Logical Validity



Marco Grossi
New College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary 2022

To my father

Acknowledgements

Thank you mum for the unconditional love and support. I would not be here if it was not for you. Thank you sister for cheering me up. Thank you Anna for your time and endless patience. Thank you Volker and Tim for all the help and for the invaluable philosophical insights you gave me. Thank you coffee for keeping me awake. Thank you New College for being so gorgeous.

Abstract

This dissertation is about logical consequence. The main points I will defend can be summarised as follows. First, I claim that logical consequence can be understood as truth-preservation under all interpretations of the language. There is no need to use modal notions to define consequence. In chapter one, I argue that domain-variation across different interpretations can be understood in purely interpretational terms, if we take quantifiers to be context-sensitive expressions.

Secondly, I am sympathetic with a substitutional account of logical consequence, where consequence is preservation of truth under all substitutions of the non-logical vocabulary. In chapter two, I extend the substitutional account to modal logic, where $\Box$ is read as 'It is substitutionally valid that'. Modal logic is seen as a way for the system to reflect upon what is valid. Different readings of $\Box$ can be provided by restricting the substitutions relevant for the truth of a modal formula.

In the third chapter, I argue that a theory of logical consequence ought to be self-applicable: it should apply to its own language. This is required by our use of logical consequence in philosophical logic, to express the debate between the monist and the pluralist about logical consequence, and to link our theory of what is valid with a theory of valid reasoning and justification.

Finally, in the fourth chapter I suggest that a theory of interpretations of a given language will never really capture all the interpretations of that language. I extend the worry even to theories that admit absolute generality and take higher-order talk to be non-reducible to first-order talk. I argue that, regardless of the order of the theory one is speaking from, there is always an order higher that one can take as one's starting point, and at any further order up the hierarchy, more interpretations of any language become available.

Contents

1 Introduction **1**

1.1 Quantification in the interpretational theory of validity 1

1.2 Substitutional validity for modal logic 3

1.3 Should theories of validity self-apply? 6

1.4 Higher order semantics and non-standard interpretations 8

2 Quantification in the Interpretational Theory of Validity **10**

2.1 Interpretations and logical validity 10

2.2 The cardinality problem 12

2.2.1 Contingency? 13

2.2.2 Asymmetry in information 14

2.3 Attempted solutions 15

2.3.1 Logical expressions: serious vs pragmatic stance 18

2.4 The context-sensitive account of the quantifiers 20

2.4.1 A methodological remark 20

2.4.2 The proposal 21

2.5 Ramifications and objections 25

2.5.1 Meta-context and generality 25

2.5.2 The Socrates problem 28

2.5.2.1 Free logic route 28

2.5.2.2 Kleene and the classical logic route 30

2.5.3 Pragmatic vs semantic uses of contexts 32

2.6 Conclusion 34

3 Substitutional Validity for Modal Logic **35**

3.1 Philosophical context of this work 35

3.1.1 Simple truth vs truth in a model 35

3.1.2 Advantages of the axiomatic approach 36

3.1.3 Intensional languages . . . 38
3.1.4 A proposal . . . 40
3.1.5 Why I don't choose a non-reductive account . . . 42
3.2 Technical preliminaries . . . 43
3.2.1 Language . . . 43
3.2.2 Substitution function . . . 45
3.2.3 Satisfaction . . . 47
3.2.4 Modal truth theories . . . 48
3.3 Substitutional logical consequence . . . 50
3.3.1 Logical necessity . . . 52
3.4 Soundness . . . 53
3.4.1 Soundness for K . . . 54
3.4.2 Soundness for other normal modal calculi . . . 55
3.5 Completeness . . . 57
3.5.1 Canonical model and adequacy . . . 57
3.6 Quantification . . . 64
3.6.1 Barcan and converse . . . 64
3.6.2 Adequacy in a variable domain semantics . . . 66
3.6.3 Adequacy for calculi without BF or CBF . . . 66
3.6.4 Adequacy for logic with BF or CBF . . . 70

4 Should Theories of Validity Self-apply? 72
4.1 Self-applicability and generality . . . 72
4.1.1 Standard set-based semantics is not general . . . 74
4.2 Self-applicability in trouble . . . 76
4.2.1 Russellian argument . . . 76
4.2.2 Size restrictions . . . 78
4.2.3 Typing . . . 79
4.2.4 Cantorian argument . . . 80
4.3 Arguments for self-applicability . . . 81
4.3.1 Cross order universality and semantic luminosity . . . 81
4.3.2 Argument from truth . . . 82
4.3.2.1 Meaning and content . . . 84
4.3.2.2 Truth vs validity . . . 86
4.3.3 Expressibility and cross order universality . . . 86
4.3.3.1 Self-applicability and normative constraints . . . 88

4.3.3.2 Schemata and universality . . . 89
4.4 The untyped theorist's perspective . . . 91
4.4.1 Against A1 . . . 91
4.4.2 An alternative test . . . 94
4.4.3 Re-evaluating the issue at hand . . . 96

5 Higher Order Semantics and Non-standard Interpretations 99
5.1 Validity, generality and ascent . . . 99
5.1.1 Semantic ascent . . . 101
5.2 Standard higher order semantics . . . 103
5.2.1 Meta-language and meta-theory . . . 103
5.2.1.1 The higher orderist's framework . . . 106
5.2.1.2 Why cumulative? . . . 108
5.2.1.3 Plurals and higher order logic . . . 109
5.2.2 A sketch of standard higher order semantics and its virtues . . . 111
5.2.2.1 Semantic ascent . . . 113
5.3 Non-standard and raising interpretations . . . 114
5.3.1 Semantic structure . . . 114
5.3.2 Raising interpretations are models of higher order logic . . . 117
5.3.3 In raising interpretations the role of an expression is exclusive 117
5.3.4 Semantic roles and intended readings . . . 119
5.3.5 Cardinality of orders and reduction . . . 121
5.3.6 Logical expressions . . . 123
5.3.6.1 Types and extensions of logical expressions in raising semantics . . . 126
5.4 Features of raising semantics . . . 128
5.4.1 Henkin models vs raising interpretations . . . 128
5.4.2 Raising semantic ascent . . . 130
5.4.3 Extensional equivalence? . . . 132

