matrix}[\mathbf{r}] \\ \vdots \\ \mathbf{s}\end{matrix}}{r \to s}}{p \wedge (q \to s) \wedge (r \to s)} \qquad (p \wedge (q \to s) \wedge (r \to s)) \to s}{\mathbf{s}}$$

[10]Note that I am using a slight generalisation of $\wedge$ introduction and elimination. This is only to make the proofs more perspicuous and is easily removed at no cost.

So, if we use these three different types of rule, some formulas, but not all formulas, are interchangeable with rules. Because of this, if we take a system which includes only atomic rules of these types, an odd thing will happen. We will have an intuitionistically invalid rule called the generalised Harrop's rule but only for a limited collection of formulas:

$$\frac{A \to (B \vee C)}{(A \to B) \vee (A \to C)} \qquad \text{(Generalised Harrop)}$$

In particular, we can only substitute disjunction-free formulas for A. And the proof of this result appeals directly to this interchangeability (see Piecha, Campos Sanz, and Schroeder-Heister 2015).

Let's look at why Generalised Harrop's is valid via an example. We will make use of the equivalence with formulas. Take

$$\frac{p \to (q \vee r)}{(p \to q) \vee (p \to r)}$$

It turns out that asking whether this inference is S-valid is the same thing, by the open case of Definition 2.1, as asking whether for all S', extending S, when $p \to (q \vee r)$ is S' valid so is $(p \to q) \vee (p \to r)$. And that in turn, via the equivalence between rules and formulas, can be answered by answering whether, when $q \vee r$ is $S' \cup \{\bar{p}\}$-valid, $(p \to q) \vee (p \to r)$ is S'-valid. But if $q \vee r$ is $S' \cup \{\bar{p}\}$-valid, then by a combination of the closed introduction case and the closed non-introduction case, there is a $S' \cup \{\bar{p}\}$-valid argument for $q \vee r$ that ends in $\vee$-introduction. And so, there is either a $S' \cup \{\bar{p}\}$-valid argument for q or for r. Assume q is $S' \cup \{\bar{p}\}$-valid then again, by the equivalence between rules and formulas, it follows that $p \to q$ is S'-valid, and so by the closed introduction case, so is $(p \to q) \vee (p \to r)$.

It is worth noting at this point that this means we are dealing with a peculiar type of logic. In particular we are dealing with logics that are not closed under uniform substitution. This means that we need to withhold a common instinct to read sentences with atomic formulas like $(p \wedge q) \rightarrow r$ as really another way of writing $(A \wedge B) \rightarrow C$. When an atomic letter like p is used, that atomic letter is really all that is meant. When a schematic letter A is used, the asserted larger expression, e.g. $(A \wedge B) \rightarrow C$, is intended to hold regardless of what one substitutes in for A.

Having generalised Harrop's rule hold for a rather circumscribed and arbitrary collection of sentences isn't ideal. However, this issue can be partially resolved by expanding the equivalence between formulas and rules. We saw that $(p \rightarrow q) \rightarrow r$ corresponds to the rule $\frac{\overset{[p]}{q}}{r}$. What then would correspond to a formula such as $((p \rightarrow q) \rightarrow r) \rightarrow s$? Schroeder-Heister (1984) proposed a generalisation of inference rules that allows us to find a corresponding rule for this formula. But it requires an expansion of what can be discharged. Note that in the formula corresponding to the discharge of an assumption, we had implication in two places: the first standing in for the proof from the assumption to the premise and the second standing for the inference from premise to conclusion. To add an extra implication, we need to add another of these. We will do this by assuming not an atomic formula but a rule. This idea may seem strange but there is no technical impediment. The resulting rule is:

$$\dfrac{\begin{array}{c}\left[\begin{array}{c}\vdots \\ \dfrac{p}{q} \\ \vdots\end{array}\right] \\ r\end{array}}{s}$$

where the square brackets represent the discharge of the assumed rule.

Let's look quickly at how we would use a rule like this in a proof. I will write rules linearly with the assumption discharged in square brackets and the inference from premises to conclusion represented by $/$. So, the above rule would be $[p/q]r/s$. Now consider how we would prove r with the two rules $[p/q]q/r$, and $/p$. To illustrate the discharge of rules it will help to write each step A, D where A is atomic and D is the set of atomic rules used so far and not discharged.

Example 2.4.

$$\dfrac{\dfrac{\dfrac{}{p, \{/p\}}}{q, \{/p, \boldsymbol{p/q}\}}}{r, \{/p, [\boldsymbol{p/q}]q/r\}}$$

This can be generalised even further. We can discharge rules which discharge assumptions themselves and in doing so get more equivalences between formulas and rules. As a second example, take the following proof of t with rule $[/p, [p/q]q/r]s/t$ and rule r/s. Note that the first three lines are identical to the proof above.

Example 2.5.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{p, \{/p\}}}{q, \{/p, p/q\}}}{r, \{/p, [p/q]q/r\}}}{s, \{\boldsymbol{/p, [p/q]q/r}, r/s\}}}{t, \{[\boldsymbol{/p, [p/q]q/r}]s/t, r/s\}}$$

If we allow rules that discharge rules of any level of complexity then every disjunction-free formula is equivalent to a rule (Piecha, Campos Sanz, and Schroeder-Heister 2015). This means that we have generalised Harrop's formula holding for all disjunction-free formulas in the antecedent. It will be shown later in Chapter 3 that the resulting logic is just intuitionistic logic with the generalised Harrop's rule restricted to disjunction-free formulas. This logic is far more uniform than the logic that results when we only consider atomic rules that do not discharge rules, but this logic is still not closed under uniform substitution as we cannot substitute a formula with a disjunction for an atomic formula in the valid instance $\frac{p \to (q \vee r)}{(p \to q) \vee (p \to r)}$. We will discuss this issue in Section 2.5. One might ask if we cannot just force this property, but this is not possible due to every valid proof of a disjunction requiring a valid proof of one of the disjuncts.

The particular kind of formulas that we cannot generate rules equivalent to have disjunction in the consequence of an implication. So, for a simple example take $p \to (q \vee r)$. We might try initially to make this equivalent to p/q because we can prove $p \to (q \vee r)$ given this rule:

$$\dfrac{\dfrac{\dfrac{[\mathbf{p}]}{\mathbf{q}}}{q \vee r}}{p \to (q \vee r)}$$

But we cannot go in the other direction. It is not possible given $p \to (q \vee r)$ to replicate the behaviour of p/q. This should be clear because $p \to (q \vee r)$ will be true when p and r are true and q is false. Having now seen the impact of different choices of atomic rules we can be more careful in our presentation of proof-theoretic validity.

As we discussed above, if we allow atomic rules to discharge other rules we end up with Harrop's rule being valid, if we do not we end up with Harrop's rule for only a restricted collection of formulas. The moral I want to take from this is that we need to add something to our definition of proof-theoretic validity. Namely, we need to restrict it to an acceptable collection of sets of atomic formulas. So, one system might be 'any set of atomic rules, which may discharge other rules', and another might be 'any set of atomic rules, which do not discharge anything'. We can write these using set theoretic notation as $\{S \mid S \text{ is a set of atomic rules}\}$ and $\{S \mid S \text{ is a set of atomic rules without assumptions}\}$. Let's call these the *complete system* and the *minimal system* and let's call any set of sets of atomic rules a *proof-theoretic system*. We can have more esoteric systems as well, such as, $\{S \mid S \text{ is in the complete system and } \bar{p} \notin S\}$ or even $\{\varnothing, \{\bar{p}\}, \{\bar{q}\}\}$. Each of these will give us a different logic or theory which is proof-theoretically valid.

2.5 Failure of Uniform Substitution

Before moving on to discuss my solution to the worry in section 2.2, I want to pause briefly to discuss the initial reaction to results showing that proof-theoretic validity notions were super-intuitionistic. As has been pointed out, most systems of proof-theoretic validity are not closed under substitution. And most examples of rules and axioms that were proof-theoretically valid but super-intuitionistic which were found were not closed under substitution. This led to the initial suggestion that what was at issue was not any problem with the definition of proof-theoretic validity but rather that one needed to take the substitution closure of the validities to find the logic of proof-theoretic validity. The substitution closure is gotten by removing any inferences that aren't closed under substitution (such as general Harrop's rule).[11]

> That a notion of validity is not closed under substitution is, of course, a highly significant result in itself, but a result which rather demonstrates that such a notion is not even a candidate for completeness. Therefore, a thorough discussion of completeness or incompleteness of intuitionistic logic should at least consider a concept of validity closed under substitution. (Piecha, Campos Sanz, and Schroeder-Heister 2015, p. 322)

Now as a matter of fact, this restriction will not help in general as even when closed under substitution prominent notions of proof-theoretic validity will still be super-intuitionistic (as we will see in Chapter 3). But there is a notion of proof-theoretic validity offered by Goldfarb (2016) which, while super-intuitionistic, results in intuitionistic logic when closed under substitution. As such it is worth considering whether failure to be closed under

[11]The reason rules are removed rather than added to ensure substitution closure is because firstly from a technically perspective adding rules is more likely to give an inconsistent system and secondly, it is assumed that a rule not closed under substitution hasn't been shown to be justified by proof-theoretic validity while all the rules that are have been.

substitution really discounts a system as suggested by the above quote and so whether the restriction to the validities that are closed under substitution is fair.

It is certainly the case that as logics are standardly conceived they are closed under substitution. This is often simply stipulated in the definition of what counts as a logic. But it is fair to ask what reason one might have for enforcing this requirement. Take the following quote from Tarski:

> Consider any class K of sentences and a sentence X which follows from the sentences of this class. [...] Moreover, since we are concerned here with the concept of logical, i.e. *formal*, consequence, and thus with a relation which is to be uniquely determined by the form of the sentences between which it folds, this relation cannot be influenced in any way by empirical knowledge, and in particular by knowledge of the objects to which the sentence X or the sentences of the class K refer. The consequence relation cannot be affected by replacing the designations of the objects referred to in these sentences by the designations of any other objects. (Tarski 1956, pp. 414–5)

Here we see Tarski presenting what has become the not an uncommon view that the form of a sentence is what makes something *logically* valid.[12] But the idea that logically valid inferences are those that are valid in virtue of their form, or without regards to content is a view embedded in the model-theoretic conception of logical validity. On that picture logical validity means preserving truth in all models. This is in sharp contrast to what proof-theoretic validity takes logical validity to be. On this picture, as we have seen, logical

[12]The view has antecedents in Bolzano's notion of analyticity. See for instance:

> But suppose that there is just a single idea in it which can be arbitrarily varied without disturbing its truth or falsity, i.e. if all the propositions produced by substituting for this idea any other idea we pleased are either true altogether or false altogether, presupposing that they have a denotation. [...] I permit myself, then, to call propositions of this kind, borrowing an expression from Kant, *analytic*. (Bolzano 1973, §148 p. 192)

validity means following from the introduction rules for the logical connectives which act as definitions. There is no mention of form or disregard of content in this view of logical validity. And it isn't clear that the advocate of proof-theoretic validity has any reason to expect logics to be closed under substitution.

There are other notions of uniform substitution that may be better suited to proof-theoretic validity. For example, Humberstone offers the following:

> [...] closure under uniform substitution of propositional variables (rather than arbitrary formulas) for propositional variables. (Humberstone 2011, p. 188)

On this picture it is only the non-logical content that needs to be substituted. Let us call this notion *propositional uniform substitution*. Most notions of proof-theoretic validity are going to be closed under propositional uniform substitution. So even if no rules are allowed that discharge assumptions, it is still the case that every combination of atomic letters as premises and conclusion are considered in the minimal system and every possible combination of such rules is an allowed extension. Because of this there will be no difference of behaviour between atomic formula and we can substitute one atomic letter for another. The same is true on the complete system and basically every other system that has been seriously considered. However, it is possible to come up with proof-theoretic systems that lack this property.

I take the following to show that the early attempts to save proof-theoretic validity from super-intuitionistic validities by restricting to the substitution closed fragment were not just technically unhelpful but also philosophically ill motivated. From this we would like to point out that the below suggestion is not equivalent to taking the substitution closure of proof-theoretic validity.

2.6 The Role of the Atomic Formulas Explained

In this section I am going to argue that the super-intuitionistic axioms that are valid follow from the treatment of the atomic formulas not the treatment of the connectives and as such should not be considered to undermine the argument, partially supported by proof-theoretic validity, that the intuitionistic elimination rule are privileged.

To start, let's illustrate how a classical axiom, which is valid on certain proof-theoretic systems, is not valid on others. We will use the following from an earlier example:

$$\frac{p \to (q \vee r)}{(p \to q) \vee (p \to r)}$$

as a simple example. We will show that it is not valid on the proof-theoretic system $\{\varnothing, \{/p, p/q\}, \{/p, p/r\}\}$. This is in some sense a toy system, but it will do nicely for the purposes of illustration.

To see that the inference is not $\varnothing$-valid on this system it first needs to be shown that:

$$\frac{p}{q \vee r}$$

is $\varnothing$-valid. This will follow, by the condition for open cases, if for every extension S of $\varnothing$ and for every proof of p that is S-valid the result of appending the proof of p to $\frac{p}{q \vee r}$ is S-valid. Now as it happens in our small system we only have one proof of p.

$$\overline{p}$$

This proof is both $\{/p, p/q\}$ and $\{/p, p/r\}$-valid by the condition for the atomic case. So, it needs to be shown that

$$\frac{\overline{p}}{q \vee r}$$

is both $\{/p, p/q\}$ and $\{/p, p/r\}$-valid. But this follows in the first case because

$$\cfrac{\cfrac{\overline{p}}{\overline{q}}}{q \vee r}$$

is $\{/p, p/q\}$-valid, by the condition for the closed introduction case, and in the second case because

$$\cfrac{\cfrac{\overline{p}}{\overline{r}}}{q \vee r}$$

is likewise $\{/p, p/r\}$-valid.

It then follows that

$$\cfrac{\cfrac{\cfrac{[p]}{q \vee r}}{p \to (q \vee r)}}{(p \to q) \vee (p \to r)}$$

is a counter example to the $\varnothing$-validity of the inference. This is because for it to be $\varnothing$-valid there must be, by the condition for the closed non-introduction case, a $\varnothing$-valid proof that ends in the introduction of $\vee$ and that requires that there is a $\varnothing$-valid proof of either $p \to q$ or $p \to r$. But neither of these are possible because one of the systems extending $\varnothing$ lets you go from p to q but not to r and vice-versa for the other one.

This example is illustrative of a general factor in the validity of Harrop's rule. That feature is that the invalidity of Harrop's rule is connected to the validity of one step inferences, like $\frac{p}{q \vee r}$ in the example.

And similarly, the validity of Harrop's rule is connected to the invalidity of one step inferences. Let's keep the example above and see how we can make $\frac{p}{q \vee r}$ invalid by changing how we treat the atomic formulas. Let's take the previous system $\{\varnothing, \{/p, p/q\}, \{/p, p/r\}\}$ and add $\{/p\}$ giving the system $\{\varnothing, \{/p\}, \{/p, p/q\}, \{/p, p/r\}\}$. To test the $\varnothing$-validity of $\frac{p}{q \vee r}$ in this system we need to check if every extension of $\varnothing$ which has a valid proof of p has a valid proof of $q \vee r$ due to the condition for open arguments. There is only one thing that has changed from our earlier discussion and that is that $\overline{p}$ is $\{/p\}$-valid. But we cannot give $\{/p\}$-valid proof of q or r by the condition for atomic conclusions. It follows from this that $\frac{p}{q \vee r}$ is not $\varnothing$-valid on this proof-theoretic system. To see that Harrop's rule now holds on this system it is sufficient to note that the inference from p to $q \vee r$ is only $\{/p, p/q\}$ and $\{/p, p/r\}$-valid and that on the first of these sets $p \rightarrow q$ is valid and on the second $p \rightarrow r$ is.

We are not adding additional types of rules in this example. But what happens when we do is very similar. Without the rules some one step inference from a formula to a disjunction will be valid, from A to $B \vee C$. And, by adding an atomic rule which is equivalent to A, this one step inferences become invalid. This is because there is now an additional set of atomic rules which has a valid proof of A (because there is a new rule that is equivalent to it) but not of either B or C. I take this to show that the validity of generalised Harrop's rule follows not from the treatment of the connectives but rather from the correspondence between atomic rules and formulas and so from the treatment of the atomic formulas. I will now consider a few objections to this point.

Now one might naturally think that while this may be an explanation of why Harrop's rule and its generalisation holds there may be many other super intuitionistic axioms that hold. However, as we will show later in Chapter 3, the complete system is axiomatized by intuitionistic logic plus generalised Harrop's rule.