A Formal Notes on Higher Order Semantics and Non-standard Interpretations 136
A.1 Meta-language and meta-theory . . . 136
A.1.1 Syntax . . . 136
A.1.2 Meta-theory . . . 137
A.1.3 Finite higher order sequences . . . 140
A.1.3.1 Uses . . . 142

A.2 Standard, Higher order Semantics 144
A.2.1 Coding 145
A.2.2 Interpretations and assignments 145
A.2.3 Standard satisfaction and standard validity 147
A.2.4 Virtues of the standard semantics 148
A.3 Raising higher order semantics 150
A.3.1 Soundness 152
A.3.2 Indefinite raisability of the logic of order n. 154
A.3.3 Categoricity and incompleteness 157
A.3.4 Other 161

Bibliography **163**

Chapter 1

Introduction

This thesis is about logical validity. If logic can mainly be divided in proof theory and semantics, logical validity is the main concept on the semantic side. Logical validity is at the crossroad of other fundamental philosophical notions, like truth and meaning. Different views on validity will be shaped by how one answers the following questions: What is truth? What does it mean to interpret a language? Do we need to provide a determined domain to interpret quantification? How should we interpret higher order quantification? This thesis is a journey into the intricate space centred around the concept of validity. It is divided in four stand-alone articles on different topics: modal logic, higher order logic, context sensitivity and paradoxes of self-applicable truth and interpretations. I now give an overview of each article.

1.1 Quantification in the interpretational theory of validity

According to a common view, logical validity is preservation of truth under all suitable interpretations. An interpretation is suitable only if it is compatible with the intended meanings of the logical expressions. We usually identify logical expressions with the connectives, the universal and existential quantifier, and identity. A common objection to the interpretational account is that it overgenerates (Etchemendy 1990, Hanson 1997, Shapiro 1998). Cardinality statements like 'There are n things' can be written using only logical vocabulary. If an interpretation is 'suitable' only if it is compatible with the intended meanings of the logical expressions, then one could conclude that there is no non-trivial way of reinterpreting statements like 'There are n things', given any positive n. So, the sentence is true in all suitable interpretations if true under the intended interpretation; it is true since there are n things, so it ought

to be logically true. Yet, the standard view is that it is not a logical truth that there are eleven things, or twelve for that matter; so there must be something wrong with the interpretational account.

This first article attempts to defend the interpretational view. First, I try to clarify what the problem is: why is it that 'There are eleven things' is not logically true? Often, it is said that it is contingent that there are eleven things. Yet, how can it be contingent? It is a truth of arithmetic that there are eleven things, and it is commonly accepted that mathematical truths are not contingent. I claim that the *contingency* at hand seems to be more about 'information asymmetry'. Logical validity seems to follow information containment – if ϕ follows from ψ, then the information in ϕ is 'contained' in ψ. The information in 'There are eleven things' does not seem to be *contained* in the information in 'There are ten things', while the opposite does seem to hold. Hence, the former does not follow from the latter, only the latter from the former. However, if both statements are logically true, they should logically co-imply each other.

The objection to the interpretational account is that, if quantifiers are logical, it is unclear how we should make sense of variations in the domain of quantification across interpretations. If 'Something' always means something, then it should always range over whatever exist, without restriction. I claim that domain variations can be made sense of in interpretational terms, if we take quantifiers to be indexical expressions. The range of quantifiers can vary from context to context. The theory is still interpretational because a context is needed to interpret indexical expressions. Consider 'Today is Saturday'. I cannot begin to interpret this sentence nor can I tell if it expresses something true, unless I am given a context with a time. Similarly, if quantifiers are indexical, I do not know what 'There are eleven things' says unless I am given a context, which supplies a range to the quantifiers. In the context where all that is relevant are the shoes in my closet, there are less than eleven relevant things, so 'There are eleven things' expresses something false in that context: it is false under some (contextual) interpretation.

Whether the content of 'There are eleven things' expresses something contingent or not depends on the context: in the context of arithmetic it does not, since it expresses the mathematical truth that there are eleven numbers; in the context of me pondering about the shoes in my closet it does: the sentence expresses something only contingently false. The asymmetry in information between the two statements is accounted for: in any context, the information I gather from the content of 'There

are ten things' is contained in the one I gather from the content of 'There are eleven things', but not vice-versa.

I discuss some ramifications of my theory. I claim that it has advantages over rival solutions to the issue, since it does not rely on any change in what we usually take logical expressions to be, nor does it need to change the definition of validity, by relying on 'representational' elements like possible worlds.[1] I claim that my proposal is compatible with absolute generality as it is neutral on the 'size' of contexts: that quantification is contextual does not imply that there is no absolutely general context.

I finally discuss what logic we naturally get from my account. 'Socrates is human' does not seem to be context-sensitive: it always means something true – that Socrates is human. 'Something is human', on the other hand, means different things in different contexts where different things are relevant – sometimes it might express something false. So, the derivation from 'Socrates is human' to 'Something is human' is not valid. I believe free logic is the most suitable for my framework, even though classical logic can be made to fit, too.

1.2 Substitutional validity for modal logic

Validity is usually defined in terms of models: validity is preservation of truth in all models. Models are usually built as sets of a particular kind, in Zermelo-Fraenkel set-theory with choice. Historically, however, there are different traditions that propose alternative notions of validity. An alternative notion is substitutional: substitutional validity is preservation of truth under all substitutions of the non-logical vocabulary.

Often, substitutional validity is built using model-theoretic machinery: truth is centred on a specific model, and then validity is defined as preservation of truth *in that model* under all substitutions. I am interested in substitutional validity as an alternative to model-theory. Substitutional validity does not need any notion of model to work: it just needs a notion of simple truth, which can be axiomatically defined. This strategy has many advantages:

1. Due to set-theoretic paradoxes, there is no intended set-based model where truth is equivalent simple truth: truth that satisfies the usual Tarskian axioms (1933). So, we are not able to 'recover' truth from the notion of truth in a model;

[1] The term is from Etchemendy 1990.

2. In the axiomatic approach, quantifiers are absolutely general if the meta-theory is absolutely general. There is no need to invoke higher order resources to provide a model where quantifiers are unrestricted, nor is it necessary to digest the idea that higher order quantifiers are not really reducible to first order quantifiers. A theory of truth where quantifiers are read homophonically is enough;

3. It is not obvious in model-theory that logical validity is truth-preserving, since there is no intended model. Reflection principles and completeness results can be invoked to show it for first order logic, but they are usually not available for incomplete logics (under standard semantics). In the axiomatic approach to substitutional validity, the link between validity and truth is obvious: if ϕ is logically true, it is true under all substitutions, therefore it is true under the identity substitution, so it is true.

In the controversial history of the substitutional account of validity, not many have tried to extend the account beyond extensional, first level languages. In this article I develop a substitutional theory of validity for modal logics. First of all, I delimit the range of my proposal. A substitutional definition of validity for a logic of action or a deontic logic does not make much sense; an extension to modal logic makes sense only if $\Box$ is interpreted in a certain way. I propose the following reading: $\ulcorner\Box\phi\urcorner$ is true exactly when it is substitutionally valid. I add this axiom to a simple type-free theory of truth. The definition of substitutional validity does not change from non-modal to modal languages: it is still preservation of truth under all suitable substitutions. Since $\Box$ means substitutional validity, it is sensible to use a substitutional account of validity that treats it as a logical expression.

The theory has interesting properties, when compared to the usual Kripke semantics. The problems highlighted above about set-based semantics exacerbate in modal logic. Suppose we try interpret $\Box$ as 'it is Kripke-valid that' in Kripke semantics. Points in a model represents interpretations: it is valid that ϕ is true at an interpretation if true at all interpretations. Here are some issues:

1. There is no 'intended' Kripke model, since there are as many Kripke models as there are sets, but only set-sized points in any Kripke model. To show that there is a model that is at least equivalent to the intended one for $\Box$, we need to rely on the logic being canonical. If we change the reading of $\Box$ to truth in some subset of Kripke models, whether there is an intended Kripke model for $\Box$ will depend on the characteristics of the logic generated;

2. For set-theoretic reasons there is no point at any Kripke model where truth means truth under the intended interpretation. It is not clear, then, why validity should be truth preserving. It is also not clear that reflection principles for ZFC can be of any help, here, since ZFC is not a modal language;

3. There are other links between $\Box$, truth and validity that should be there but they are missing. If $\Box$ means 'It is valid that', then it should be obvious that ϕ is logically true exactly when 'It is valid that ϕ' is true. It should also be obvious that if 'It is valid that ϕ' is true then ϕ is true. None of this can be shown in standard Kripke semantics.[2]

On the other hand, in my system all the features above can be shown just by looking at the definitions. About (1): the 'intended' reading of $\Box$ as substitutional validity is the only one we have, by definition. About (2): (substitutional) validity is obviously truth-preserving because it implies truth under the identity substitution. About (3): $\Box\phi$ is true exactly when it is true under all substitutions, i.e. when ϕ is valid; $\Box\phi$ is true only if ϕ is logically true, which happens only if ϕ is true.

I further specify my proposal and explore its limits in the rest of the paper. First, using a similar proposal, one can specify different truth-theories that differ in their understanding of $\Box$. A truth-theory may pick a specific subset of the substitutions and interpret the $\Box$ as truth in only those substitutions. Substitutional validity has different extensions in these different truth-theories. For example, if the null set is chosen, the schema $\Box\phi$ will be valid; if, for any ϕ, a relevant substitution swaps it with an equivalent ψ, then $\Box\phi \rightarrow \phi$ will be valid. I discuss some general soundness conditions for common modal calculi like K, T, B, S4 and S5.

Several questions can be asked: is there a truth-theory where validity is adequate (i.e. sound and complete) for the usual modal logics dicussed in the literature, like K, S4, S5 etc? I prove that for any canonical modal logic (with Barcan and converse) there is a truth-theory where validity is adequate for that logic. However, in order to adapt filtration methods to prove completeness for Kripke complete modal logics that are not canonical, we would likely need to change our definition of validity. A simple way is to define a notion of super-validity as validity in all truth-theories of a certain class. However, I do not believe that this notion of super-validity is in line with the substitutional account.

[2]Similar argument works when we try to model 'It is substitutionally valid that' in Kripke semantics, since there is no intended model and truth cannot be recovered from the concept of truth in a model.

I also discuss variable domain semantics. Variable domain semantics does not fit in well within a substitutional account. Barcan and converse become banalities when we read $\Box$ as 'It is substitutionally valid that' – everything satisfies all the relevant substitutions of a condition ϕ exactly when 'Everything is ϕ' is satisfied under all relevant substitutions. What if we wished to provide an adequate notion of substitutional validity for modal logic without Barcan or converse, anyway? I explore different ways of doing it; most of them require *ad hoc* changes in the definition of validity. My conclusion is that a constant domain semantics makes much more sense under a substitutional understanding of validity.

1.3 Should theories of validity self-apply?

A theory of validity self-applies if it applies to its own language. The standard model-theoretic definition of validity in set-theory self-applies: it is defined for all first level languages, in particular for the language of set-theory. Some philosophers have argued against it, though. They claim that a theory of validity for a given language L ought to be general: it ought to capture all suitable interpretations of L. They claim that generality and self-applicability are inconsistent, due to some paradoxes of interpretations. They offer a Cantorian objection and a Russellian objection.