Now one might think to solve this we should just remove the equivalence between formulas and rules to get a system in which the treatment of the atomic formulas does not have this impact. This will not work for two reasons. The first is that even with the simplest of rules for atomic formulas, the axioms, there is still an equivalence between the axiom $/p$ and the formula p. And a system with no atomic rules would have no validities.[13] The second is

[13]One could change the treatment of open formulas as well and there are versions of proof-theoretic validity that behave like this. They have not been more successful in capturing intuitionistic logic however.

that it isn't this particular treatment of the atomic formulas that causes the problem but rather any treatment. It has been shown that for all common notions the resulting logic will be super-intuitionistic (Piecha and Schroeder-Heister 2019). For example, if we take the minimal system, that is the one in which no atomic rules can discharge assumptions, it turns out that $\neg\neg p \to p$ is valid.[14]

At this point the reader might complain that, regardless of how the super-intuitionistic validities may appear to follow from the atomic rules, if every proof-theoretic system has its own super-intuitionistic validities then they cannot all be attributed to the treatment of the atomic formulas. However, in response to this criticism I would like to bring up a parallel with Tarskian model-theoretic semantics. In Tarskian model-theoretic semantics for classical logic, the classical validities are those which hold in all models. But there is no one model in which all and only the classical validities hold. In fact, despite no atomic formula p or it's negation $\neg p$ being a classical validity, every classical model proves either p or $\neg p$ for all atomic p. Does this mean that these super-classical validities do not follow from the treatment of the atomic formulas? Clearly not, it is to be expected that every model with have many super-classical validities. I put forward that proof-theoretic semantics can appeal to a similar point. While every proof-theoretic system has some super-intuitionistic validities, they are still a result of the treatment of the atomic formulas not the treatment of the connectives.

2.7 Conclusion

Let's reiterate the argument of the chapter. Proof-theoretic validity is supposed to be a formal method for finding those inferences that follow from the intuitionistic introduction rules. It aims to explicate Gentzen's claim that the intuitionistic elimination rules are consequences

[14]Remember that these logics are not closed under substitutions so this does not imply that $\neg\neg A \to A$ is valid which would mean the logic was classical.

of the intuitionistic introduction rules and it was conjectured by Prawitz and Dummett that no intuitionistically invalid inference were proof-theoretically valid. Instead, it has been shown that super-intuitionistic inferences are proof-theoretically valid, and these inferences are not in harmony with the intuitionistic introduction rule or closed under substitution. This looks like a bad situation for proof-theoretic validity as it seems tenuous to claim that inharmonious inferences follow from the introduction rules.

However, careful examination of the definition of proof-theoretic validity and the proof of the super-intuitionistic validities showed the vital role played by the treatment of the atomic formulas. In particular whether an instance of super-intuitionistic rule was valid or not depended on whether there was an atomic rule corresponding to the formulas it contained. On this basis it was argued that it was the treatment of the atomic formulas that was leading to the super-intuitionistic inferences being valid.

Chapter 3

The Logic of Proof-Theoretic Validity

Prawitz (1971) conjectured that proof-theoretic validity offers a semantics for intuitionistic logic. This conjecture has recently been proven false by Piecha and Schroeder-Heister (2019). This article resolves one of the questions left open by this recent result by showing the extensional alignment of proof-theoretic validity and general inquisitive logic. General inquisitive logic is a generalisation of inquisitive semantics, a uniform semantics for questions and assertions. The chapter further defines a notion of quasi-proof-theoretic validity by restricting proof-theoretic validity to allow double negation elimination for atomic formulas and proves the extensional alignment of quasi-proof-theoretic validity and inquisitive logic.

3.1 Introduction

Proof-Theoretic Validity was proposed by Prawitz (1971) as an explication of Gentzen's famous claim that the introduction rules can be viewed as definitions of the connectives.[1] Prawitz's formal definition of proof-theoretic validity (henceforth PTV) is complex but,

[1]For the quotation from Gentzen, see the beginning of Section 3.4.2

roughly, a proof is valid if it stands in the correct relationship to the introduction rules (see Definition 3.17). Prawitz conjectured in the early 1970s that:

Conjecture 3.1 (Prawitz)**.** *PTV aligns extensionally with the validities of intuitionistic logic.*

This conjecture remained open for many years but was recently disproven:

Theorem 3.1 (Piecha and Schroeder-Heister 2019, Corollary 3.9)**.** *PTV is a proper superset of the validities of intuitionistic logic.*

For those still sympathetic to Gentzen and Prawitz, this leaves open the question of what super-intuitionistic logic defines the same set of validities as PTV, and hence how far from true Prawitz's conjecture was.

Let quasi-PTV be PTV with double-negation for atomic formulas. For the formal definition of PTV see Definition 3.17 and for the formal definition of quasi-PTV see Definition 3.23. The main result of this chapter is:

Theorem 3.2. *PTV aligns extensionally with the validities of general inquisitive logic and quasi-PTV aligns extensionally with the validities of inquisitive logic.*

Inquisitive logic has been studied extensively in recent years, see for example Ciardelli and Roelofsen (2011), Ciardelli, Groenendijk, and Roelofsen (2018), and Punčochář (2016). It arises naturally out of an effort to capture the idea that propositions have both informative and inquisitive content. Inquisitive semantics unifies the picture of propositions as sets of possible worlds, with the picture of questions as sets of answers. It does this by, roughly, treating assertions as questions with only one answer.

The main result of this chapter is important because it shows which logic would be justified by proof-theoretic validity. Further, quasi-PTV is as classical as one can make PTV since inquisitive logic is the maximal weak intermediate logic with the disjunction property and double negation holding for atomic formulas (Ciardelli and Roelofsen 2011, p. 18). So, adding more instances of double negation elimination will either not change the logic or the resulting system will be stronger than classical logic.

One, of course, would want to know more about the connection between PTV and inquisitive logic discovered here, and whether anything in the underlying motivations for these logics may be responsible for this. There are two things worth mentioning in this connection. First, one traditional semantics for intuitionistic logic was Kolmogorov's problem interpretation. This interpretation takes formulas as stating problems in need of a solution. For example, Kolmogorov (1932, p. 329) states that $a \vee b$ means there is a way to "solve at least one of the problems a and b". While Kolmogorov had a broad range of problems in mind, including those that involve constructions, we can identify many problems with the questions they are supposed to answer. When considered this way we can think of the Kolmogorov interpretation in terms of questions. For example, $a \vee b$ might be "Can you answer the question b or a?" This offers a connection between constructive proof and the semantics of questions. Second, Dummett's *The Logical Basis of Metaphysics* is a defence of Prawitz's idea, and a key component of Dummett's philosophy of language was that knowing a sentence's meaning involved knowing how to use the sentence (Dummett 1991b, p. 103). In particular, Dummett thought of this as involving the ability to recognise when a statement has been verified. Perhaps we can see a verification as an answer to a query about the truth of the statement. This is not the place to explore these connections, but this should suffice to dissipate the concern that the extensional alignment of quasi-PTV and inquisitive logic is a mere mathematical accident.

This chapter is organised as follows. Since neither inquisitive logic nor PTV are closed under uniform substitution, in Section 3.2 we discuss weak or non-structural logics which are not necessarily closed under substitution. Inquisitive logic and generalised inquisitive logic are discussed in more detail in Section 3.3. We define PTV both as Prawitz's originally did on derivations and as a consequence relation and discuss the relationship between the two in Section 3.4. In Section 3.5 we discuss how to modify PTV by changing how one deals with atomic formulas. We prove several results about how modified systems relate to PTV and each other. We define quasi-PTV in Section 3.6. And it is there that our main theorems are proven and the consequences of these results for PTV are drawn out.

3.2 Logics Without Closure Under Substitution

Before we explore PTV and inquisitive semantics, we need to modify the definition of logics to include logics which aren't closed under substitution. Logics which aren't closed under substitution are called weak or non-structural (we will use weak). After defining weak logics, we will focus in on those weak logics which have the disjunction property (Definition 3.3). The chapter then considers characterisations of weak logics found in Ciardelli and Roelofsen (2011) and Punčochář (2016). In later sections, this will allow us to show the extensional identity of quasi-PTV and inquisitive semantics and of PTV and generalised inquisitive semantics.

3.2.1 Weak Logics

The set of propositional formulas is built from the set of atomic formulas $\{p_i \mid i \in \mathbb{N}\} \cup \{\bot\}$ and the recursive application of the connectives $\wedge, \vee, \rightarrow$. Let a *logic* L be a subset of the propositional formulas closed under modus ponens and substitution. Note that we take $\bot$ to be an atomic proposition. Let a *weak logic* L be a set of propositional formulas closed under modus ponens, which need not be closed under substitution.

We define (weak) intermediate logics as follows:

Definition 3.1. *A (weak) logic L is a (weak) intermediate logic if $IPC \subset L \subset CPC$.*

Here we take IPC to be the deductive closure of the axioms of the intuitionistic predicate calculus, and similarly, CPC is the deductive closure of the axioms of the classical predicate calculus. While weak intermediate logics aren't necessarily closed under substitution, because they have IPC as a sublogic they contain all substitution instances of IPC.

Definition 3.2. *Let L be a (weak) intermediate logic then:*

$$\vdash_L \varphi \Leftrightarrow \varphi \in L$$

$$\Gamma \vdash_L \varphi \Leftrightarrow \text{ there are } \psi_1, \ldots, \psi_n \in \Gamma, (\psi_1 \wedge \cdots \wedge \psi_n) \to \varphi \in L$$

Notice that we have essentially defined the consequence relation so that the deduction theorem holds. We can see that the relation is well behaved in other ways as the following properties hold:

Lemma 3.1. *Let L be a (weak) intermediate logic.*

$$\nvdash_L \bot, \qquad \text{(Falsum Property)}$$

$$\vdash_L \varphi \wedge \psi \Longleftrightarrow \vdash_L \varphi \text{ and } \vdash_L \psi, \qquad \text{(Conjunction Property)}$$

$$\vdash_L \psi \to \varphi \Longleftrightarrow \psi \vdash_L \varphi, \qquad \text{(Weak Deduction Theorem)}$$

$$\varphi \vdash_L \varphi, \qquad \text{(Reflexivity)}$$

$$\Gamma \vdash_L \varphi \text{ and } \varphi \vdash_L \psi \Longrightarrow \Gamma \vdash_L \psi, \qquad \text{(Transitivity)}$$

$$\Gamma \vdash_L \varphi \Longleftrightarrow \exists \varphi_0, \ldots, \varphi_n \in \Gamma,\ \varphi_0, \ldots, \varphi_n \vdash_L \varphi. \qquad \text{(Compactness)}$$

Above we have a condition for $\bot$, $\wedge$, and $\to$ but no condition for $\vee$. The natural one is the following:

Definition 3.3. *A logic L has the disjunction property if*

$$\vdash_L \varphi \vee \psi \quad \Longleftrightarrow \quad \vdash_L \varphi \text{ or } \vdash_L \psi.$$

However, while this property holds in IPC, it does not hold in CPC and there are weak intermediate logics both with and without it.

3.2.2 Equality Between Weak Logics with the Disjunction Property

From now on we are interested in logics which do have the disjunction property. We will see that given two weak logics with the disjunction property, if both logics agree that every formula is equivalent to one which is the disjunction of negated formulas, then they are equal and similarly if they agree on the disjunction free formulas and every formula is equivalent to one which is the disjunction of disjunction free formulas. If a formula is a disjunction of negated formulas (i.e. $\neg\varphi_1 \vee \cdots \vee \neg\varphi_n$) then we say it is in *disjunctive negation form.* If it is a disjunction of disjunction free formulas (i.e. $\varphi_1 \vee \cdots \vee \varphi_n$ with φ_i disjunction free for all i) then we say it is in *disjunctive form.* We will give a translation into disjunctive and disjunctive negation form below in Definition 3.4 and 3.5. But first we will generalise results of Ciardelli and Roelofsen (2011, Theorem 3.2.36) and Punčochář (2016, Theorem 4).

Let $DF(L)$ be all disjunction free formulas of L.

Theorem 3.3. *Suppose L_1, L_2 are weak intermediate logics such that they both have the disjunction property. Assume that both logics satisfy the same condition, either:*

(3.3.1.) for all φ there is a ψ in disjunctive negation form *such that $\psi \equiv \varphi \in L_1, L_2$, or*

(3.3.2.) for all φ there is a ψ in disjunctive form *such that $\psi \equiv \varphi \in L_1, L_2$ and $DF(L_1) = DF(L_2)$.*

Then $L_1 = L_2$.

Proof. We prove the first case first. Without loss of generality assume $\varphi \in L_1$. There is a ψ in disjunctive negation form such that $\psi \equiv \varphi \in L_1, L_2$ and so $\psi \in L_1$. Let ψ be

$\neg\varphi_1 \vee \cdots \vee \neg\varphi_n$. As L_1 has the disjunction property there is an $i \leq n$ such that $\neg\varphi_i \in L_1$. Because L_1 is a sublogic of CPC it follows that $\neg\varphi_i$ is a tautology of classical logic and so by Glivenko's theorem $\neg\neg\neg\varphi_i$ is a tautology of IPC and so $\neg\varphi_i$ is a tautology of IPC. It follows then that $\neg\varphi_1 \vee \cdots \vee \neg\varphi_n$ is a tautology of IPC by disjunction introduction. And as IPC is a sublogic of L_2 it follows that $\neg\varphi_1 \vee \cdots \vee \neg\varphi_n \in L_2$ and so $\varphi \in L_2$ as $\psi \equiv \varphi \in L_2$. So L_1 is a sublogic of L_2. And as the same reasoning goes through with L_1 switched with L_2 it follows that the two logics must be equal.

The proof of the second is very similar. The difference is that one assumes ψ is $\varphi_1 \vee \cdots \vee \varphi_n$ where φ_i is disjunction free for all i. And rather than going through IPC one uses the assumption that $DF(L_1) = DF(L_2)$, to get that there is a $\varphi_i \in L_2$ for some i. □

This theorem gives us two ways to characterise weak intermediate logic with the disjunction property.

We now want to define when a logic is such that every formula is equivalent to one in disjunctive or disjunctive negation form. It will be helpful to have a particular transformation of formulas into disjunctive and disjunctive negation form. The following definition of a disjunctive negation translation, which we call DNT, is a variation on Maksimova (1986) employed by Ciardelli and Roelofsen (2011). The definition of disjunctive translation is the obvious modification.

Definition 3.4. *Let the disjunctive translation DT be as follows:*

$$DT(p) = p \tag{3.1}$$

$$DT(\varphi \vee \psi) = DT(\varphi) \vee DT(\psi) \tag{3.2}$$

$$DT(\varphi \wedge \psi) = \bigvee\{\varphi_i \wedge \psi_j \mid 0 < i \leq n, 0 < j \leq m\} \tag{3.3}$$

$$\textit{where } DT(\varphi) = \varphi_1 \vee \cdots \vee \varphi_n \textit{ and } DT(\psi) = \psi_1 \vee \cdots \vee \psi_m$$

$$DT(\varphi \to \psi) = \bigvee\{\bigwedge_{0<j\leq n} (\varphi_j \to \psi_{i_j}) \mid (i_1, \ldots, i_n) \in \{1, \ldots, m\}^n\} \tag{3.4}$$

$$\textit{where } DT(\varphi) = \varphi_1 \vee \cdots \vee \varphi_n \textit{ and } DT(\psi) = \psi_1 \vee \cdots \vee \psi_m$$

Definition 3.5. *Let the disjunctive negation translation DNT be as follows:*

$$DNT(p) = \neg\neg p \tag{3.5}$$

$$DNT(\varphi \vee \psi) = DNT(\varphi) \vee DNT(\psi) \tag{3.6}$$

$$DNT(\varphi \wedge \psi) = \bigvee\{\neg(\varphi_i \vee \psi_j) \mid 0 < i \leq n, 0 < j \leq m\} \tag{3.7}$$

$$\textit{where } DNT(\varphi) = \neg\varphi_1 \vee \cdots \vee \neg\varphi_n \textit{ and } DNT(\psi) = \neg\psi_1 \vee \cdots \vee \neg\psi_m$$

$$DNT(\varphi \to \psi) = \bigvee\{\neg\neg \bigwedge_{0<j\leq n} (\psi_{i_j} \to \varphi_j) \mid (i_1, \ldots, i_n) \in \{1, \ldots, m\}^n\} \tag{3.8}$$

$$\textit{where } DNT(\varphi) = \neg\varphi_1 \vee \cdots \vee \neg\varphi_n \textit{ and } DNT(\psi) = \neg\psi_1 \vee \cdots \vee \neg\psi_m$$

The definition of DNT looks confusing, particularly for $\wedge$ and $\to$. Some examples will hopefully illuminate the translation. The $\wedge$ translation says that there is some φ_i and ψ_i such that $\neg(\varphi_i \vee \psi_i)$ is true. So $(\neg\varphi_1 \vee \neg\varphi_2) \wedge (\neg\psi_1 \vee \neg\psi_2)$ becomes

$$\neg(\varphi_1 \vee \psi_1) \vee \neg(\varphi_1 \vee \psi_2) \vee \neg(\varphi_2 \vee \psi_1) \vee \neg(\varphi_2 \vee \psi_2).$$

The DNT translation for $\to$ is the disjunction of all the ways the φ_i's could be implied by some of the ψ_j's with a double negation on the front. For example, the DNT translation of

$(\neg\varphi_1 \vee \neg\varphi_2) \to (\neg\psi_1 \vee \neg\psi_2)$ is

$$\neg\neg((\psi_1 \to \varphi_1) \wedge (\psi_1 \to \varphi_2)) \vee \neg\neg((\psi_1 \to \varphi_1) \wedge (\psi_2 \to \varphi_2))$$
$$\vee\neg\neg((\psi_2 \to \varphi_1) \wedge (\psi_1 \to \varphi_2)) \vee \neg\neg((\psi_2 \to \varphi_1) \wedge (\psi_2 \to \varphi_2))$$

Now that we have particular translations in mind, we can think about what properties a logic would have to have to prove $\varphi \leftrightarrow DT(\varphi)$ or $\varphi \leftrightarrow DNT(\varphi)$. We are interested in what properties it would have to have on top of IPC. In both cases, we need $\to$ to commute over $\vee$, for disjunction free antecedents in the case of DT and for negated antecedents in the case of DNT.