I argue that self-applicability is pivotal for our theorising about validity and for the philosophical uses of validity. Intuitive reasoning about validity is often impossible unless validity is self-applicable. For example, if I uttered: 'Everything I say is logically true', I would be lying – 'Everything I say is logically true' does not seem to have the logical structure of a logical truth. Yet, this very sentence does not make sense unless validity self-applies, because it tries to talk about what is valid in the language of validity I am using.

We use validity in philosophical logic to make generalisations. Yet, such generalisations are impossible if validity does not self-apply, because they can never cover the language we are using. Intuitively a classical monist about logical validity will want to say that excluded middle is unrestrictedly valid; from their claim it should follow that 'Excluded middle is valid or not' is also valid, being an instance of excluded middle. It is part of what they are conveying that their statement covers the very language they are using, so their statement can be expressed only if validity is self-applicable. Much discussion in philosophical logic is centred on 'bridge principles': ways we can bridge from reasoning about what is logically valid to ways we should act and guide our beliefs. Yet, if validity is not self-applicable, then there is no way to bridge from

the theory of validity in which I am theorising to my context at hand. The monist is unable to use their belief that *modus ponens* is unrestrictedly valid to guide their beliefs about what is valid.

In the second half of the paper, I try to re-examine the issue at hand. I first argue that the Russellian objection to self-applicability relies on a disputable principle: that if some predicate P means that ϕ under an interpretation, then something satisfies P under that interpretation exactly when it ϕs. This assumption is usually harmless, but becomes suspicious if we apply it to statements about an untyped 'Truth under an interpretation' predicate. For example, the statement forces truth to be 'transparent': it can be iterated and discharged at will – this is quite a strong axiom in a type-free theory of truth. Another issue is that, using this principle, a 'mirror-language' where P means that not-P would be impossible; for then 'True under i' would mean not true under i, and we easily get a contradiction. My impression is that we get the contradiction too easily: there seems nothing wrong with a mirror-language *per se*, and nothing that warrants such a flat inconsistency. I argue that the principle fails to be a necessary condition for checking meaning, and I offer an alternative test that does not bear such weird consequences. Using this new principle one cannot derive a contradiction using the Russellian reasoning, so the objection is stopped.

I then discuss the Cantorian objection. I first argue that the Cantorian objection is really compelling only if one is already embracing a non-reductive account of higher order quantification. The Cantorian objection works only if we 'ascend' to a language of order higher than the one we started from, and claim that quantification in the new language is irreducible to quantification in the old language; a first orderist will simply deny that the higher orderist's reasoning is grounded. Granted the higher orderist's view on semantic ascent, however, the Cantorian reasoning is sound. Yet, the conclusion – that validity is not self-applicable – looks suspicious, since the Cantorian line of reasoning assumes that there is a viable notion of validity where validity is *not* self-applicable, to conclude that validity cannot be self-applicable because it is general

I conclude the paper by advocating the need for a re-evaluation of the issue at hand. Logical validity looks like a technical concept; a bold appeal to pre-formal intuitions when it comes to technical concepts looks unwarranted. Of course, we have some features that we'd like logical validity to have. Yet, those features also pertain to what we'd like validity to do for us and what theoretical role it has in our discussion in philosophical logic. If so, then whether we ought to prefer generality over self-applicability will need to be assessed holistically, given our theoretical needs.

To suit our theoretical needs in philosophical logic, a self-applicable theory of validity like the usual model-theoretic one might simply be a better choice.

1.4 Higher order semantics and non-standard interpretations

This article is about semantic theories of higher level languages in higher order theories. The philosophical assumption of this semantic program is that higher order theories are not reducible to set theory: higher order quantification is not reducible to first order quantification. More in general, quantification of order n is irreducible to quantification of strictly lower orders. I call someone who embraces this perspective 'higher orderist'.

I develop a higher order semantics for cumulative languages of arbitrary finite level. In standard higher order semantics, quantification of order n is interpreted homophonically, using quantification of order n in the meta-language. An interpretation of a predicate that applies at most to variables of order $n-1$ is a concept of order n (where a concept of order n is whatever we quantify over at order n). Satisfaction under an interpretation for a language of order $n \geq 2$ can be defined explicitly at order $n+1$. The semantics has interesting properties. Given a language L of level n with $n > 1$, there are exactly order-n-many interpretations of L, where a concept has cardinality order-n-many exactly when it can be bijected with the concepts of order n. In set-based semantics, on the other hand, there are exactly order-1-many interpretations of a predicate, regardless of its order. For any language L of order n, there is an interpretation under which truth satisfies the axioms for truth in Tarski 1933: this is the intended interpretation of the language; an equivalent interpretation cannot be specified in the set-based semantics for familiar reasons.

In the second part of the paper I discuss non-standard higher order interpretations of higher level languages. I focus on what I call 'raising interpretations': interpretations that raise uniformly the values of variables and predicates by some fixed order e. I provide a framework that justifies raising interpretations. In this framework each expression receives a semantic role, which is determined holistically, against the role of other expressions. The hierarchy of roles is as follows:

- formula: truth-value-role
- constants, first level variables: object-role

– second level variables: second-order-concept-role

– ....

– n-th level variables: n-th-order-concept-role.

We can conceive of raising semantics as the result of lifting the requirement that to play the role of an X *is* to be an X. This fits our idea of what a role is. I can play the role of King in the game of chess even though I am not a king. In a set-based semantics, objects (sets) play the role of concepts, even though they are not concepts in the eye of the higher orderist. In intensional logics, whether we are dealing with topic logic, temporal logic or epistemic logic, our definition of a model does not change: any set can be the set of points in the structure. Yet, if to play the role of a place, an instant or an epistemic alternative meant to *be* a place, an instant or an epistemic alternative, we would need to check whether the elements of any set can be mapped to genuine Xs: locations, instants or epistemic alternatives etc. That seems wrong.