In the case of negated antecedents this condition is the widely studied Kreisel-Putnam axiom[2]:

Definition 3.6.

$$(\neg\varphi \to \psi \vee \chi) \to [(\neg\varphi \to \psi) \vee (\neg\varphi \to \chi)] \tag{KP}$$

For disjunction free antecedents we offer the natural generalisation of the Kreisel-Putnam axiom:

Definition 3.7. *For all disjunction free φ,*

$$(\varphi \to \psi \vee \chi) \to [(\varphi \to \psi) \vee (\varphi \to \chi)] \tag{GKP}$$

Note that, even though we are working with weak logics, we will only be concerned with logics which contain every substitution instance of the standard or generalised Kreisel-Putnam axiom. The converse of both the generalised and standard Kreisel-Putnam axiom is provable

[2]The inference rule that goes from the antecedent to the consequence of the Kreisel-Putnam axiom is called Harrop's rule. And it is this name that is often used in work on proof-theoretic validity.

in IPC, so we can replace the main $\to$ with a $\leftrightarrow$. If one simply adds the Kreisel-Putnam axiom to IPC the result is Kreisel-Putnam logic KP. If one adds the generalised Kreisel-Putnam axiom to intuitionistic logic we will write $IPC + GKP$.[3] Note that in IPC, $\neg\varphi$ is equivalent to some formula φ^* not containing disjunction (Kleene 1952, Sec. 26-7) so the generalised Kreisel-Putnam axiom implies the Kreisel-Putnam axiom.

For DNT we also need the equivalence of p and $\neg\neg p$. While $IPC \vdash p \to \neg\neg p$, we also have $IPC \nvdash \neg\neg p \to p$. So, we are going to need a logic which has $\neg\neg p \to p$ for all atomic formulas. In fact, in the presence of double negation elimination for atomic formulas, KP implies GKP (Punčochář 2016).[4]

We can show that any logic extending $IPC + GKP$ proves the equivalence of every formula with its DT translation and if it also contains double negation elimination for atomic formulas it proves the equivalence of every formula with its DNT translation.

Lemma 3.2. *If L is a weak intermediate logic with the disjunction property and the generalised Kreisel-Putnam axiom then:*

(3.2.1.) for all φ, we have that $\vdash_L \varphi \leftrightarrow DT(\varphi)$,

(3.2.2.) if $\vdash_L \neg\neg p \to p$ for all atomic p, then for all φ, we have $\vdash_L \varphi \leftrightarrow DNT(\varphi)$.

Hence, by the previous remark about the relation between KP and GKP, one has that (3.2.2.) holds for any weak intermediate logic with the disjunction property and the Kreisel-Putnam axiom.

[3] This logic is equivalent to one where GKP restricts φ to the Harrop formulas (roughly formulas where the right most implication always has a disjunction free formula in the consequent). In this second guise it has been studied by Punčochář 2016 and Miglioli et al. 1989.

[4] Note that using DNT it is easy to see (via Lemma 3.2) that every disjunction free formula is equivalent to one starting with a negation *if* atomic formulas are.

Proof. First, we prove the equivalence with DT. The base case is trivial, as is the disjunction case. The conjunctive case simply uses the distributive properties of conjunction and disjunction provable in IPC.

For the case of implication, note that $\varphi_1 \vee \cdots \vee \varphi_n \to \psi$ is equivalent to $(\varphi_1 \to \psi) \wedge \cdots \wedge (\varphi_n \to \psi)$ and as φ_i for all i is disjunction free, $\varphi_i \to \psi_1 \vee \cdots \vee \psi_m$ is equivalent to $(\varphi_i \to \psi_1) \vee \cdots \vee (\varphi_i \to \psi_m)$ by application of GKP. This means that $\varphi_1 \vee \cdots \vee \varphi_n \to \psi_1 \vee \cdots \vee \psi_m$ is equivalent to $[(\varphi_1 \to \psi_1) \vee \cdots \vee (\varphi_1 \to \psi_m)] \wedge \cdots \wedge [(\varphi_n \to \psi_1) \vee \cdots \vee (\varphi_n \to \psi_m)]$. And by the distributive properties of IPC this is equivalent to $\bigvee\{\bigwedge_{0<j\leq n}(\varphi_j \to \psi_{i_j}) \mid (i_1, \ldots, i_n) \in \{1, \ldots, m\}^n\}$.

We turn now to the second case. Note that by hypothesis $\vdash_L \neg\neg p \to p$ so $DNT(p) \equiv_L p$. The disjunction case is trivial.

To show the conjunction case in the right to left direction holds in IPC we first assume $\bigvee\{\neg(\varphi_i \vee \psi_j) \mid 0 < i \leq n, 0 < j \leq m\}$ and then applying $\vee$ elimination as follows. Note that $\neg(\varphi_i \vee \psi_j)$ is equivalent in IPC to $(\neg\varphi_i \wedge \neg\psi_j)$. So, by disjunction introduction on $\neg\varphi_i$ and $\neg\psi_j$, it follows that $\neg\varphi_1 \vee \cdots \vee \neg\varphi_n$ and $\neg\psi_1 \vee \cdots \vee \neg\psi_m$. From which it follows by the induction hypothesis that $\vdash_L \varphi \wedge \psi$.

For the other direction we again work in L, using the induction hypothesis and basic facts about IPC. Assume $\varphi \wedge \psi$. By the induction hypothesis $(\neg\varphi_1 \vee \cdots \vee \neg\varphi_n) \wedge (\neg\psi_1 \vee \cdots \vee \neg\psi_m)$. So, by using the distribution properties in IPC we get $\bigvee\{(\neg\varphi_i \wedge \neg\psi_j) \mid 0 < i \leq n, 0 < j \leq m\}$. But $\neg\varphi_i \wedge \neg\psi_j$ is equivalent to $\neg(\varphi_i \vee \psi_j)$ in IPC. So, we get we get $\bigvee\{\neg(\varphi_i \vee \psi_j) \mid 0 < i \leq n, 0 < j \leq m\}$.

This leaves $\to$. This follows from the induction hypothesis and the following derivation in GKP.

$$\vdash_{GKP} (\neg\varphi_1 \vee \cdots \vee \neg\varphi_n) \to (\neg\psi_1 \vee \cdots \vee \neg\psi_m)$$

Which is equivalent to its DT translation:

$$\vdash_{GKP} \bigvee\{ \bigwedge_{0<j\leq n} (\neg\varphi_j \to \neg\psi_{i_j}) \mid (i_1, \ldots, i_n) \in \{1, \ldots, m\}^n\}$$

(This just says that one of all the possible ways the $\neg\varphi_i$'s implies the $\neg\psi_j$'s must be true.) if and only if (because in IPC we have $(\neg A \to \neg B) \leftrightarrow (\neg\neg B \to \neg\neg A)$)

$$\vdash_{GKP} \bigvee\{ \bigwedge_{0<j\leq n} (\neg\neg\psi_{i_j} \to \neg\neg\varphi_j) \mid (i_1, \ldots, i_n) \in \{1, \ldots, m\}^n\}$$

if and only if (because in IPC double negation commutes over implication and conjunction)

$$\vdash_{GKP} \bigvee\{\neg\neg \bigwedge_{0<j\leq n} (\psi_{i_j} \to \varphi_j) \mid (i_1, \ldots, i_n) \in \{1, \ldots, m\}^n\}. \quad \square$$

This result allows us to generalise Theorem 3.3.

Corollary 3.1. *Given two weak intermediate logics L_1, L_2 such that one of the following conditions is satisfied by both logics:*

(3.1.1.) disjunction property and all formulas equivalent to their DNT translation,

(3.1.2.) disjunction property, the Kreisel-Putnam axiom and $\neg\neg p \to p$,

(3.1.3.) disjunction property, the generalised Kreisel-Putnam axiom and $\neg\neg p \to p$.

It follows that $L_1 = L_2$.

Corollary 3.2. *Further, given two weak intermediate logics L_1, L_2 such that one of the following conditions is satisfied by both logics:*

(3.2.1.) disjunction property, $DF(L_1) = DF(L_2)$, and all formulas equivalent to their DT translation,

(3.2.2.) disjunction property, $DF(L_1) = DF(L_2)$*, and the generalised Kreisel-Putnam axiom.*

It follows that $L_1 = L_2$*.*

3.3 Inquisitive Logic

Inquisitive semantics is a formal semantics designed to offer a uniform treatment of assertions and questions. This is motivated by observations such as the mutual embedding of sentences and questions. For example:

Xiao asked if Anna is here.	(embedded question)	(3.9)
Who told you that Anna is here?	(embedded assertion)	(3.10)
Xiao asked me who told you that Anna is here.	(two-level embedding)	(3.11)

Further motivations include the use of logical connectives in both questions and assertions, that answers to questions are interpreted with contextual information given in the question, and that propositional attitudes can have questions as their objects (Ciardelli, Groenendijk, and Roelofsen 2018, Ch. 1).

Traditionally assertions have been modelled by sets of possible worlds (Stalnaker 1976) while questions are treated as the set of possible answers to the question (Karttunen and Peters 1980; Groenendijk and Stokhof 1984). As answers to questions are assertions, this means that a question is a set of assertions. So, questions can be treated as sets of sets of possible worlds. These different treatments rule out a uniform treatment of “Is Anna here?" and “Anna is here." The solution proposed by inquisitive semantics is to treat every assertion similarly to the traditional treatment of questions.

In inquisitive semantics, propositions are treated as sets of sets of possible worlds closed under subsets. We will now set this out more precisely. First, we need a collection of possible worlds W and using this we can define an information state.

Definition 3.8. *An* information state $s \subseteq W$ *is a set of possible worlds.*

Note that in the more traditional setting an information state would be a proposition. In that setting, you can think of a proposition as having more information the fewer worlds it contains. This is because you can think of a proposition as containing all the information the worlds have in common and fewer worlds mean more shared information. It follows that if a proposition p is a subset of another proposition q, then p has more information. This justifies calling a set of worlds that is a subset of another an enhancement of it.

Definition 3.9. *A state t is an* enhancement *of s if $t \subseteq s$.*

In the setting of inquisitive logic, a proposition is then defined as a set of information states, but not any sets only those closed under enhancements.

Definition 3.10. *A* proposition *is a non-empty set P of information states which is closed downwards under enhancements.*

A proposition must be non-empty because containing no information is associated with the proposition that is always false, which contains one set namely the empty one. Being closed under enhancements means that an inquisitive proposition is a set which contains every traditional proposition which implies any other traditional proposition in the set.

A proposition in inquisitive logic is an assertion if its union is a member of it. What this amounts to is that it can be treated as a set of possible worlds because it has a least informative set in it. More formally if P is a set of possible worlds, $\mathscr{P}(P) = \{Q : Q \subseteq P\}$ is its translation into inquisitive semantics. Propositions that don't have this property represent questions (here they are called *inquisitive*). They contain at least two distinct maximal sets of worlds which can be thought of as representing distinct answers to a question. Formally:

Definition 3.11. *P is* inquisitive *iff* $\bigcup P \notin P$.

The language of propositional logic can be interpreted in this setup.

Definition 3.12. *Let $V : ATOM \to \mathscr{P}(W)$ be a valuation function on atomic propositions, then define $[\ \cdot\ ]_{\langle W,V \rangle} : PROP \to \mathscr{P}(\mathscr{P}(W))$ as a function from the language of propositional logic to sets of information states meeting the following conditions:*

$$[p]_{\langle W,V \rangle} = \{s \mid s \subseteq V(p)\} \tag{3.12}$$

$$[\varphi \wedge \psi]_{\langle W,V \rangle} = [\varphi]_{\langle W,V \rangle} \cap [\psi]_{\langle W,V \rangle} \tag{3.13}$$

$$[\varphi \vee \psi]_{\langle W,V \rangle} = [\varphi]_{\langle W,V \rangle} \cup [\psi]_{\langle W,V \rangle} \tag{3.14}$$

$$[\varphi \to \psi]_{\langle W,V \rangle} = \{s \mid \forall t \subseteq s\ (t \in [\varphi]_{\langle W,V \rangle} \Rightarrow t \in [\psi]_{\langle W,V \rangle})\} \tag{3.15}$$

Here $\bot$ is not true at any world, so that $[\bot]_{\langle W,V \rangle} = \{\emptyset\}$.[5] *We can prove by an easy induction that for all φ the set $[\varphi]_{\langle W,V \rangle}$ will be a proposition. When W and V are clear from context, they will be omitted and we will write $[\varphi]$ instead of $[\varphi]_{\langle W,V \rangle}$.*

A cursory examination will show only formulas containing $\vee$ are inquisitive. What explanation is there for this identification of disjunctions with questions? Ciardelli, Groenendijk, and Roelofsen (2018, pp. 73–4) appeal to cross-linguistic evidence that the same 'words' are used for questions and disjunctions. For example, in Japanese the particle 'ka' is used at the end of a sentence to signal a question ('Anna wa kita-**ka**' Did Anna come?) and attached to each noun to signal a disjunction ('Anna-**ka** Xiao-**ka**' Anna or Xiao). It has been proposed that inquisitive semantics can account for this data (Szabolesi 2015). Reflection on the role of disjunction might also suggest a relation to questions. A disjunction involves in

[5] It is also worth noting that the definition of $\to$ is equivalent to the more traditional definition of $sup\{[\chi]_{\langle W,V \rangle} \mid [\chi \wedge \varphi]_{\langle W,V \rangle} \subseteq [\psi]_{\langle W,V \rangle}\}$ from Heyting algebras (Troelstra and Dalen 1988, ch. 13).

some sense a loss or lack of information–you do not know which of the two disjuncts are true. Questions similarly involve a lack of information. While this might be thought to be a pragmatic feature of disjunction, inquisitive semantics accounts for it in the semantics.

With $[\cdot]_{\langle W,V\rangle}$ defined we can define inquisitive logic as follows:

Definition 3.13. *$\varphi \in L_{Inq}$ if and only if for all W, V one has $[\varphi]_{\langle W,V\rangle} = \mathscr{P}(W)$.*

It also turns out that negated formulas are uninquisitive.

Lemma 3.3. *Given a W, for all φ we have that $\bigcup[\neg\varphi] \in [\neg\varphi]$. In other words, $\neg\varphi$ is uninquisitive.*

Proof. We will show that $\bigcup\{s \mid \forall t \in [\varphi]\ t \cap s = \varnothing\} \in \{s \mid \forall t \in [\varphi]\ t \cap s = \varnothing\}$. Assume otherwise. Then there would be a $t \in [\varphi]$ such that $t \cap \bigcup\{s \mid \forall t \in [\varphi]\ t \cap s = \varnothing\} \neq \varnothing$ but then there would have to be a $s \in \{s \mid \forall t \in [\varphi]\ t \cap s = \varnothing\}$ such that $t \cap s \neq \varnothing$ but this is a contradiction. □

It follows that inquisitive logic has double negation elimination for atomic formulas.

Lemma 3.4 (Ciardelli and Roelofsen 2011, Rmk 3.8, p. 10)**.** *$\neg\neg p \to p$ holds in inquisitive logic*

Proof. Atomic propositions aren't inquisitive, and this result holds for all uninquisitive formulas. It follows from the fact that φ is not inquisitive then $[\neg\neg\varphi] = [\varphi]$, which we will now prove. Note that

$$[\neg\neg\varphi] = \{s \mid \forall t[(\forall v \in [\varphi](t \cap v = \varnothing)) \Rightarrow t \cap s = \varnothing]\}. \tag{3.16}$$

Clearly $[\varphi] \subseteq [\neg\neg\varphi]$. To show the other direction we will show that $\bigcup[\neg\neg\varphi] = \bigcup[\varphi]$. Assume for a contradiction that $w \in \bigcup[\neg\neg\varphi]$ and assume $w \notin \bigcup[\varphi]$. It follows that for all $v \in [\varphi]$ we have $\{w\} \cap v = \varnothing$ but $\{w\} = \{w\} \cap \bigcup[\neg\neg\varphi]$. From this it follows by equation 3.16

that $\bigcup[\neg\neg\varphi] \notin [\neg\neg\varphi]$. But as $[\neg\neg\varphi]$ isn't inquisitive by the Lemma above and so contains $\bigcup[\neg\neg\varphi]$ this is a contradiction. As φ isn't inquisitive, it has $\bigcup[\varphi]$ as its maximal element. As such the maximal elements of both sets are the same and they are both closed downwards, so they are the same set. □

Note that inquisitive logic is a weak logic as $\neg\neg\varphi \to \varphi$ does not hold for inquisitive propositions. For example, $[\neg\neg(\varphi\vee\psi)]$ is the downwards closure of $\bigcup[\varphi\vee\psi]$ while $[\varphi\vee\psi] = [\varphi]\cup[\psi]$. Inquisitive logic also satisfies the axioms of IPC (Ciardelli and Roelofsen 2011, Prop 3.19, p. 14), the disjunction property (Ciardelli and Roelofsen 2011, Prop 3.9, p. 10), and the Kreisel-Putnam axiom (Ciardelli and Roelofsen 2011, Rmk 3.8, p. 10). As such it is characterised by the second condition of Corollary 3.1.

There is another way of rendering the validities of inquisitive semantics by defining a support relation between states and formulas. We will here present a generalised version of the semantics found in Punčochář 2016 where, instead of taking the set of all states, we can restrict to a subset $I \subseteq \mathscr{P}(\mathscr{P}(W))$.

Definition 3.14. *Given a set of worlds W, a valuation function V, an $I \subseteq \mathscr{P}(\mathscr{P}(W))$, and a state s in I, we say s supports φ in $\langle W, I, V\rangle$ or $s \vDash^{Inq}_{\langle W,I,V\rangle} \varphi$ as follows:*

(3.14.1.) $s \vDash^{Inq}_{\langle W,I,V\rangle} p$ *iff* $s \subseteq V(p)$,

(3.14.2.) $s \vDash^{Inq}_{\langle W,I,V\rangle} \bot$ *iff* $s = \varnothing$,

(3.14.3.) $s \vDash^{Inq}_{\langle W,I,V\rangle} \varphi \wedge \psi$ *iff* $s \vDash^{Inq}_{\langle W,I,V\rangle} \varphi$ *and* $s \vDash^{Inq}_{\langle W,I,V\rangle} \psi$,

(3.14.4.) $s \vDash^{Inq}_{\langle W,I,V\rangle} \varphi \vee \psi$ *iff* $s \vDash^{Inq}_{\langle W,I,V\rangle} \varphi$ *or* $s \vDash^{Inq}_{\langle W,I,V\rangle} \psi$,

(3.14.5.) $s \vDash^{Inq}_{\langle W,I,V\rangle} \varphi \to \psi$ *iff for all* $t \subseteq s$ *if* $t \in I$ *and* $t \vDash^{Inq}_{\langle W,I,V\rangle} \varphi$ *then* $t \vDash^{Inq}_{\langle W,I,V\rangle} \psi$.

As pointed out by Punčochář (2016, p. 410), this new definition connects to Definition 3.12 via the following result:

Proposition 3.1. *$s \vDash^{Inq}_{\langle W,\mathscr{P}(W),V\rangle} \varphi$ if and only if whenever $s \subseteq W$ it follows that $s \in [\varphi]_{\langle W,V\rangle}$. So $\varphi \in L_{Inq}$ if and only if for all W and V and all states $s \subseteq \mathscr{P}(W)$, it holds that $s \vDash^{Inq}_{\langle W,\mathscr{P}(W),V\rangle} \varphi$.*

We can use this more general framework to get a general inquisitive semantics which does not have double negation elimination for atomic formulas. From Punčochář (2016, p. 412) we get the following equivalence with Kripke models:

Lemma 3.5 (Punčochář 2016, p. 412)**.** *Given $\langle W, I, V\rangle$, the Kripke semantics $\langle I - \{\varnothing\}, \supseteq, V^*\rangle$ where $V^*(p) = \{s \in I - \{\varnothing\} \mid s \subseteq V(p)\}$ defines the same satisfaction relation.*

It follows from this that any general inquisitive semantics will satisfy IPC.

We will write $L_{\langle W,I,V\rangle}$ for the set of formulas supported by every $s \in I$ in $\langle W, I, V\rangle$. From now on we will look at those cases where I is a topology on W. That is, closed under finite intersections and arbitrary unions. From this, we can show when a generalised inquisitive semantics will have the disjunction property. This is a generalization of (Ciardelli and Roelofsen 2011, Prop 3.9, p. 10).

Lemma 3.6. *If I is a topology on W, then $L_{\langle W,I,V\rangle}$ has the disjunction property.*

Proof. Assume $\varphi \vee \psi \in L_{\langle W,I,V\rangle}$. As I is a topology there is $t \in I$ that for all $s \in I$, it follows that $s \subseteq t$. It follows that $t \vDash^{Inq}_{\langle W,I,V\rangle} \varphi$ or $t \vDash^{Inq}_{\langle W,I,V\rangle} \chi$. As propositions are downwards closed it follows that for all $s \in I$ either $s \vDash^{Inq}_{\langle W,I,V\rangle} \varphi$ or $s \vDash^{Inq}_{\langle W,I,V\rangle} \psi$. So $\varphi \in L_{\langle W,I,V\rangle}$ or $\psi \in L_{\langle W,I,V\rangle}$. □

What is more if I is a topology then $\langle W, I, V\rangle$ satisfy the generalised Kreisel-Putnam axiom:

Lemma 3.7 (Punčochář 2016, p. 420)**.** *Every instance of the generalized Kreisel-Putnam axiom holds in every $\langle W, I, V\rangle$ where I is a topology on W.*

Proof. Note that if φ is disjunction free then it is uninquisitive so with a topological I there is a maximal information state s such that $s \vDash^{Inq}_{\langle W,I,V\rangle} \varphi$. Let $s \in I$ be such that that $s \vDash^{Inq}_{\langle W,I,V\rangle} \varphi \rightarrow \psi \vee \chi$ where φ is disjunction free. Assume for a contradiction that $s \nvDash^{Inq}_{\langle W,I,V\rangle} (\varphi \rightarrow \psi) \vee (\varphi \rightarrow \chi)$. It then follows that there are $t, r \in I$ which are subsets of s, and t supports φ and ψ but not χ, and r supports φ and χ but not ψ. As I is topological $t \cup r \in I$ and because φ is uninquisitive it follows that $t \cup r$ supports φ. Inspection of Definition 3.14 shows that if a state supports a proposition then all its substates will to and so. Hence $t \cup r$ supports $\varphi \rightarrow \psi \vee \chi$ and so it either supports ψ or χ. But then so must t and r which contradicts our assumption. □

Let L_{GInq} or general inquisitive logic be every formula which is valid on all $\langle W, I, V\rangle$ when I is a topology on W. It follows from Corollary 5 of Punčochář 2016 that L_{GInq} the same logic as $IPC + GKP$. From this, it follows that the disjunction free fragment of L_{GInq} is the same as the disjunction free fragment of intuitionistic logic. We will use this result later along with Corollary 3.2 to show that PTV results in the same logic as general inquisitive logic.

3.4 Proof-Theoretic Validity

Thus far we have been discussing weak logics and inquisitive logic. We now turn to the other main topic of this chapter, namely proof-theoretic semantics. While it ostensibly looks different, our main results again describe the relation between proof-theoretic semantics and inquisitive logics. We begin in this section by introducing the notion of an atomic system and a supersystem and by defining two notions of proof-theoretic validity relative to a supersystem. The first definition is due to Prawitz (1973). The second is due to Piecha, Campos Sanz, and Schroeder-Heister (2015). We will show that these two definitions aren't equivalent (see Lemma 3.14).

3.4.1 Atomic Systems

To define proof-theoretic validity we first need to define a very general notion of a proof rule for an atomic formula:

Definition 3.15. *(Schroeder-Heister 1984) We define atomic rule formulas and their levels as follows:*

(3.15.1.) *A* level-0 atomic rule *is an axiom consisting of a single atomic formula. That is a rule with no premises or hypotheses. It is written as $\bar{p}$ or $/p$,*

(3.15.2.) *A* level-1 atomic rule *is a rule which has premises but does not discharge hypotheses. Written $p_0, \ldots, p_n/q$ for an inference from $p_0, \ldots, p_n$ to q,*

(3.15.3.) *A* level-2 atomic rule *is a rule which discharge hypotheses. Written*

$$[p_{0_0}, \ldots, p_{m_0}]q_0, \ldots, [p_{0_n}, \ldots, p_{m_n}]q_n/r$$

for an inference from $q_0, \ldots, q_n$ to r, which for each q_i discharges $p_{0_i}, \ldots, p_{m_i}$,

(3.15.4.) *A* level-n atomic rule *(for $n > 2$) is a rule which discharge rules of level-$(n-2)$. Written*

$$([R_{0_0}, \ldots, R_{m_0}]q_0), \ldots, ([R_{0_n}, \ldots, R_{m_n}]q_n)/r$$

for a level-n inference from $q_0, \ldots, q_n$ to r, which for each q_i discharges rules $R_{0_i}, \ldots, R_{m_i}$, where the level of each R_j is at most level-$(n-2)$.

This definition can seem confusing, particularly for rules above 2.

Example 3.1. *Consider the following examples:*

$$\textit{level 0: } \overline{p} \qquad \textit{level 1: } \frac{p}{q} \qquad \textit{level 2: } \begin{array}{c} [p] \\ \vdots \\ q \\ \hline r \end{array} \qquad \textit{level 3: } \begin{array}{c} \left[\begin{array}{c} \vdots \\ \frac{p}{q} \\ \vdots \end{array}\right] \\ q \\ \hline r \end{array}$$

Using this we can give an example of a proof using a level-3 rule.

Example 3.2. *Consider the system containing the level 3 rule above, written $[p/q]q/r$, and $\bar{p}$. Note that we have written each step φ, D where φ is atomic and D is the set of atomic rules used so far and not discharged:*

$$\dfrac{\dfrac{\dfrac{}{p, \{\bar{p}\}}}{q, \{\bar{p}, \boldsymbol{p/q}\}}}{r, \{\bar{p}, [\boldsymbol{p/q}]q/r\}}$$

We define $\vdash_S$ as a relation between sets of hypothesis and rules, and atomic propositions such that $p_1, \ldots, p_n, R_1, \ldots, R_m \vdash_S p$ means there is a proof of p with open hypothesis $p_1, \ldots, p_n$ containing only the rules in $\{R_1, \ldots, R_m\} \cup S$. We will write $q \vdash_S p$ and $\bar{q} \vdash_S p$ interchangeably as there is no real effect to swapping axioms and atomic assumptions. Note the following observation:

Fact 3.1. *Rules are only removed when a level 2 or higher rule is applied. It follows that a sub-proof of a proof with level-2 or higher rules in it may have new assumptions and new rules. However, a proof with only level 0 and 1 rules does not have sub-proof that use more rules (though they may have more assumptions).*

This means that we can split proofs containing only level-0 or level-1 rules. First we give an illustrative example, and then we prove the general splitting result.

Example 3.3. *For example, given the following proof of $p/q \vdash_{\{/s; s/p; q/r\}} r$:*

$$\begin{array}{c} \bar{s} \\ \hline p \\ \hline q \\ \hline r \end{array}$$

We can split the proof into $\frac{\bar{s}}{p}$ *of* $\vdash_{\{/s;s/p;q/r\}} p$ *and* $\frac{q}{r}$ *of* $/q \vdash_{\{/s;s/p;q/r\}} r$.

Lemma 3.8. *If* $\mathcal{D}$ *is a proof in an atomic system* S*, from distinct rules* $p/q, R_0, \ldots, R_n$ *with conclusion* r*–that is if* $\mathcal{D}$ *witnesses* $p/q, R_0, \ldots, R_n \vdash_S r$*–and all the rules in* S *and* $R_0, \ldots, R_n$ *are level 0 or 1 rules, then if* $\mathcal{D}$ *contains* p/q *it follows that there are* $\mathcal{D}_1$ *and* $\mathcal{D}_2$ *witnessing* $R_0, \ldots, R_n \vdash_S p$ *and* $q, R_0, \ldots, R_n \vdash_S r$*, respectively.*

Proof. The proof proceeds by induction on the number of instances of p/q in $\mathcal{D}$.

For the base case, assume there are no instances of p/q in $\mathcal{D}$ it follows that we are done simply because the antecedent is not satisfied.

Now let us assume for all $i < m$ that the induction hypothesis holds and that $\mathcal{D}$ contains m instances of p/q. Take any such occurrence and split $\mathcal{D}$ into two proofs; as follows:

$$\cfrac{\cfrac{\cfrac{\mathcal{D}_1}{p}}{q}}{\cfrac{\mathcal{D}_2}{r}}$$

If $\mathcal{D}$ contained m occurrences of p/q then $\mathcal{D}_1$ and $\mathcal{D}_2$ contain i and j respectively where $m = i + j + 1$ to account for the application of p/q dividing $\mathcal{D}_1$ and $\mathcal{D}_2$. There are four cases we need to deal with depending on whether or not $i, j > 0$. That is $m = 1$ and $i = j = 0$; $i = 0$ and $j = m - 1$; $i = m - 1$ and $j = 0$; or both $i, j > 0$. However, it is easier to consider the subcases $i > 0$, $j > 0$, $i = 0$, and $j = 0$ separately. When considering i we show how to get a proof witnessing $R_0, \ldots, R_n \vdash_S p$ and for j we show how to get a proof witnessing $q, R_0, \ldots, R_n \vdash_S r$. Then for the four cases above one simply combines the relevant pieces of the proof.

First we consider $i > 0$. We note by Fact 3.1 that $\mathcal{D}_1$ does not contain any rules $\mathcal{D}$ didn't. What is further, because the top of the derivation remains unchanged, no new assumptions are added. This means it witnesses $p/q, R_0, \ldots, R_n \vdash_S p$. We then use the induction hypothesis to get two proofs: $\mathcal{D}_{1_1}$ witnessing $R_0, \ldots, R_n \vdash_S p$ and $\mathcal{D}_{1_2}$ witnessing $q, R_0, \ldots, R_n \vdash_S p$. The derivation $\mathcal{D}_{1_1}$ is the first proof we were looking for and we are done with $i > 0$.

Second we do $j > 0$. We note by Fact 3.1 that $\mathcal{D}_2$ does not contain any rules $\mathcal{D}$ didn't, however it does have a new assumption q. So $\mathcal{D}_2$ witnesses $q, p/q, R_0, \ldots, R_n \vdash_S r$. So by the induction hypothesis we can split it into two proofs: $\mathcal{D}_{2_1}$ witnessing $q, R_0, \ldots, R_n \vdash_S p$ and $\mathcal{D}_{2_2}$ witnessing $q, R_0, \ldots, R_n \vdash_S r$. The derivation $\mathcal{D}_{2_2}$ is the second proof we were looking for and we are done with $j > 0$.

Third assume $i = 0$ then we still know $\mathcal{D}_1$ witnesses $p/q, R_0, \ldots, R_n \vdash_S p$ but we also know it contains no instances of p/q, so in fact it witnesses $R_0, \ldots, R_n \vdash_S p$. The derivation $\mathcal{D}_1$ is the third proof we were looking for and we are done with $i = 0$.

Fourth assume $j = 0$ then just as before we have $\mathcal{D}_2$ witnessing $q, p/q, R_0, \ldots, R_n \vdash_S r$. But as $\mathcal{D}_2$ contained no instances of p/q it witnesses $q, R_0, \ldots, R_n \vdash_S r$. The derivation $\mathcal{D}_2$ is the fourth proof we were looking for and we are done with $j = 0$. And so, we are done. □

We will use this result later in Lemma 3.21 to show how provability in supersystems changes. This result doesn't hold when there are higher-level rules since discharged rules may be separated from the rule that discharges them.

Atomic rules can now be used to define atomic systems and supersystems.

Definition 3.16. *Let the set of all atomic rules of any level be denoted as $\mathbb{S}$.*

(3.16.1.) Call a set of atomic rules $S \subseteq \mathbb{S}$ an atomic system.

(3.16.2.) Call a set $\mathfrak{S} \subseteq \mathscr{P}(\mathbb{S})$ an atomic supersystem.

We use these supersystems as a base to define an abstract notion of proof-theoretic validity.

3.4.2 Prawitz's Definition of Proof-Theoretic Validity

Proof-Theoretic Validity was proposed by Prawitz (1971) as an explication of Gentzen's claim that:

> The introductions represent, as it were, the 'definitions' of the symbols concerned, and the eliminations are no more, in the final analysis than the consequences of these definitions. This fact may be expressed as follows: In eliminating a symbol, we may use the formula with whose terminal symbol we are dealing only 'in the sense afforded it by the introduction of that symbol'. (Gentzen 1935, p. 80)

There are well-known problems with taking this explanation literally. As Prawitz points out the introduction rules are not, in fact, explicit definitions, in the sense of definiendum-definiens pairs. Because of this Prawitz suggests, instead, that the introduction rules are used to inductively define a notion of 'validity'. In doing this, Prawitz is making the notion of consequence central to his interpretation of Gentzen. So, instead of showing that the elimination rules follow in some way from the introduction rules, it would instead be shown that they preserve 'validity'. In this definition, S is an atomic system and $\mathcal{J}$ is a set of transformations on derivations which preserve the conclusion and do not add open assumptions, though they may delete them.