I show that raising interpretations are models of the usual axioms of higher order logic. Also, contrary to Henkin models, they preserve the meta-theoretic properties of standard semantics: the logic is incomplete, non-compact, etc. I then discuss how quantifiers are treated in raising semantics. Under raising interpretations, every quantifier is taken to signify its characteristic logical operation across all interpretations. These logical operations are simply the extension of the usual logical operations for the quantifiers, where we take into account the larger pool of possible first order domains of raising interpretations.

I discuss how in raising semantics there is a phenomenon of indefinite raisability: irrespective of the order from which we choose to define validity, there are always more interpretations we are missing which are reachable from strictly higher orders. At no order can we say to have gathered *all* the interpretations of a given language of some level or other.

Is validity in raising semantics the same as in standard higher order semantics? I show how the two validities diverge if we added quantifiers like 'There are order-2-many xs', where an interpretation makes true 'There are order-2-many xs' when there are at least as many xs in the first order domain as second order concepts. In standard higher order semantics, 'There are not order-2-many xs that ϕ' is a logical truth. It is not in raising semantics. I discuss the result and its philosophical bearings.

In the appendix I discuss the technical details of standard and non-standard higher order interpretations.

Chapter 2

Quantification in the Interpretational Theory of Validity

According to the interpretational theory of validity, logical consequence is preservation of truth under all interpretations compatible with the intended meanings of logical expressions. A common complaint is that the theory fails to account for domain variation, which is needed to show the invalidity of statements about the cardinality of the universe. I discuss ways to phrase the arguments and the most popular solutions in the literature, which I find all wanting. I then develop a context-sensitive solution, and I argue that it solves the problem adequately. I finally discuss the consequences of my proposal and some of its ramifications.

2.1 Interpretations and logical validity

According to the interpretational theory of validity, logical consequence is preservation of truth across all acceptable interpretations, where an interpretation is acceptable only if compatible with the intended meanings of logical expressions.[1] The interpretational account is attractive for many reasons. It fits perfectly with the idea that a theory of validity is a semantic theory. A semantic theory for a language is a way to

[1] I borrow the term 'interpretational' and, later, 'representational' from Etchemendy (1990). Some think we should say 'under all interpretations of the non-logical vocabulary'. I find this way of putting things misleading, because, put like that, it might seem like we are not interpreting the logical expressions at all. Yet, what we are interpreting is a language. To interpret a language, one needs to interpret all its sentences. The meaning of a sentence is compositional, so to interpret a sentence one needs to interpret all of its constituents. Some constituents will be connectives and quantifiers, which are logical expressions. To assume that we are interpreting only the non-logical vocabulary is to ignore the definition of truth in a model, which does interpret (often homophonically) the logical expressions.

interpret it; a theory of validity is a generalisation of a semantic theory: it quantifies over all acceptable ways a language can be interpreted. Also, the interpretational account is parsimonious: all that is needed to define validity are interpretations and truth, there is no need to rely on modal notions.

The interpretational view, however, is not at all uncontroversial. A popular objection is that a 'representational' ingredient is missing in the definition of validity. According to a representational account, logical consequence should be framed using the concept of 'truth-in-a-case', where a 'case' is something yet to be specified, but that it is taken to be representational in content: a situation, a way the world might be or might have been.[2]

Truth-in-a-case is not truth under an interpretation. To say that a sentence is true 'under an interpretation' seems to me just a way to say that it is true, so-and-so interpreted. For example, if 'grass' meant snow and 'blue' white then 'Grass is blue' would be true, so-and-so interpreted. We need to specify an interpretation because only interpreted sentences can be said to be true or false, because sentences are only pieces of syntax that do not have a meaning of their own.[3] The notion of 'truth-in-a-case' is different. I can imagine a situation about dogs. Talking about that situation, we can say that every animal is a dog 'in it'; or, equivalently, that 'Every animal is a dog' is true 'in' that situation. It is not that 'Every animal is a dog' is true, interpreted under that situation; rather, so-and-so interpreted in English, it is true-in-the-situation. Finding a case is not finding an interpretation. In fact, it is central to my pretence that I keep the interpretation of sentences fixed: I am still speaking in English when I say that every animal is a dog, in the situation I am conceiving.

For the 'pure' representationalist, validity is truth in all cases, and what we call 'models' for a language are only representations of cases that serve as counterexamples to invalidities. Models do not reinterpret the language: an interpretation is prefixed, and we simply apply it to different situations. As Etchemendy puts it: 'the sentence is ours, even though the world depicted by the model is not'(1990, 21).[4] For the interpretationalist, validity is truth under all interpretations. A hybrid view is often preferred, where validity is preservation of truth-in-a-case under all interpretations (Hanson 1997, Shapiro 1998, Sher 2001).

[2]Sometimes 'case' is used very generally, as a place-holder for anything that might be relevant to validity. Beall et al. (2006) seem to use it in this way. Here, by 'case' I mean a representational case.

[3]The need for a concept of truth relative to an interpretation might also arise for technical reasons. In model theory, for example, one cannot have but truth relative to a set, for otherwise truth-in-a-structure cannot be defined in set theory.

[4]A pure representationalist seems to be Etchemendy. See 1988, 93 and 2008, 286-7.

2.2 The cardinality problem

Here is one reason why people think that the interpretational theory is wrong and we need to add a representational element. Take the following schema, where n is any positive integer[5]:

CN: There are at least n things.