Definition 3.17. *(Prawitz 1973, p. 236; Schroeder-Heister 2006, p. 560; Schroeder-Heister 2018, Suplement 1) A derivation $\mathcal{D}$ being an* $(S, \mathcal{J})$-valid derivation *for an atomic system S and set of justifications $\mathcal{J}$ is defined inductively as follows:*

(3.17.1.) If $\mathcal{D}$ is a closed derivation in S then it is $(S, \mathcal{J})$-valid.

(3.17.2.) If $\mathcal{D}$ is a closed derivation ending in an introduction rule then it is $(S, \mathcal{J})$-valid if the derivations of the premises of the introduction rule are $(S, \mathcal{J})$-valid.

(3.17.3.) If $\mathcal{D}$ is a closed derivation which does not end in an introduction rule then it is $(S, \mathcal{J})$-valid if it $\mathcal{J}$-reduces to an $(S, \mathcal{J})$-valid derivation which does end in an introduction rule.

(3.17.4.) If $\mathcal{D}$ is an open derivation of φ with open assumptions $\varphi_0, \ldots, \varphi_n$ then it is $(S, \mathcal{J})$-valid if for all atomic systems S' extending S, all justifications $\mathcal{J}'$ extending $\mathcal{J}$ and all closed $(S', \mathcal{J}')$-valid derivations $\mathcal{D}_0, \ldots, \mathcal{D}_n$ of $\varphi_0, \ldots, \varphi_n$, the following derivation is $(S', \mathcal{J}')$-valid:

$$\dfrac{\begin{matrix}\mathcal{D}_0 & \cdots & \mathcal{D}_n \\ \varphi_0 & \cdots & \varphi_n\end{matrix}}{\begin{matrix}\mathcal{D} \\ \varphi\end{matrix}}$$

Example 3.4. *This definition can be illustrated by considering an example. Take the following proof of $p \to \neg\neg p$ or $p \to ((p \to \bot) \to \bot)$ let $S = \varnothing$ and $\mathcal{J}$ be the standard reductions used in the proof of normalisation:*

$$\dfrac{\dfrac{\dfrac{[p] \quad [p \to \bot]}{\bot}}{(p \to \bot) \to \bot}}{p \to ((p \to \bot) \to \bot)}$$

This is a closed proof and it ends on an introduction rule so by condition 3.17.2 this proof is $(S, \mathcal{J})$-valid if its immediate sub-proof is:

$$\dfrac{\dfrac{p \quad [p \to \bot]}{\bot}}{(p \to \bot) \to \bot}$$

This proof is open and so by condition 3.17.4 it is $(S, \mathcal{J})$-valid if given any closed proof of p in every extension S' and $\mathcal{J}'$ of S and $\mathcal{J}$ the proof generated by replacing the assumption p with its derivation is $(S', \mathcal{J}')$-valid. For the purposes of illustration let us consider the simplest case where we simply extend S with the axiom $\bar{p}$. By 3.17.1 the proof $\bar{p}$ is $(\{\bar{p}\}, \mathcal{J})$-valid. The new proof ends in an introduction rule and so by 3.17.2 will be $(\{\bar{p}\}, \mathcal{J})$-valid just if its immediate sub proof is:

$$\frac{\bar{p} \qquad p \to \bot}{\bot}$$

This is again an open proof so by condition 3.17.4 we need to consider all expansions of $\{\bar{p}\}$ *and* $\mathcal{J}$*. This time there are two obvious expansions of* $\{\bar{p}\}$*. The first by* $\bar{\bot}$ *and the second by* $\frac{p}{\bot}$*. Let's use the second. We then get the following proof which we need to check is* $(\{\bar{p}, \frac{p}{\bot}\}, \mathcal{J})$*-valid:*

$$\frac{\bar{p} \qquad \dfrac{\dfrac{[p]}{\bot}}{p \to \bot}}{\bot}$$

This proof does not end in an introduction rule, so we finally get to use condition 3.17.3. Careful inspection of the proof will show that we introduced $\to$ *only to eliminate it directly afterwards. But we can use one of the reductions from the proof of normalisation to remove this. This results in the proof:*

$$\frac{\bar{p}}{\bot}$$

Which is $(\{\bar{p}, \frac{p}{\bot}\}, \mathcal{J})$*-valid by 3.17.1. To make this illustration into a rigorous argument, one would simply follow this line of reasoning backwards, starting with the small proofs and building up, and replacing our consideration of the simplest cases by the more general cases. But like with other inductive definitions, such as that of well-formed formula, sometimes they are best illustrated by breaking a familiar example down rather than building up from the base cases.*

We can replace Prawitz's conditions on derivations with a consequence relation. To do this we need to define a consequence relation directly from derivations. We take this definition from Schroeder-Heister (2006, p. 561) who is explicating Prawitz's conjecture (Prawitz 1973, p. 246).

Definition 3.18. *Let* $\Gamma \Vdash_S^{\mathfrak{S}} \varphi$ *hold if there is a derivation* $\mathcal{D}$ *of* φ *with open assumptions in* Γ *which is* $(S, \mathcal{J}_{MAX})$*-valid where* $\mathcal{J}_{MAX}$ *is the maximal set of justifications. Given a supersystem* $\mathfrak{S}$ *let the Prawitz semantics associated with it define the consequence relation* $\Vdash^{\mathfrak{S}}$ *where* $\Gamma \Vdash_{\mathfrak{S}} \varphi$ *if and only if for all* $S \in \mathfrak{S}$ *it follows* $\Gamma \Vdash_S^{\mathfrak{S}} \varphi$*.*

Here we understand justifications in the sense of Schroeder-Heister (2006, p. 558) where a justification is any map from proofs to proofs, which preserve conclusions and does not add assumptions, that can be applied even if the proof is a sub-proof of a larger one. All the reductions used in normalization are justifications. As Schroeder-Heister points out, this differs from Prawitz's who places a constraint of 'consistency' on justifications. However, Schroeder-Heister argues that Prawitz's constraints were based on worries about normalization, not validity, and so they can be removed. We will use Schroeder-Heister's definition here.

Lemma 3.9. *Given a supersystem* $\mathfrak{S}$ *Prawitz's validity notion from Definition 3.18 satisfies the following:*

$$\vdash_S p \Longleftrightarrow \Vdash_S^{\mathfrak{S}} p, \quad \text{(Autonomy of Atoms)}$$

$$\Vdash_S^{\mathfrak{S}} \varphi \textit{ and } \Vdash_S^{\mathfrak{S}} \psi \Longleftrightarrow \Vdash_S^{\mathfrak{S}} \varphi \wedge \psi, \quad \text{(Conjunction Property)}$$

$$\Vdash_S^{\mathfrak{S}} \varphi \textit{ or } \Vdash_S^{\mathfrak{S}} \psi \Longleftrightarrow \Vdash_S^{\mathfrak{S}} \varphi \vee \psi, \quad \text{(Disjunction Property)}$$

$$[\forall S' \supseteq S(S' \in \mathfrak{S} \textit{ and } \Vdash_{S'}^{\mathfrak{S}} \psi \Rightarrow \Vdash_{S'}^{\mathfrak{S}} \varphi)] \Longleftrightarrow \Vdash_S^{\mathfrak{S}} \psi \to \varphi. \quad \text{(Weak Monotonicity)}$$

$$\exists \text{finite}\Delta \subseteq \Gamma[\forall S' \supseteq S(S' \in \mathfrak{S} \textit{ and } \Vdash_{S'}^{\mathfrak{S}} \Delta \Rightarrow \Vdash_{S'}^{\mathfrak{S}} \varphi)] \Longleftrightarrow \Gamma \Vdash_S^{\mathfrak{S}} \varphi. \quad \text{(Compact Monotonicity)}$$

Proof. (Autonomy of Atoms): By definition, if $\vdash_S p$ there is a derivation which is S-valid. Assume there is a closed S-valid derivation $\mathcal{D}$ of an atomic proposition p. Now $\mathcal{D}$ is either in S, ends in an introduction rule, or reduces to a proof $\mathcal{D}'$ meeting one of the previous two conditions. So, we can assume we have a derivation either ending in an introduction rule or a derivation in S. But it cannot end with an introduction rule as it is a proof of an atomic formula. This leaves only that it is in S and so $\Vdash_S^{\mathfrak{S}} p \Leftrightarrow \vdash_S p$.

The proof of the conjunction and disjunction properties are relatively simple and so we exclude them here.

(Compact Monotonicity): It needs to be shown that $\Gamma \Vdash_S^{\mathfrak{S}} \varphi$ if and only if $\exists finite \Delta \subseteq \Gamma \forall S' \supseteq S[\Vdash_{S'}^{\mathfrak{S}} \Delta \Rightarrow \Vdash_{S'}^{\mathfrak{S}} \varphi]$. Assume $\Gamma \Vdash_S^{\mathfrak{S}} \varphi$ then we know that there is some finite subset $\Delta = \{\psi_0, \ldots, \psi_n\}$ and a derivation

$$\begin{array}{c} \psi_0, \ldots, \psi_n \\ \mathcal{D} \\ \varphi \end{array}$$

and given any S' extending S and any closed S'-valid derivations $\begin{array}{c} \mathcal{D}_i \\ \psi_i \end{array}$ then the closed proof

$$\begin{array}{ccc} \mathcal{D}_0 & & \mathcal{D}_n \\ \psi_0 & \ldots & \psi_n \\ & \mathcal{D} & \\ & \varphi & \end{array}$$

is S' valid. From which monotonicity clearly follows.

So now assume that there is finite $\Delta \subseteq \Gamma$ and for every S' extending S if there are closed S'-valid derivation of Δ then there is a closed S'-valid derivation of φ. Assume for a contradiction that $\Gamma \Vdash_S^{\mathfrak{S}} \varphi$ is false. What follows from this is that there is no derivation in S with open assumptions in Γ that satisfies Definition 3.17.4. So, for the derivation

$$\frac{\Delta}{\varphi}$$

Since this does not satisfy Definition 3.17.4, there is a S' extending S such that there are S'-valid derivations $\mathcal{D}_i$ of every formula $\psi_i \in \Delta$ but the composition of those derivations with the original derivation of φ:

$$\dfrac{\begin{matrix}\mathcal{D}_0 & & \mathcal{D}_n \\ \psi_0 & \ldots & \psi_n\end{matrix}}{\varphi}$$

is not S'-valid. But we have all justifications in our system, and we are dealing with closed proofs we have the justification that simply takes this derivation to *any* S'-valid derivation of φ. So, this comes down to the idea that there isn't a derivation of φ in S'. But by assumption and the fact that there are S'-valid derivations of Δ we know there is a derivation of φ.

(Weak Monotonicity): By Compact Monotonicity it is sufficient to show that $\psi \Vdash_S^{\mathfrak{S}} \varphi$ if and only if $\Vdash_S^{\mathfrak{S}} \psi \to \varphi$.

Assume $\psi \Vdash_S^{\mathfrak{S}} \varphi$ then there is an open S-valid derivation of ψ from premise φ. It follows that by condition 2 of Definition 3.17 the following is an S-valid derivation of $\varphi \to \psi$:

$$\dfrac{\begin{matrix}[\varphi] \\ \mathcal{D} \\ \psi\end{matrix}}{\varphi \to \psi}$$

so $\Vdash_S^{\mathfrak{S}} \psi \to \varphi$.

Assume $\Vdash_S^{\mathfrak{S}} \psi \to \varphi$ then $\mathcal{D}'$ is S-valid derivation of $\varphi \to \psi$. As $\varphi \to \psi$ is not an atomic formula it follows that either condition 3 holds of $\mathcal{D}$, in which case there is a $\mathcal{D}'$ of which condition 2 holds, or condition 2 holds. So, we assume condition 2 holds. Then $\mathcal{D}'$ is of the form:

$$\dfrac{\begin{matrix}[\varphi] \\ \mathcal{D} \\ \psi\end{matrix}}{\varphi \to \psi} \quad \text{From this it follows that:} \begin{matrix}\varphi \\ \mathcal{D} \\ \psi\end{matrix}$$

is an S-valid derivation of ψ from φ so $\psi \Vdash_S^{\mathfrak{S}} \varphi$. □

It can also be shown that any two consequence relations satisfying the conditions in Lemma 3.9 for a Prawitz semantics will be identical.

Lemma 3.10. *Given a supersystem $\mathfrak{S}$ and two binary relations satisfying the conditions of Lemma 3.9, $(\mathfrak{S}, \Vdash_{\mathfrak{S}})$ and $(\mathfrak{S}, \Vdash'_{\mathfrak{S}})$, it follows that $\Vdash_S^{\mathfrak{S}} = \Vdash_S'^{\mathfrak{S}}$ for all $S \in \mathfrak{S}$.*

Though we do not include the proof here, it is a simple induction on formula complexity.

3.4.3 PCS Semantics

While the above formulation is essentially that found in Prawitz's work, most of the results on what is provable in proof-theoretic semantics use a different formulation. We call it PCS semantics because it appears in Piecha, Campos Sanz, and Schroeder-Heister (2015). It is PCS semantics that we will use for the rest of the chapter. The two definitions differ only on the condition for $\Gamma \vDash_S^{\mathfrak{S}} \varphi$. While it appears not to have been addressed in the literature, we will show that this notion is not equivalent to Prawitz's. It is hoped that in future work we will be able to show that the definitions coincide on the relevant supersystems.

Definition 3.19. *A pair $(\mathfrak{S}, \vDash_{\mathfrak{S}})$ is a PCS semantics if $\mathfrak{S}$ is a supersystem (Definition 3.16), and if for every S in $\mathfrak{S}$ one has that there is a $\vDash_S^{\mathfrak{S}}$ which satisfies the following:*

$$\vdash_S p \Longleftrightarrow \vDash_S^{\mathfrak{S}} p, \qquad \text{(Autonomy of Atoms)}$$

$$\vDash_S^{\mathfrak{S}} \varphi \textit{ and } \vDash_S^{\mathfrak{S}} \psi \Longleftrightarrow \vDash_S^{\mathfrak{S}} \varphi \wedge \psi, \qquad \text{(Conjunction Property)}$$

$$\vDash_S^{\mathfrak{S}} \varphi \textit{ or } \vDash_S^{\mathfrak{S}} \psi \Longleftrightarrow \vDash_S^{\mathfrak{S}} \varphi \vee \psi, \qquad \text{(Disjunction Property)}$$

$$[\forall S' \supseteq S(S' \in \mathfrak{S} \textit{ and } \vDash_{S'}^{\mathfrak{S}} \psi \Rightarrow \vDash_{S'}^{\mathfrak{S}} \varphi)] \Longleftrightarrow \vDash_S^{\mathfrak{S}} \psi \to \varphi. \qquad \text{(Weak Monotonicity)}$$

$$[\forall S' \supseteq S(S' \in \mathfrak{S} \textit{ and } \vDash_{S'}^{\mathfrak{S}} \Gamma \Rightarrow \vDash_{S'}^{\mathfrak{S}} \varphi)] \Longleftrightarrow \Gamma \vDash_S^{\mathfrak{S}} \varphi. \qquad \text{(Monotonicity)}$$

and further $\vDash_{\mathfrak{S}}$ *is defined from* $\vDash_S^{\mathfrak{S}}$ *as follows:*

$$\Gamma \vDash_{\mathfrak{S}} \varphi \Longleftrightarrow \forall S \in \mathfrak{S}, \Gamma \vDash_S^{\mathfrak{S}} \varphi. \tag{3.17}$$

It turns out that each supersystem has a unique PCS semantics.

Lemma 3.11. *Given a supersystem* $\mathfrak{S}$ *there is a consequence relation* $\vDash_{\mathfrak{S}}$ *such that* $(\mathfrak{S}, \vDash_{\mathfrak{S}})$ *is a PCS semantics and given two PCS semantics* $(\mathfrak{S}, \vDash_{\mathfrak{S}})$ *and* $(\mathfrak{S}, \vDash'_{\mathfrak{S}})$ *it follows that* $\vDash_S^{\mathfrak{S}} = \vDash_S'^{\mathfrak{S}}$ *for* $S \in \mathfrak{S}$.

We will now show explicitly that all PCS semantics are extensions of minimal logic. For the axioms of minimal logic, we use a Hilbert system (Troelstra and Schwichtenberg 2000, p. 51).

Lemma 3.12. *Given a PCS semantics* $(\mathfrak{S}, \vDash_{\mathfrak{S}})$ *we have that the following are satisfied:*

$$\begin{aligned}
&(a)\varphi \to (\psi \to \varphi), \quad (b)(\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi)) \\
&(c)\varphi \to \varphi \vee \psi, \quad (d)\psi \to \varphi \vee \psi; \\
&(e)(\varphi \to \chi) \to ((\psi \to \chi) \to (\varphi \vee \psi \to \chi)) \\
&(f)\varphi \wedge \psi \to \varphi \quad (g)\varphi \wedge \psi \to \psi, \quad (h)\varphi \to (\psi \to (\varphi \wedge \psi))
\end{aligned}$$

Proof. (a) Let $S \in \mathfrak{S}$ be such that $\vDash_S^{\mathfrak{S}} \varphi$ then for all $S' \supseteq S$ in $\mathfrak{S}$ such that $\vDash_{S'}^{\mathfrak{S}} \psi$ it follows because $S' \supseteq S$ and Monotonicity that $\vDash_{S'}^{\mathfrak{S}} \varphi$.

(b) Let $S \in \mathfrak{S}$ be such that $\vDash_S^{\mathfrak{S}} \varphi \to (\psi \to \chi)$ and let $S'' \supseteq S' \supseteq S$ be in $\mathfrak{S}$ and such that $\vDash_{S'}^{\mathfrak{S}} (\varphi \to \psi)$ and $\vDash_{S''}^{\mathfrak{S}} \varphi$. It is sufficient to show that $\vDash_{S''}^{\mathfrak{S}} \chi$. By monotonicity $\vDash_{S''}^{\mathfrak{S}} \varphi \to (\psi \to \chi)$ and $\vDash_{S''}^{\mathfrak{S}} (\varphi \to \psi)$. But then by the weak deduction theorem and $\vDash_{S''}^{\mathfrak{S}} \varphi$ it follows that $\vDash_{S''}^{\mathfrak{S}} (\psi \to \chi)$ and $\vDash_{S''}^{\mathfrak{S}} \psi$. From which it follows that $\vDash_{S'}^{\mathfrak{S}} \chi$.

(c-d) Follow from Disjunction Property.

(e) Let $S \in \mathfrak{S}$ be such that $\vDash^{\mathfrak{S}}_{S} (\varphi \to \chi)$ and let $S'' \supseteq S' \supseteq S$ be in $\mathfrak{S}$ and such that $\vDash^{\mathfrak{S}}_{S'} (\psi \to \chi)$ and $\vDash^{\mathfrak{S}}_{S''} \varphi \vee \psi$ then by the disjunction property $\vDash^{\mathfrak{S}}_{S''} \varphi$ or $\vDash^{\mathfrak{S}}_{S''} \psi$. By monotonicity $\vDash^{\mathfrak{S}}_{S'} (\varphi \to \chi)$ and $\vDash^{\mathfrak{S}}_{S'} (\psi \to \chi)$. So, whichever case holds by the weak deduction theorem $\vDash^{\mathfrak{S}}_{S''} \chi$.

(f-h) Follow from Conjunction Property. □

If the supersystem includes $\bot/p$ for all atomic formulas p then it will satisfy intuitionistic logic.

Lemma 3.13. *If every $S \in \mathfrak{S}$ contains $\bot/p$ for all atomic formulas p then $\vDash_{\mathfrak{S}} \bot \to \varphi$ for all formulas φ.*

Proof. Let $S \in \mathfrak{S}$ we will proceed by induction on complexity of formulas. This is trivial for the atomic case and simple for the case of $\wedge$ and $\vee$. So, we only show $\to$. We want to show $\vDash^{\mathfrak{S}}_{S} \bot \to (\varphi \to \chi)$. Let $S' \supseteq S$ be such that $\vDash^{\mathfrak{S}}_{S'} \bot \wedge \varphi$. By the induction hypothesis $\vDash^{\mathfrak{S}}_{S'} \bot \to \chi$ so $\vDash^{\mathfrak{S}}_{S'} \chi$ and we are done. □

It can be shown that PCS and Prawitz notion from Section 3.4.2 are distinct because Monotonicity isn't equivalent to Compact Monotonicity.

Lemma 3.14. *Let $\mathfrak{S} = \{\{p_0, \ldots, p_n\} \mid n \in \mathbb{N}\} \cup \{\{q, p_0, p_1, \ldots\}\}$ then $\{p_0, p_1, \ldots\} \vDash_{\mathfrak{S}} q$ but $\{p_0, p_1, \ldots\} \nVdash_{\mathfrak{S}} q$.*

Proof. Note that $\{p_0, p_1, \ldots\} \vDash_{\mathfrak{S}} q$ if for all $S \in \mathfrak{S}$ when $\vDash^{\mathfrak{S}}_{S} \{p_0, p_1, \ldots\}$ then $\vDash^{\mathfrak{S}}_{S} q$. As the only S such that $\vDash^{\mathfrak{S}}_{S} \{p_0, p_1, \ldots\}$ is $\{q, p_0, p_1, \ldots\}$ this follows. However, $\{p_0, p_1, \ldots\} \Vdash_{\mathfrak{S}} q$ if there is a derivation $\mathcal{D}$ of q with (finite) open assumptions in $\{p_0, p_1, \ldots\}$ such that any S which has a derivation of those open assumptions has a derivation of q. Assume this is so for a contradiction. Then there is some n such that $\{p_0, \ldots, p_n\}$ contains all the open assumptions in $\mathcal{D}$. But then $S = \{p_0, \ldots, p_n\}$ has a derivation of $p_0, \ldots, p_n$ and yet it contains (by there being no rules ending in q) no derivation of q. It follows that $\{p_0, p_1, \ldots\} \nVdash_{\mathfrak{S}} q$. $\square$

The cause of this is that while $\vDash$ has the monotonicity condition $[\forall S' \supseteq S(\vDash^{\mathfrak{S}}_{S'} \Gamma \Rightarrow \vDash^{\mathfrak{S}}_{S'} \varphi)] \iff \Gamma \vDash^{\mathfrak{S}}_{S} \varphi$, for $\Vdash$ there is instead compact monotonicity condition $\exists \text{finite} \Delta \subseteq \Gamma [\forall S' \supseteq S(\vDash^{\mathfrak{S}}_{S'} \Delta \Rightarrow \vDash^{\mathfrak{S}}_{S'} \varphi)] \iff \Gamma \vDash^{\mathfrak{S}}_{S} \varphi$. A straightforward result of this proof is that $\Vdash$ does not always satisfy monotonicity as it still holds in the example above that if $\Vdash^{\mathfrak{S}}_{S} \{p_0, p_1, \ldots\}$, then $\Vdash^{\mathfrak{S}}_{S} q$. However, when monotonicity does hold Lemma 3.9 straightforwardly gives us the following result:

Lemma 3.15. *Given a supersystem $\mathfrak{S}$ for which $\Vdash$ is monotone, then for all $S \in \mathfrak{S}$ and φ and Γ we have*

$$\Gamma \Vdash^{\mathfrak{S}}_{S} \varphi \Leftrightarrow \Gamma \vDash^{\mathfrak{S}}_{S} \varphi$$

Proof. By Lemma 3.9 and the assumption of monotonicity we know that it meets the criteria to be a PCS semantics and then by Lemma 3.11 we know that any two PCS semantics with the same supersystem are identical. $\square$

On the other side of the equation what is lacking is compactness. That is $\Gamma \vDash^{\mathfrak{S}}_{S} \varphi \Leftrightarrow \exists \text{finite} \Delta \subseteq \Gamma \ \Delta \vDash^{\mathfrak{S}}_{S} \varphi$. It can be shown that whenever the PCS semantics is compact it agrees with Prawitz's notion.

Lemma 3.16. *Let $\vdash_L$ be compact and monotonic then it has compact monotonicity.*

Proof. $\Gamma \vdash_L \varphi \Leftrightarrow_{\text{Compactness}} \exists \text{finite} \Delta \subseteq \Gamma \ \Delta \vdash_L \varphi \Leftrightarrow_{\text{Monotonicity}} \exists \text{finite} \Delta \subseteq \Gamma [\forall S' \supseteq S(\Vdash^{\mathfrak{S}}_{S'} \Delta \Rightarrow \Vdash^{\mathfrak{S}}_{S'} \varphi)]$. $\square$

This allows us to show:

Lemma 3.17. *Given a supersystem $\mathfrak{S}$ for which $\vDash$ is compact, then for all $S \in \mathfrak{S}$ and φ and Γ we have*

$$\Gamma \vDash_S^{\mathfrak{S}} \varphi \Leftrightarrow \Gamma \Vdash_S^{\mathfrak{S}} \varphi$$

Proof. Given that $\vDash_S^{\mathfrak{S}}$ is compact it follows that it has compact monotonicity by Lemma 3.16 as well as the first 4 conditions of a PCS semantics. By Lemma 3.9 we know that $\Vdash_S^{\mathfrak{S}}$ satisfies the first 4 conditions of a PCS semantics and compact monotonicity. Then by Lemma 3.10 it follows that they must be identical. □

We do not currently know which supersystems are compact or monotone. And so, we do not know on which supersystems PCS semantics and Prawitz's validity notion align. Characterising the set of supersystems which are compact will be the subject of future work. For our purposes, we continue to work with PCS semantics given the usefulness of monotonicity. When we move to the weak logics defined by these systems we will have compactness built-in.

3.5 Results about PCS semantics

3.5.1 Relations between Supersystems

The supersystem which is chosen can have a considerable impact on the logic that results. We consider $\mathfrak{S}_\infty^M = \mathscr{P}(\mathbb{S})$ to be the largest possible supersystem. Here M stands for minimal as there is no special treatment of $\perp$ in this system. The system $\mathfrak{S}_\infty = \{S \subseteq \mathbb{S} \mid (\perp/p) \in S \text{ for all atomic } p\}$ restricts $\mathfrak{S}_\infty^M$ to ensure that $\perp$ behaves as it does in intuitionistic logic. All other supersystems are defined as subsets of these supersystems. The following result is worth noting:

Lemma 3.18. *Let φ be $\to$ free and $\mathfrak{S}' \subseteq \mathfrak{S}$, then for all $S \in \mathfrak{S}'$*

$$\vDash^{\mathfrak{S}}_{S} \varphi \Leftrightarrow \vDash^{\mathfrak{S}'}_{S} \varphi$$

Proof. For the atomic case note that: $\vDash^{\mathfrak{S}}_{S} p \Leftrightarrow \vdash_{S} p \Leftrightarrow \vDash^{\mathfrak{S}'}_{S} p$. Now for the induction cases, assume $\vDash^{\mathfrak{S}}_{S} \varphi \Leftrightarrow \vDash^{\mathfrak{S}'}_{S} \varphi$ and $\vDash^{\mathfrak{S}}_{S} \psi \Leftrightarrow \vDash^{\mathfrak{S}'}_{S} \psi$. The $\wedge$ and $\vee$ follow from the induction hypothesis. □

Corollary 3.3. *Let φ be $\to$ free and $\mathfrak{S}' \subseteq \mathfrak{S}$, then*

$$\vDash_{\mathfrak{S}} \varphi \Rightarrow \vDash_{\mathfrak{S}'} \varphi$$

Given a supersystem $\mathfrak{S}$ and an atomic system $K \in \mathfrak{S}$, two natural ways to get new supersystems are as follows:

1. $\{S \in \mathfrak{S} \mid S \subseteq K\}$

   This defines an ideal on $\mathfrak{S}$. Examples are $\mathfrak{S}^M_n$ and $\mathfrak{S}_n$, where these are $\mathfrak{S}^M_n = \{S \in \mathfrak{S}^M_\infty \mid S \subseteq \{R \mid R \text{ is level } n\}\}$ and $\mathfrak{S}_n = \{S \in \mathfrak{S}_\infty \mid S \subseteq \{R \mid R \text{ is level } n\}\}$. That is $\mathfrak{S}^M_n$ and $\mathfrak{S}_\infty$ respectively restricted to rules of at most level n.

2. $\{S \in \mathfrak{S} \mid K \subseteq S\}$

   This defines a filter on $\mathfrak{S}$. An example is $\mathfrak{S}_\infty$ which is $\{S \in \mathfrak{S}^M_\infty \mid \{(\bot/p) \mid p \text{ atomic}\} \subseteq S\}$. In other words, $\mathfrak{S}_\infty$ is every $S \in \mathfrak{S}^M_\infty$ such that $(\bot/p) \in S$ for every atomic p.

A supersystem $\mathfrak{S}' \subseteq \mathfrak{S}$ is closed under supersets from $\mathfrak{S}$ if whenever $S \in \mathfrak{S}'$, $S' \in \mathfrak{S}$ and $S \subseteq S'$ then $S' \in \mathfrak{S}'$. Filters 2 are closed under supersets from $\mathfrak{S}$. If a supersystem is a subset of another and is closed under supersets from it, then it proves anything the larger supersystem proves:

Lemma 3.19. *If $\mathfrak{S}' \subseteq \mathfrak{S}$ is closed under supersets from $\mathfrak{S}$, then for all $S \in \mathfrak{S}'$*

$$\vDash_S^{\mathfrak{S}} \varphi \Leftrightarrow \vDash_S^{\mathfrak{S}'} \varphi$$

Proof. Induction on φ. All cases except for $\to$ are covered by the inductive steps in the proof of Lemma 3.18.

For the induction hypothesis, assume we have for all $S \in \mathfrak{S}'$ that $\vDash_S^{\mathfrak{S}} \varphi \Leftrightarrow \vDash_S^{\mathfrak{S}'} \varphi$ and $\vDash_S^{\mathfrak{S}} \psi \Leftrightarrow \vDash_S^{\mathfrak{S}'} \psi$. Assume $\vDash_{S'}^{\mathfrak{S}} \varphi \to \psi$ and $S' \in \mathfrak{S}'$, then by the weak monotonicity and monotonicity of Definition 3.19 it follows that for all $S \in \mathfrak{S}$ extending S' if $\vDash_S^{\mathfrak{S}} \varphi$ then $\vDash_S^{\mathfrak{S}} \psi$. Take $S'' \in \mathfrak{S}'$ extending S' such that $\vDash_{S''}^{\mathfrak{S}'} \varphi$. Clearly $S'' \in \mathfrak{S}$. Now by the induction hypothesis $\vDash_{S''}^{\mathfrak{S}} \varphi$ from which it follows that $\vDash_{S''}^{\mathfrak{S}} \psi$ and so again by the induction hypothesis $\vDash_{S''}^{\mathfrak{S}'} \psi$.

Assume $\vDash_S^{\mathfrak{S}'} \varphi \to \psi$, then for all $S' \in \mathfrak{S}'$ extending S such that $\vDash_{S'}^{\mathfrak{S}'} \varphi$ it follows that $\vDash_{S'}^{\mathfrak{S}'} \psi$. Take $S' \in \mathfrak{S}$ extending S such that $\vDash_{S'}^{\mathfrak{S}} \varphi$. As S' extends $S \in \mathfrak{S}'$ it follows by assumption of closure under supersets that $S' \in \mathfrak{S}'$. Now by the induction hypothesis $\vDash_{S'}^{\mathfrak{S}'} \varphi$ from which it follows that $\vDash_{S'}^{\mathfrak{S}'} \psi$ and so again by the induction hypothesis $\vDash_{S'}^{\mathfrak{S}} \psi$. □

Lemma 3.20. *If $\mathfrak{S}' \subseteq \mathfrak{S}$ is closed under supersets from $\mathfrak{S}$, then*

$$\vDash_{\mathfrak{S}} \varphi \Rightarrow \vDash_{\mathfrak{S}'} \varphi$$

Proof. Assume $\vDash_{\mathfrak{S}} \varphi$ then for all $S \in \mathfrak{S}$, $\vDash_S^{\mathfrak{S}} \varphi$. As all $S' \in \mathfrak{S}'$ are also in $\mathfrak{S}$ it follows by Lemma 3.19 that $\vDash_{S'}^{\mathfrak{S}'} \varphi$ and so $\vDash_{\mathfrak{S}'} \varphi$. □

Straight away we get from this that as $\mathfrak{S}_\infty$ is a filter on $\mathfrak{S}_\infty^M$ it follows that $\vDash_{\mathfrak{S}_\infty^M} \varphi$ implies $\vDash_{\mathfrak{S}_\infty} \varphi$ and so the 'intuitionistic' system prove everything the 'minimal' one does. The hypothesis of closure under supersets is crucial for Lemma 3.20. This can be shown by considering $\mathfrak{S}_1$ and $\mathfrak{S}_\infty$. Note that $\mathfrak{S}_1 \subseteq \mathfrak{S}_\infty$ but $\mathfrak{S}_1$ is not closed under supersets in $\mathfrak{S}_\infty$.

Every instance of the generalised Kreisel-Putnam axiom is provable in $\mathfrak{S}_\infty$ (see Lemma 3.25) but we can construct instances of the generalised Kreisel-Putnam axiom not provable in $\mathfrak{S}_1$, as we show in the proof below. So, we see that the smaller supersystem $\mathfrak{S}_1$ does not prove everything the larger one $\mathfrak{S}_\infty$ does.

Lemma 3.21. *For distinct r, p, q not equal to falsum: $\vDash_{\mathfrak{S}_1} ((p \to q) \to r) \to (r \vee (p \wedge (q \to r)))$ but $\nvDash_{\mathfrak{S}_1} ((p \to q) \to r) \to r$ and $\nvDash_{\mathfrak{S}_1} ((p \to q) \to r) \to (p \wedge (q \to r))$.*

Proof. First we show that $\vDash_{\mathfrak{S}_1} ((p \to q) \to r) \to (r \vee (p \wedge (q \to r)))$. This follows if for all $S \in \mathfrak{S}_1$ if $\vDash_S^{\mathfrak{S}_1} ((p \to q) \to r)$ then $\vDash_S^{\mathfrak{S}_1} r \vee (p \wedge (q \to r))$. Let $S \in \mathfrak{S}_1$ be such that $\vDash_S^{\mathfrak{S}_1} ((p \to q) \to r)$. Now either $\vdash_S r$ or not. If $\vdash_S^{\mathfrak{S}_1} r$ then clearly $\vDash_S^{\mathfrak{S}_1} r \vee (p \wedge (q \to r))$. So, assume not. We know that $\vdash_{S \cup \{p/q\}}^{\mathfrak{S}_1} r$ because $\vDash_{S \cup \{p/q\}}^{\mathfrak{S}_1} p \to q$. So as we have at most level 1 rules there must be proofs, by Lemma 3.8, such that $\vdash_S p$ and $q \vdash_S r$, so $\vDash_S^{\mathfrak{S}_1} r \vee (p \wedge (q \to r))$.