Logic students know that CN can be expressed for any n using only connectives, existential quantification and identity. They also know that CN is usually not a theorem of first-order logic with identity. Are CN instances logically true? Many argue that they are, in a purely interpretational theory. They often reason as follows: according to the interpretational theory, a sentence is logically true when true under all reinterpretations compatible with the intended meaning of logical expressions. The meaning of such expressions ought to be kept fixed.[6] The intended meaning of an expression is the meaning the expression has as we are using it. '$\exists$', '$\neg$' and '$=$' are all logical expressions, and since any instance of CN is composed of these logical expressions only, it has only one acceptable interpretation: that there are n things. Thus, it is logically true if true under the intended interpretation; that is, it is logically true if true. It is true, because there are n things, so it is logically true.[7]

This is taken to be a central problem for the interpretational theory of validity, because people deny that CN instances are logical truths. There is no unique reasoning behind this thought, however, nor is there a specific methodology that people follow when making this claim. Of course, in model-theory a counterexample to C30 – say – is given by any structure with cardinality less than 30. The point, though, is how to make sense of such counterexamples: the ground for justification cannot be in the mere existence of a mathematical structure, but in what it is taken to model. If validity is truth under all acceptable interpretations and there is only one acceptable interpretation of the sentence, what is the structure modelling, and why should it matter to validity? If structures ought to be interpretations, and there is no

[5]Sometimes an exception is made for $n = 1$, since the domain is usually assumed to be non-empty. See Bencivenga 1986 for a discussion.

[6]There is a discussion in the literature about what exactly 'keeping them fixed' amounts to. I have some suggestions later in this paper. For an insightful discussion, see Sagi 2018.

[7]This argument was famously made by Etchemendy (1990) against Tarski's account of logical consequence in 1936a and 1936b. Etchemendy's argument sparked a heated discussion in the literature, with many like Sher (1996), Ray (1996) and Gómez-Torrente et al. (1996) attacking Etchemendy's exegesis of Tarski's text. Here, I am not interested in the historical discussion on Tarski's account. The argument is interesting and worth discussing for any interpretational account of validity, in its own right.

interpretation compatible with the meanings of logical expressions that the structure can model, then it should be discarded, and to consider it relevant just because it is definable in model-theory is to beg the question against the interpretational account.

2.2.1 Contingency?

Usually, arguments against instances of CN being logical truths hinge on features that people take logical truths to have and CN instances to lack. For example, a common worry is that CN instances are contingent when logical truths ought to be necessary.[8] Aristotle's famous passage in the *Prior analytics* is often quoted: 'A syllogism is speech in which, certain things being supposed, something other than the things supposed results of necessity because they are so' (24b18-20). The quote suggests that the premises strictly entail the conclusion of a valid argument and, when no premises are needed, the conclusion ought to be necessary. 'Contingent' itself has many faces, which correspond to many ways to spell out the problem. Defenders of the interpretational view claim that the notion of contingency at issue cannot be metaphysical contingency, because CN instances are provable already in Peano arithmetic, and as such, they are all as necessary as any other theorem of mathematics. Since mathematical truths are metaphysically necessary and 'There are n things' for any n is a mathematical truth, each instance of CN is metaphysically necessary. Calling for a special, stricter sense of 'logical' possibility is quite unhelpful, and might be convincing only if someone has already assumed that CN instances are not logically true, which begs the question. For, arguably, we understand something to be logically necessary when it is logically true.

Hanson concedes that CN instances are true in all possible complete situations, since possible complete situations will contain all mathematical truths. Yet, he claims that they are not true in some partial cases, in 'sub-worlds' of possible worlds. He asks us to consider 'just the Washington Monument and the White House'. If we focus on these two things alone, we have an (actual) case where it is false that there are three things (1997, 388). However, that partial worlds are relevant counterexamples is not so straightforward. Consider just London and what happens in it: is it true in London that Trump is wearing a wig? That might be, but in Washington. Yet, one might argue that it cannot be false either, for then it would be true in London that Trump is not wearing a wig. It stands to reason that partial situations call for partial evaluations: what happens in London does not settle what happens in

[8]Some believe that there are contingent logical truths, though. See Zalta 1988 for an example.

general, and in particular what relationship Trump might have with head accessories. It seems somewhat weird to point to what happens in London and claim that we have found a cogent counterexample to excluded middle. Besides, if we stipulate that only complete cases are relevant because excluded middle is valid, we are back to the problem that all (metaphysically) possible cases validate CN. One needs some non *ad hoc* criteria for the relevancy of sub-worlds to validity, yet no such criteria were put forward.

2.2.2 Asymmetry in information

Sometimes 'contingency' is interpreted in different ways. Take the following two CN instances:

C30: There are at least 30 things;

C31: There are at least 31 things.

In standard first-order logic, C30 follows from C31, but not vice-versa. There seems to be an asymmetry in information between these sentences, for what C30 says is 'contained' in what C31 says but not conversely. One can correctly conclude C30 solely in virtue of the information provided by C31, but not vice-versa. Yet, according to the argument above, both C30 and C31 are logical truths in an interpretational theory, and thus they trivially co-entail each other.

There are many ways to further specify what one means by 'information containment '. Many will be happy to say that analytic containment is what we are spotting: in virtue of what the existential quantifier, conjunction, negation and identity mean, there being 30 things is contained in there being 31 things, and to deny C30 on the hypothesis that C31 is a contradiction *in terms*. Someone who understands the meaning of logical expressions understands sentences like C30 and C31, since logical expressions are all they are made out of. And someone who understands what C31 means must conclude that C30 is true, if C31 is. Yet, the reverse does not seem to be unintelligible in the same way: even if one does conclude C31 after understanding and assuming C30, one does not do so because of what the logical expressions (and thus C30) mean.

The general, pre-given assumption behind the analytic-containment argument is that logical consequence between a premise and a conclusion implies the analytic containment between the premise and the conclusion. As a limit case, the assumption is that logical truths are analytic. Philosophers disagree on the definition of 'Analytic',

and even on whether the notion makes sense, at all.[9] Under a prominent definition of 'analytic' all of this is pointless. If we follow Frege (1884b) and define an analytic statement as something that can be converted in a logical truth by swapping synonyms with synonyms, all logical truths are trivially analytic, and analytic containment is perfectly useless as an argument against the validity of the inference from C30 to C31. For to first establish whether ⌜If C30 then C31⌝ is analytic, one would need to establish if it is a logical truth, since there is no interesting swapping to be done, as the statement is purely made out of logical terms. Yet, ⌜If C30 then C31⌝ is logically true only if the argument from C30 to C31 is valid.

If one is sceptical about the notion of analyticity, or if someone is sympathetic to the Fregean definition, arguably one should not reject altogether the idea that there is an asymmetry in information between C30 and C31, but should rather try to spell it out in different terms. One other way to spell out an asymmetry in information is to appeal to warranty in hypothetical reasoning. One seems warranted to infer C30 from C31; yet, if we start from C30, C31 does not seem to be guaranteed in the same way. In this sense, if C31 is true then C30 *must* be true on that basis alone, but not vice-versa. Incidentally, this suggests that C31 cannot be logically true, for a logical truth comes with a 'self-warranty' for its own truth, so to speak: regardless of where we start, we are always warranted to infer it. Yet, by starting from C30, we do not seem to be warranted to infer C31.

I doubt warranty should be framed in epistemic terms. I am more symphatetic to Beall and Restall who link it to the general applicability of logic: 'The applicability of logic is not a contingent matter; it works come what may, whatever hypotheses we care to entertain' (2006, 16). Whatever hypothesis we assume, we can always conclude a logical truth. Spelling out the problem in terms of warranty is useful if one is generally sceptical about the concept of analyticity, or of the notion of something being true 'in virtue' of the meaning of some expressions.

2.3 Attempted solutions

Before evaluating the problem any further, it is useful to survey the attempted solutions scattered throughout the literature. Etchemendy suggests two strategies (1990, 65-9). In the first 'something' is treated as a variable (i.e. non-logical) term. Yet, its range of admissible reinterpretations is somehow restricted to generalised existential quantifiers like 'some-dogs', 'some-humans' etc. As Etchemendy later points out, this

[9]The literature is vast and imposing. For a general discussion, see Rey 2018.

strategy looks quite *ad hoc*. If 'something' is not logical, it seems that we should be able to swap it with any other (unary) quantifier, like 'everything'. If we cannot, then we are restricting the range of legitimate interpretations. In the interpretational view, this must be because of the logical form of sentences and the meaning of logical expressions. Yet, the existential quantifier is not logical, so it cannot be the culprit and we cannot account for the restriction.

The second strategy Etchemendy suggests is a more refined version of the first one. The existential quantifier, like any other quantifier, is composed of a determiner – 'some' – and a noun-phrase – 'thing'. Maybe what is logical is the determiner alone, and not the whole quantifier. If so, then the noun-phrase can change arbitrarily, causing a shift in the range of quantification. This strategy is not directly applicable to the syntax of standard first-order logic where there are no generalised quantifiers, yet it can be implemented by exploiting the equivalence between ⌜Some ϕ is ψ⌝ and ⌜Something is ϕ and ψ⌝. Call L a language of FOL and L^+ the result of adding generalised quantification to L. A determiner in L^+ works like a binary function which takes two noun phrases and gives out a sentence (Barwise and Cooper 1981). In the syntax, we will have formulae like: $\ulcorner\exists(\phi,\psi)\urcorner$, where ϕ is the noun-phrase that restricts the range of quantification. The idea is to uniformly translate formulae of the form $\ulcorner\exists x\phi\urcorner$ of L in L^+. We then reinterpret 'thing' in L^+, to then translate the sentences back into the original L, via the equivalence of $\ulcorner\exists(\phi,\psi)\urcorner$ with $\ulcorner\exists x(\phi\wedge\psi)\urcorner$.[10]

Etchemendy claims that the proposal undergenerates: existential generalisation is invalid. If we interpret 'some-thing' as 'some-dog', the generalisation from 'Socrates is a philosopher' to 'Something is a philosopher' fails, because nothing is a dog and a philosopher. Etchemendy's point is important and valid, but it should not be taken as a knockout argument. As already said, while it is feasible to start with some pre-given results in mind about which argument is valid and which is not, these pre-given results should not be set in stone before even attempting to give a theory of logical consequence. Rather, that the theory invalidates something we pre-theoretically deem as valid ought to be assessed holistically against all other facts about the theory as a whole. We ought to stay open to changes in our pre-given ideas about the extension of validity, if we want our theory of validity to be explanatory and not just an elaborate exercise in reverse-engineering.

However, I see other issues with the proposal which I deem more serious. Usually, 'Socrates is a thing' is formalised as 'Something is identical to Socrates', which is

[10]This strategy will not work for any quantifier: 'Most dogs are four-legged' is not equivalent to 'Most things, if dogs, they are four-legged'.

why often in the literature, when we translate L into the language of generalised quantifiers, we translate 'Something is Socrates' as $\exists(x = x, x = S)$. This is feasible: to be a thing is to be self-identical, and to be self-identical is to be a thing. Yet, if we follow this natural translation, the solution does not work anymore since – *modulo* the logicality of identity – '$x = x$' has no non-trivial reinterpretations. This shows that enriching the language with generalised quantifiers is not really what makes the strategy work. We need a further assumption, as Etchemendy suggests in a note: we need to translate existential sentences via a restriction to an unspecified, non-logical predicate 'U' (1990, 166).

Having clarified how generalised quantifiers were not enough for the solution, I object that no ground has been given for translating 'Something is F' in the language of generalised quantifiers as $\exists(Ux, Fx)$ rather than $\exists(x = x, Fx)$. I am not saying that no justification is possible, but rather that generalised quantifiers by themselves do not offer any reason to prefer either translation. Yet the solution hinges on choosing the first and rejecting the second. Secondly, even if we offered such a ground, and we refused to translate 'thing' as $x = x$, we could avoid using 'thing' and, instead, use a new word explicitly for $x = x$, like 'self-identical'. Under Etchemendy's proposal 'There are two self-identicals' would still be logically true, as it should be translated as '$\exists(x = x, \exists(y = y, x \neq y))$', and we are back where we started. Even more convincingly, Etchemendy's proposal would not work if we developed a theory of validity directly in the language with generalised quantifiers L^+, for then '$\exists(x = x, \exists(y = y, x \neq y))$' would be logically true.