But $\nvDash_{\mathfrak{S}_1} ((p \to q) \to r) \to r$ as $S = \{\bar{p}, q/r\} \cup \{\bot/a \mid a \text{ atomic}\}$ demonstrates. Note that $\vDash_S^{\mathfrak{S}_1} ((p \to q) \to r)$ as assume S' extending S proves $\vDash_{S'}^{\mathfrak{S}_1} p \to q$ then as $\bar{p} \in S'$ it follows that $\vDash_{S'}^{\mathfrak{S}_1} q$ and as $q/r \in S'$ then $\vDash_{S'}^{\mathfrak{S}_1} r$. We need to show that $\nvDash_S^{\mathfrak{S}_1} r$ which is the case only if $\vdash_S r$. Assume so then there is a derivation ending in r but this can be the case only by application of q/r. So, we would have a derivation of q but no rule ends in q and so no such demonstration is available. And $\nvDash_{\mathfrak{S}_1} ((p \to q) \to r) \to (p \wedge (q \to r))$ as $S = \{\bar{r}\} \cup \{\bot/a \mid a \text{ atomic}\}$ demonstrates. □

We can give an even simpler example of how the adding of systems can radically alter what a supersystem proves. Consider the supersystems $\{\varnothing\}$, $\{\{p/q\}\}$, $\{\{\bar{p}\}\}$, $\{\{\bar{p}, p/q\}\}$, $\{\varnothing, \{p/q\}, \{\bar{p}\}, \{\bar{p}, p/q\}\}$. Clearly the first four are subsets of the fifth one. And yet no clear relationship holds between what is true on the subsupersystems and the larger system.

	$\vDash_{\{\varnothing\}}$	$\vDash_{\{\{p/q\}\}}$	$\vDash_{\{\{\bar{p}\}\}}$	$\vDash_{\{\{\bar{p},p/q\}\}}$	$\vDash_{\{\varnothing,\{p/q\},\{\bar{p}\},\{p/q,\bar{p}\}\}}$
p	no	no	yes	yes	no
$p \to q$	yes	yes	no	yes	no
$(p \to q) \to p$	no	no	yes	yes	no

This highlights both an interesting feature of supersystems and why they can be so hard to work with.

3.5.2 $\mathfrak{S}_\infty$ Satisfies the Kreisel-Putnam axiom

Piecha, Campos Sanz, and Schroeder-Heister (2015) show that for $\mathfrak{S}_\infty$ we can replace any disjunction-free formula with a rule in $\mathbb{S}$ and vice-versa. This will help us to pick out supersystems which prove sets of formulas. So, we will spell it out in some detail.

Definition 3.20. *We define* rule-formulas *and their associated* levels *inductively as follows:*

- *an atomic formula p is a rule-formula of level* 0
- *where $\varphi_1, \ldots, \varphi_n$ are rule formulas of level $i_1, \ldots, i_n$ and p is an atomic formula, then $(\varphi_1 \wedge \cdots \wedge \varphi_n) \to p$ is a rule-formula of level* $\max(i_1, \ldots, i_n) + 1$.

A formula is *disjunction-free* if the connective $\vee$ does not occur in it.

Lemma 3.22. *(Piecha, Campos Sanz, and Schroeder-Heister 2015) In IPC, every disjunction-free formula is equivalent to a conjunction of rule-formulas.*

Proof. We must show that for every disjunction free formula ψ there are rule-formulas $\varphi_1, \ldots, \varphi_n$ such that $IPC \vdash \psi \equiv \varphi_1 \wedge \cdots \wedge \varphi_n$. Base case: Assume p is an atomic formula then p is a rule-formula and $IPC \vdash p \equiv p$. Induction case: Assume $IPC \vdash \varphi \equiv \varphi_1 \wedge \cdots \wedge \varphi_n$

and $IPC \vdash \psi \equiv \psi_1 \wedge \cdots \wedge \psi_m$. For the inductive step associated to conjunction, note $IPC \vdash \varphi \wedge \psi \equiv \varphi_1 \wedge \cdots \wedge \varphi_n \wedge \psi_1 \wedge \cdots \wedge \psi_m$. There are no inductive steps associated to disjunction since we are working under the assumption of disjunction freeness. The only difficult case is the inductive step associated to $\varphi \to \psi$. We know $IPC \vdash (\varphi \to \psi) \equiv (\varphi_1 \wedge \cdots \wedge \varphi_n \to \psi_1 \wedge \cdots \wedge \psi_m)$ and $IPC \vdash (\chi \to \theta_1 \wedge \theta_2) \equiv ((\chi \to \theta_1) \wedge (\chi \to \theta_2))$. So, it is sufficient to show that each $\varphi_1 \wedge \cdots \wedge \varphi_n \to \psi_i$ is a rule-formula. Note that as $\psi_1 \wedge \cdots \wedge \psi_m$ are rule-formulas we know that either they are atomic formulas or of the form $(\chi_1 \wedge \cdots \wedge \chi_j) \to q$. If the former we are done so assume $\psi_i = (\chi_1 \wedge \cdots \wedge \chi_j) \to q$. But as $IPC \vdash (\theta_1 \to (\theta_2 \to \theta_3)) \equiv (\theta_1 \wedge \theta_2 \to \theta_3)$, it follows that $IPC \vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \to ((\chi_1 \wedge \cdots \wedge \chi_j) \to q) \equiv ((\varphi_1 \wedge \cdots \wedge \varphi_n \wedge \chi_1 \wedge \cdots \wedge \chi_j) \to q)$ which is a rule-formula. $\square$

In what follows if R is of the form p then R/r is p/r and if R is of the form $[R']q/p$ with $[R']$ possibly empty then R/r is $[R'/q]p/r$. With this we associate rules with rule formulas one-to-one as follows:

1. $\bar{p}^* = p$

2. $(p_0, \ldots, p_n/q)^* = p_0 \wedge \cdots \wedge p_n \to q$

3. If $R_{0,0}, \ldots, R_{0,m_0}, \ldots, R_{n,0}, \ldots, R_{n,m_n}$ are associated to $\varphi_{0,0}, \ldots, \varphi_{0,m_0}, \ldots, \varphi_{n,0}, \ldots, \varphi_{n,n_m}$ then

   $(([R_{0,0}, \ldots, R_{0,m_0}]q_0), \ldots, ([R_{n,0}, \ldots, R_{n,m_n}]q_n)/r)^* = (((\varphi_{0,0} \wedge \cdots \wedge \varphi_{0,m_0}) \to q_0) \wedge \cdots \wedge ((\varphi_{n,0} \wedge \cdots \wedge \varphi_{n,m_n}) \to q_n) \to r)$

We associate rule formulas with rules one-to-one as follows:

4. $p^+ = \bar{p}$

5. If $\varphi_0, \ldots, \varphi_n$ are associated to $R_0, \ldots, R_n$ then $((\varphi_0 \wedge \cdots \wedge \varphi_n) \to r)^+ = ((R_0, \ldots, R_n)/r)$.

It can be checked that $\varphi^{+*} = \varphi$ and $R^{*+} = R$. Let $S^* = \{R^* \mid R \in S\}$ and $\Gamma^+ = \{\varphi^+ \mid \varphi \in \Gamma\}$. Given this we have the following result from Piecha, Campos Sanz, and Schroeder-Heister (2015):

Lemma 3.23 (Piecha, Campos Sanz, and Schroeder-Heister 2015, Lemma 2 (C4)). *Let Δ be a set of disjunction-free formulas.*

$$\Gamma, \Delta \vDash_C^{\mathfrak{S}_\infty} \varphi \Leftrightarrow \Gamma \vDash_{C \cup \Delta^+}^{\mathfrak{S}_\infty} \varphi \textit{ and } \Gamma, S^* \vDash_C^{\mathfrak{S}_\infty} \varphi \Leftrightarrow \Gamma \vDash_{C \cup S}^{\mathfrak{S}_\infty} \varphi$$

We have been talking about logics with the disjunction property, but it turns out there is a generalisation of this property, which for PCS semantics implies the Kreisel-Putnam axiom.

Definition 3.21. *A consequence relation $\vDash$ has the generalised disjunction property if for all sets Γ of disjunction free formulas:*

$$\Gamma \vDash \varphi \vee \psi \Leftrightarrow [\Gamma \vDash \varphi \textit{ or } \Gamma \vDash \psi].$$

We can now show that $\mathfrak{S}_\infty$ satisfies the general disjunction property:

Lemma 3.24. *(Piecha and Schroeder-Heister 2019) For all $S \in \mathfrak{S}_\infty$, the consequence relation $\vDash_S^{\mathfrak{S}_\infty}$ has the generalised disjunction property.*

Proof. Let Γ be disjunction free and $\Gamma \vDash_S^{\mathfrak{S}_\infty} \varphi \vee \psi$. By Lemmas 3.12 and 3.13 we know that $\vDash_S^{\mathfrak{S}}$ satisfies IPC, so by Lemma 3.22 we can assume that Γ is a set of rule-formulas. It follows by Equation 3.17 of Definition 3.19 for all $S \in \mathfrak{S}_\infty$ that $\Gamma \vDash_S^{\mathfrak{S}_\infty} \varphi \vee \psi$. By Lemma 3.23, $\vDash_{S \cup \Gamma^+}^{\mathfrak{S}_\infty} \varphi \vee \psi$. So, by Definition 3.19 disjunction property $\vDash_{S \cup \Gamma^+}^{\mathfrak{S}_\infty} \varphi$ or $\vDash_{S \cup \Gamma^+}^{\mathfrak{S}_\infty} \psi$. Which again by Lemma 3.23 gets us $\Gamma \vDash_S^{\mathfrak{S}_\infty} \varphi$ or $\Gamma \vDash_S^{\mathfrak{S}_\infty} \psi$. □

Because for all $S \in \mathfrak{S}_\infty$ it follows that $S \cup \Gamma^+ \in \mathfrak{S}_\infty$ it follows that:

Corollary 3.4. *(Piecha and Schroeder-Heister 2019)* $\mathfrak{S}_\infty$ *has the generalised disjunction property.*

This means that $\mathfrak{S}_\infty$ has the disjunction property (trivially). Piecha and Schroeder-Heister (2019, Lemma 2.1 p. 4) have shown, any PCS semantics with the generalised disjunction property satisfies the generalised Kreisel-Putnam axiom.

Lemma 3.25. *If a PCS semantics has the generalised disjunction property then it satisfies the generalised Kreisel-Putnam axiom.*

Proof. Let $S, S' \in \mathfrak{S}$ and $S \supseteq S'$. Let $\vDash_S^{\mathfrak{S}} \varphi \to \psi_1 \vee \psi_2$ such that φ is disjunction free. It follows that $\varphi \vDash_S^{\mathfrak{S}} \psi_1 \vee \psi_2$ by the weak deduction theorem. By the generalised disjunction property $\varphi \vDash_S^{\mathfrak{S}} \psi_i$ for some $i \in \{1, 2\}$. So $\vDash_S^{\mathfrak{S}} \varphi \to \psi_i$. From which it follows that $\vDash_S^{\mathfrak{S}} (\varphi \to \psi_1) \vee (\varphi \to \psi_2)$. Which by monotonicity implies $\vDash_{S'}^{\mathfrak{S}} [\varphi \to \psi_1 \vee \psi_2] \to [(\varphi \to \psi_1) \vee (\varphi \to \psi_2)]$. □

We now have everything we need in place to define quasi-PTV and so it aligns with inquisitive semantics and show that $\mathfrak{S}_\infty$ aligns with general inquisitive semantics.

3.6 Proof and Consequences of Main Theorem

Our goal is to show the equivalence of the two systems of inquisitive semantics we considered in Section 3.3 with supersystems of proof-theoretic validity. To do this we need to ensure that the supersystem meets the conditions set out in Corollary 3.1 and 3.2. We have already seen that $\mathfrak{S}_\infty$ satisfies the generalised disjunction property and the generalised Kreisel-Putnam axiom. That means the equivalence of $L_{\mathfrak{S}_\infty}$ to generalised inquisitive logic will follow if $\mathfrak{S}_\infty$'s disjunction free fragment is the same as intuitionistic logic's. Further, if we could define a filter on $\mathfrak{S}_\infty$ that satisfies double negation elimination for atomic formulas we would show that the defined system was equivalent to inquisitive logic.

Our first result follows from the following lemma:

Lemma 3.26 (Piecha, Campos Sanz, and Schroeder-Heister 2015, p. 331 Lem. 4)**.** *For all disjunction free* φ*,* $\vDash_{\mathfrak{S}_\infty} \varphi \Leftrightarrow \vdash_{IPC} \varphi$*.*

Proof. Assume Γ, φ are disjunction free and $\Gamma \vDash_S^{\mathfrak{S}_\infty} \varphi$, it follows that $\vDash_{S \cup \Gamma^+}^{\mathfrak{S}_\infty} \varphi$. We will show that $\vDash_S^{\mathfrak{S}_\infty} \varphi \Leftrightarrow S^* \vdash_{IPC} \varphi$. We will not prove the base case: $\vdash_S p \Leftrightarrow S^* \vdash_{IPC} p$. The proof is an induction of the level of atomic rules which is long but not difficult. The conjunction case is obvious. The implication case follows via: $\vDash_S \varphi \to \psi \Leftrightarrow \varphi \vDash_S \psi \Leftrightarrow \vDash_{S \cup \{\varphi^+\}} \psi \Leftrightarrow S^* \cup \{\varphi^{+*}\} \vdash_{IPC} \psi \Leftrightarrow S^* \cup \{\varphi\} \vdash_{IPC} \psi \Leftrightarrow S^* \vdash_{IPC} \varphi \to \psi$. □

From which we get via Corollary 3.2:

Theorem 3.4. $L_{\mathfrak{S}_\infty} = L_{GInq} = IPC + GKP$.[6]

What remains is to try and offer a proof-theoretic validity notion for inquisitive logic. It turns out we have a general mechanism for defining new supersystems which prove sets of disjunction-free formulas:

Definition 3.22. *Given a set of disjunction free formulas* Δ *let* $\mathfrak{S}_\Delta = \{S \in \mathfrak{S}_\infty \mid \Delta^+ \subseteq S\}$

Lemma 3.27. *For any set of disjunction free formulas* Δ *it follows that* $\mathfrak{S}_\Delta$ *is a filter on* $\mathfrak{S}_\infty$ *and* $\vDash_{\mathfrak{S}_\Delta} \Delta$*.*

Proof. Clearly $\mathfrak{S}_\Delta$ is a filter on $\mathfrak{S}_\infty$ so by Lemma 3.20 we have for all $S \in \mathfrak{S}_\Delta$ that $\vDash_S^{\mathfrak{S}_\infty} \Delta \Rightarrow \vDash_S^{\mathfrak{S}_\Delta} \Delta$. Note that for all $S \in \mathfrak{S}_\infty$, $\Delta \vDash_S^{\mathfrak{S}_\infty} \Delta$. So, by Lemma 3.23 it follows that

[6] I would like to thank an anonymous reviewer for pointing me towards Punčochář 2016 and suggesting this result.

$\vDash^{\mathfrak{S}_\infty}_{S\cup\Delta^+} \Delta$. Let $S \in \mathfrak{S}_\Delta$, it follows that $S \in \mathfrak{S}_\infty$ and that $S = S \cup \Delta^+$. So $\vDash^{\mathfrak{S}_\infty}_{S\cup\Delta^+} \Delta$ which implies $\vDash^{\mathfrak{S}_\Delta}_{S\cup\Delta^+} \Delta$ and so $\vDash^{\mathfrak{S}_\Delta}_{S} \Delta$. From which it follows that $\vDash_{\mathfrak{S}_\Delta} \Delta$. □

If we consider the set of formulas true on a particular $\mathfrak{S}_\Delta$ sometimes this will be the set of all formulas because $\bot$ is among them. Of the consistent sets, many will not be closed under substitution. Some of these will include formulas which cannot be proven in classical logic, for example, $\mathfrak{S}_{\{p\}}$. This makes it clear that these systems can be weak logics that are either intermediate or stronger than classical logic.[7]

We have now collected all the necessary pieces to define quasi-PTV.

Definition 3.23. *Define $\mathfrak{S}_Q$ (Q for Quasi) based on $\mathfrak{S}_\infty$ as follows:*

$$\mathfrak{S}_Q = \mathfrak{S}_{\{\neg\neg p \to p | p \ atomic\}} = \{S \in \mathfrak{S}_\infty \mid ([p/\bot]\bot/p) \in S \text{ for all atomic } p\} \tag{3.18}$$

As $\mathfrak{S}_Q$ is a filter on $\mathfrak{S}_\infty$ we can show it satisfies the Kreisel-Putnam axiom as $\mathfrak{S}_\infty$ does. This is the last piece needed to show that $\mathfrak{S}_Q$ satisfies a condition from Corollary 3.1.

Lemma 3.28. *$\mathfrak{S}_Q$ is a filter on $\mathfrak{S}_\infty$ and $\mathfrak{S}_Q$ satisfies both the Kreisel-Putnam axiom and $\neg\neg p \to p$ for all atomic p.*

Proof. As $\mathfrak{S}_Q = \mathfrak{S}_{\{\neg\neg p \to p | \text{patomic}\}}$ it follows by Lemma 3.27 that $\mathfrak{S}_Q$ is a filter on $\mathfrak{S}_\infty$ and $\vDash_{\mathfrak{S}_Q} \neg\neg p \to p$ for all atomic p. As $\mathfrak{S}_Q$ is a filter on $\mathfrak{S}_\infty$ it follows by Lemma 3.20 that $\mathfrak{S}_Q$ proves everything $\mathfrak{S}_\infty$ does. By Lemma 3.24 and Lemma 3.25 it follows that $\mathfrak{S}_\infty$ satisfies the Kreisel-Putnam axiom and so, so does $\mathfrak{S}_Q$. □

This proof shows that $\mathfrak{S}_Q$ defines a logic $L_{\mathfrak{S}_Q} = \{\varphi \mid \vDash_{\mathfrak{S}_Q} \varphi\}$ which meets the requirements found in Corollary 3.1. By Lemmas 3.4 and 3.7 we have that $L_{\mathfrak{S}_Q}$ is the same logic as inquisitive logic. That is:

[7] They cannot be classical as they all have the disjunction property which CPL does not.

Theorem 3.5. $L_{\mathfrak{S}_Q} = L_{Inq}$

As we are working with weak logics we can ask the question what logic characterises the fragment closed under substitution. To do this we define a logics schematic fragment.

Definition 3.24. *If L is a weak logic, its schematic fragment is:*

$$Schm(L) = \{\varphi \in L \mid \text{ for all atomics } p_1, \ldots, p_n \text{ in } \varphi \text{ and all } \varphi_1, \ldots, \varphi_n,\ \varphi[p_1/\varphi_1, \ldots, p_n/\varphi_n] \in L\}$$

That is $Schm(L)$ is every formula of L for which substitution is valid. If L is a weak intermediate logic then $Schm(L)$ is an intermediate logic. It follows from these results that the schematic fragment of both logics is Medvedev's logic.

Definition 3.25. *Medvedev's logic of finite problems ML is defined semantically as the logic of all finite frames of the form $(\mathcal{P}(X) - \{X\}, \subseteq)$.*

Ciardelli and Roelofsen (2011) have shown that the schematic fragment of L_{Inq} is Medvedev's logic of finite problems discussed above, which is not known to be axiomatizable. Punčochář (2016) has generalised this result. From this it follows that: $Schm(L_{\mathfrak{S}_Q}) = Schm(L_{\mathfrak{S}_\infty}) = ML$.

3.7 Conclusion

To summarise: We have seen that inquisitive semantics has the disjunction property, the generalised Kreisel-Putnam axiom and double negation elimination for atomic formulas, while general inquisitive semantics has the disjunction property, the generalised Kreisel-Putnam axiom and shares a disjunction free fragment with IPC. We have generated a system of PTV, namely quasi-PTV or $\mathfrak{S}_Q$, which is extensionally equivalent to inquisitive semantics and have shown that PTV or $\mathfrak{S}_\infty$ is extensionally equivalent to general inquisitive semantics.

This answers an important question about proof-theoretic validity as presented in the works of Piecha, de Campos Sanz, and Schroeder-Heister. It also highlights a fascinating connection between inquisitive semantics and proof-theoretic semantics. Despite very different motivations, this connection suggests that inquisitive semantics can be given a constructive justification and that proof-theoretic validity might have wider applicability than previously though.

It remains open what logic is captured by the $\mathfrak{S}_n$s for $n \in \mathbb{N}$. These logics are going to be fewer well behaviours than the ones considered here because they will have the generalised Kreisel-Putnam axiom for formulas only below a certain complexity. They may also validate additional axioms as is seen by $\mathfrak{S}_1$'s having double negation elimination for atomic formulas (Piecha 2016).

Bibliography

Beaney, Michael (1997). *The Frege Reader*. Wiley-Blackwell.

Bell, John L. (1999). “Frege’s Theorem in a Constructive Setting”. In: *Journal of Symbolic Logic* 64.2, pp. 486–488.

Belnap, Nuel D (1962). “Tonk, Plonk and Plink”. In: *Analysis* 22.6, pp. 130–134.

Benacerraf, Paul (1965). “What Numbers Could not Be”. In: *Philosophical Review* 74.1, pp. 47–73.

Bolzano, Bernard (1973). *Theory of Science*. Ed. by Jan Berg. Dordrecht: Riedel.

Boolos, George (1986). “Saving Frege from Contradiction”. In: *Proceedings of the Aristotelian Society* 87, pp. 137–151.

— (1997). “Is Hume’s principle analytic?” In: *Language, Thought and Logic*. Ed. by Richard Kimberly Heck. Oxford: Clarendon Press, pp. 245–261.

— (1998). *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press.

Burgess, John (1984). “Review of wright (1983)”. In: *Philos. Rev.* 93.1984, pp. 638–640.

Burgess, John P. (2005). *Fixing Frege*. Princeton: Princeton University Press.

Button, Tim and Sean Walsh (2018). *Philosophy and Model Theory*. Oxford: Oxford University Press.

Ciardelli, Ivano, Jeroen Groenendijk, and Floris Roelofsen (2018). *Inquisitive Semantics*. Oxford Surveys in Semantics and Pragmatics. Oxford: Oxford University Press.

Ciardelli, Ivano and Floris Roelofsen (Feb. 2011). “Inquisitive Logic”. In: *Journal of Philosophical Logic* 40.1, pp. 55–94.

Cook, Roy T (2007). "Introduction". In: *The Arché Papers on the Mathematics of Abstraction*. Ed. by Roy T Cook. Vol. 71. The Western Ontario Series in Philosophy of Science. Berlin: Springer, pp. xv–xxxvii.

Demopoulos, William (June 1994). "Frege and the rigorization of analysis". In: *J. Philos. Logic* 23.3, pp. 225–245.

Dummett, Michael (1991a). *Frege: Philosophy of Mathematics*. en. Harvard University Press.

— (1991b). *The Logical Basis of Metaphysics*. Cambridge MA: Harvard University Press.

Feferman, Solomon (2005). "Predicativity". In: *The Oxford Handbook of Philosophy of Mathematics and Logic*. Ed. by Stewart Shapiro. Oxford: Oxford University Press, pp. 590–624.

Fitting, Melvin and Richard L. Mendelsohn (1998). *First-Order Modal Logic*. Dordrecht: Kluwer Academic Publishers.

Frege, Gottlob (1884). *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl.* (translated as *The Foundations of Arithmetic: A logico-mathematical enquiry into the concept of number*, by J.L. Austin, Oxford: Blackwell, second revised edition, 1974.) Breslau: W. Koebner.

— (1893). *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet.* (translation as *Basic Laws of Arithmetic: Derived using concept-script* by P. Ebert and M. Rossberg (with C. Wright), Oxford: Oxford University Press, 2013.) Jena: Verlag Hermann Pohle.

Gentzen, G. (1935). "Untersuchungen über das logische Schließen". In: *Mathematische Zeitschrift* 39, pp. 176–210, 405–431.

Goldfarb, Warren (2016). "On Dummett's "Proof-theoretic justifications of logical laws"". In: *Advances in proof-theoretic semantics*. Springer, Cham, pp. 195–210.

Groenendijk, Jeroen and M J B Stokhof (1984). "Studies on the Semantics of Questions and the Pragmatics of Answers". PhD thesis.

Hájek, Petr and Pavel Pudlák (1998). *Metamathematics of First-Order Arithmetic*. Perspectives in Mathematical Logic. Berlin: Springer.

Hale, Bob and Crispin Wright (2001a). *The Reason's Proper Study*. Oxford: Oxford University Press.

— (2001b). *To Bury Caesar* .

Heck, Richard Kimberly (1993). "The development of arithmetic in Frege's Grundgesetze der Arithmetik". In: *Journal of Symbolic Logic* 58.2. (originally published under the name "Richard G. Heck, Jr"), pp. 579–601.

Heijenoort, Jean van (1967). *From Frege to Gödel : A Source Book in Mathematical Logic, 1879-1931*. Cambridge: Harvard University Press.

Hodes, Harold (1984). "Logicism and the Ontological Commitments of Arithmetic". In: *The Journal of Philosophy* 81.3, pp. 123–149.

— (1990a). "Ontological Commitment: Thick and Thin". In: *Meaning and Method: Essays in Honor of Hilary Putnam*. Cambridge: Cambridge University Press, pp. 235–260.

— (1990b). "Where Do The Natural Numbers Come From?" In: *Synthese* 84.3, pp. 347–407.

— (1991). "Where Do Sets Come From?" In: *The Journal of Symbolic Logic* 56.1, pp. 150–175.

Hodges, Wilfrid (Mar. 1993). *Model Theory*. Cambridge: Cambridge University Press.

Humberstone, Lloyd (2011). *The Connectives*. en. MIT Press.

Karttunen, Lauri and Stanley Peters (1980). "Interrogative quantifiers". In: *Time, tense, and quantifiers* 83, pp. 181–205.

Kim, Joongol (2015). "A Logical Foundation of Arithmetic". In: *Studia Logica* 103.1, pp. 113–144.

Kleene, Stephen Cole (1952). *Introduction to metamathematics*. Vol. 483. New York: van Nostrand.

Kocurek, Alexander W. (2016). "The Problem of Cross-world Predication". In: *Journal of Philosophical Logic* 45.6, pp. 697–742.

Kolmogorov, A. N. (1932). "Zur Deutung der intuitionistischen Logik". In: *Mathematische Zeitschrift* 35, pp. 58–65.

Linnebo, Øystein (2013). "The Potential Hierarchy of Sets". In: *The Review of Symbolic Logic* 6.2, pp. 205–228.

— (2018). *Thin objects: An abstractionist account*. Oxford: Oxford University Press.

Linnebo, Øystein and Stewart Shapiro (Mar. 2019). "Actual and Potential Infinity". In: *Noûs* 53.1, pp. 160–191.

Maksimova, Larisa (1986). "On maximal intermediate logics with the disjunction property". In: *Studia Logica* 45, pp. 69–75.

Miglioli, P. et al. (1989). "Some results on intermediate constructive logics". In: *Notre Dame Journal of Formal Logic* 30(4), pp. 543–562.

Mostowski, Marcin (2001). "On Representing Concepts in Finite Models". In: *Mathematical Logic Quarterly* 47.4, pp. 513–523.

Negri, Sara and Jan von Plato (July 2008). *Structural Proof Theory*. en. Cambridge: Cambridge University Press.

Parsons, Charles (1983). "Sets and Modality". In: *Mathematics in Philosophy: Selected Essays*. Ithaca: Cornell University Press, pp. 298–341.

Piecha, Thomas (2016). "Completeness in Proof-Theoretic Semantics". In: *Advances in Proof-Theoretic Semantics*. Ed. by Thomas Piecha and Peter Schroeder-Heister. Cham: Springer International Publishing, pp. 231–251.

Piecha, Thomas, Wagner de Campos Sanz, and Peter Schroeder-Heister (June 2015). "Failure of Completeness in Proof-Theoretic Semantics". In: *Journal of Philosophical Logic* 44.3, pp. 321–335.

Piecha, Thomas and Peter Schroeder-Heister (Feb. 2019). "Incompleteness of Intuitionistic Propositional Logic with Respect to Proof-Theoretic Semantics". In: *Studia Logica* 107.1, pp. 233–246.

Prawitz, Dag (1971). "Ideas and Results in Proof Theory". In: *Proceedings of the Second Scandinavian Logic Symposium*. Ed. by J.E. Fenstad. Vol. 63. Studies in Logic and the Foundations of Mathematics. Elsevier, pp. 235–307.

Prawitz, Dag (1973). “Towards A Foundation of A General Proof Theory”. In: *Proceedings of the Fourth International Congress for Logic, Methodology and Philosophy of Science, Bucharest, 1971*. Ed. by Patrick Suppes et al. Vol. 74. Studies in Logic and the Foundations of Mathematics Supplement C. Elsevier, pp. 225–250.

— (2006). “Meaning Approached via Proofs”. In: *Synthese* 148.3, pp. 507–524.

— (2014). “An Approach to General Proof Theory and a Conjecture of a Kind of Completeness of Intuitionistic Logic Revisited”. In: *Advances in Natural Deduction: A Celebration of Dag Prawitz’s Work*. Ed. by Luiz Carlos Pereira, Edward Hermann Haeusler, and Valeria de Paiva. Dordrecht: Springer Netherlands, pp. 269–279.

Prior, A N (Dec. 1960). “The Runabout Inference-Ticket”. In: *Analysis* 21.2, pp. 38–39.

Punčochář, Vıt (Aug. 2016). “A Generalization of Inquisitive Semantics”. In: *J. Philos. Logic* 45.4, pp. 399–428.

Putnam, Hilary (1967a). “Mathematics Without Foundations”. In: *Journal of Philosophy* 64.1. reprinted in Putnam 1979, pp. 5–22.

— (1967b). “The Thesis that Mathematics is Logic”. In: *Bertrand Russell: Philosopher of the Century*. London: Allen and Unwin.

Read, Stephen (2010). “General-Elimination Harmony and the Meaning of the Logical Constants”. In: *J. Philos. Logic* 39.5, pp. 557–576.

Russell, Bertrand (1903). *The Principles of Mathematics*. Cambridge: Cambridge University Press.

— (1967). “Letter to Frege”. In: *From Frege to Gödel*. Ed. by van Heijenoort. Cambridge, MA: Harvard University Press, pp. 124–125.

Schroeder-Heister, Peter (1984). “A Natural Extension of Natural Deduction”. In: *J. Symbolic Logic* 49.4, pp. 1284–1300.

— (Feb. 2006). “Validity Concepts in Proof-theoretic Semantics”. In: *Synthese* 148.3, pp. 525–571.

Schroeder-Heister, Peter (2018). “Proof-Theoretic Semantics”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2018. Metaphysics Research Lab, Stanford University.

Shapiro, Stewart and Øystein Linnebo (2015). “Frege Meets Brouwer”. In: *Review of Symbolic Logic* 8.3, pp. 540–552.

Simpson, Steve (2009). *Subsystems of Second Order Arithmetic*. Perspectives in Logic. Cambridge: Cambridge University Press.

Stalnaker, Robert C (1976). “Possible Worlds”. In: *Noûs* 10.1, pp. 65–75.

Stanley, Jason (1997). “Names and Rigid Designation”. In: *A Companion to the Philosophy of Language*. Ed. by Bob Hale and Crispin Wright. Malden: Blackwell, pp. 555–585.

Steinberger, Florian (2011). “What Harmony Could and Could Not Be”. In: *Australasian Journal of Philosophy* 89.4, pp. 617–639.

Studd, JP (2016). “Abstraction Reconceived”. In: *The British Journal for the Philosophy of Science* 67.2, pp. 579–615.

Tarski, Alfred (1956). *Logic, Semantics, and Metamathematics*. Oxford: Clarendon Press.

Troelstra, A. S. and Dirk van Dalen (1988). *Constructivism in Mathematics. An Introduction. Volume 2.* Amsterdam: Elsevier.

Troelstra, A. S. and H. Schwichtenberg (2000). *Basic Proof Theory*. Second. Vol. 43. Cambridge Tracts in Theoretical Computer Science. Cambridge: Cambridge University Press.

Urbaniak, Rafal (2016). “Potential infinity, abstraction principles and arithmetic (Leniewski Style)”. In: *AXIOMS* 5.2, p. 20.

Walsh, Sean (2012). “Comparing Peano arithmetic, Basic Law V, and Hume’s Principle”. In: *Ann. Pure Appl. Logic* 163.11, pp. 1679–1709.

— (2016). “The Strength of Abstraction with Predicative Comprehension”. In: *Bull. Symb. Log.* 22.1, pp. 105–120.

Wehmeier, Kai F (1999). “Consistent fragments of Grundgesetze and the existence of non-logical objects”. In: *Synthese* 121.3, pp. 309–328.

Wehmeier, Kai F (2004). "Russell's paradox in consistent fragments of Frege's Grundgesetze der Arithmetik". In: *One hundred years of Russell's paradox* 6, pp. 247–257.

Whitehead, Albert and Bertrand Russell (1910). *Principia Mathematica.* Vol. 1. Cambridge: Cambridge University Press.

Williamson, Timothy (2013). *Modal Logic as Metaphysics*. en. Oxford: Oxford University Press.

— (June 2020). *Suppose and Tell: The Semantics and Heuristics of Conditionals*. en. Oxford University Press.

Wright, Crispin (1984). "Frege's Conception of Numbers as Objects". In: *J. Philos.* 81.12, pp. 778–783.