I also take issue with the assumption that U is non-logical: this is not something we can just stipulate. If U is to work as a suitable translation for 'thing' without sneaking in any extra-commitment, then it is likely that 'self-identical' and 'thing' must receive the same semantic value. Arguably, then, it seems to me that U is logical if self-identity is. Yet self-identity is a strong candidate for logicality, thus so should be U. Logicality is about the semantic value of a term, not about how we introduced it. I talk about tests for logicality later on, but I can give one example here. A popular test is invariance under isomorphisms (Sher 1991): a universal predicate like U that applies to anything is invariant under isomorphisms, so it is logical according to the test.

A different reaction to the cardinality problem is to deny that identity is logical. Quine (1986) famously held such a view. If identity is not logical, then any instance of CN is not a logical truth, either. Etchemendy notes that this strategy, too, comes at a cost. Seemingly logically valid arguments like the following turn out to be invalid:

Fa, $a = b$, so Fb (1990, 166). Again, Etchemendy's argument has the same limitations as before, for it relies on a predetermined extension of 'validity'. It would be better to back up the argument with some independent reasons to deny that identity is logical. Quine (1986) held the view that logical expressions link sentences together to build complex ones, in the recursive definition of well-formed formula. Under this grammatical criterion, identity is not logical. Yet many think Quine's definition overgenerates, and does not hold up to scrutiny (MacFarlane 2017, 2).

2.3.1 Logical expressions: serious vs pragmatic stance

All solutions so far tampered with the extension of 'logical expression'. It would be useful to first explain what one means by 'logical expression' first, to then re-evaluate the problem at hand. In the literature, we find a loose and a strict interpretation. The factivity of knowledege is a truth of the logic for knowledge, yet most think it is not a logical truth *simpliciter*. Some claim that the 'real' logical truths are true in virtue of the meanings of 'strictly logical expressions'. It is a non-trivial matter which expressions are logical, in this sense. For example, the invariance tradition identifies strict logicality with preservation of meaning under some kind of transformation of the universe, usually permutations or isomorphisms.[11] Others identify strict logicality with characterisability by some special, harmonious introduction and eliminations rules.[12]

A loose sense of 'logical expression' is also quite common. The set of loose logical expressions is determined by the purpose at hand: what counts as an acceptable choice of logical expressions is ultimately a pragmatic matter. Tarski at some point held such a view:

> No objective grounds are known to me which permit us to draw a sharp boundary between the two groups of terms. It seems to me possible to include among logical terms some which are usually regarded by logicians as extra-logical without running in consequences which stand in sharp contrast with ordinary usage (1983, 418).

A pragmatist is open to expansions of the usual choice of logical terms: if they are doing a logic of X, they will take X as logical.[13] A loosely logical expression

[11] The test was suggested in Tarski 1986, isomorphisms in Sher 1991, McGee 1996. For an optimistic overview of invariance tests, see Bonnay 2014.

[12] For a discussion, see MacFarlane 2017, Chapter 6.

[13] Tarski talks about extending the set of logical terms. It is unclear whether there is a lower bound, and whether there are expressions that ought to be taken as logical in all contexts, as Warmbrod 1999 suggests. Regardless, this point has no bearing to the argument I put forward.

is whatever we do not reinterpret arbitrarily. We do not reinterpret it for various reasons. It might be that we wish to enrich the logical apparatus of the theory (Warmbrod 1999), or that we are interested in the rules of reasoning that pertain to that expression. As Williamson aptly puts it: 'varying the extension of "logical constant" amounts to varying what one is investigating the general structural features of' (2017).

An interpretational theory of validity is neutral with respect to either of these two interpretations of what 'logical expression' means. I see two feasible stances one can take. One might think that there are only loosely logical expressions, and that the boundaries of what is valid varies from context to context, with the purpose at hand: I call this the 'pragmatic stance'. On the other hand, one might think that there is no conflict between the loose and the strict sense of logical expressions, but rather a change of subject and a bad choice of vocabulary. Logical expressions are the strictly logical ones, and the loosely logical expressions are not logical expressions at all, but rather something treated as such, for the purpose at hand. This explains why we are happy to treat the knowledge operator as logical in a logic for knowledge, without having to accept that the factivity of knowledge is a logical truth. I call this the 'serious stance'.

If, on the other hand, one takes the pragmatic stance, the problem with CN cannot be that its instances are not logical truth *simpliciter*, for there aren't any, but that they are not truths of the logic of connectives, existential quantifier and identity. If so, then both the Quinean and the noun-determiner strategy are pseudo-solutions, and to say that the existential quantifier or identity are not logical is to change the subject: they are not logical in a context where they are not, but here we are discussing a context where they are.

If one takes a serious stance on logical expressions, there are two ways to spell out the problem. One is to say that CN instances are not logically true *simpliciter*, the other is to say that they are not truths of the logic of connectives, existential quantifier and identity. The attempted solutions address only the first problem but are helpless against the second. The attempted solutions are also problematic in their own right, because we have independent reasons to believe that the existential quantifier and identity are in fact strictly logical expressions: most theories of logicality label them as such. For example, according to all invariance theories the existential quantifier is logical. Identity is also logical unless one consider the strictest tests, like Feferman's homomorphisms invariance (1999). Not many consider this test a viable one. By denying the logicality of these notions, both strategies set themselves apart from the

majority of the literature on the subject, which might be an intolerable burden for a meager result.

That CN instances are not truths of the connectives, quantification and identity can equally be defended with the arguments above. Even someone who takes the serious stance should agree that, regardless of whether the existential quantifier and identity are strictly logical, in a logic for such expressions we ought to treat them as such. In the same way, even if we denied that a knowledge operator is logical, in epistemic logic we surely ought to treat it as such. Hanson and Etchemendy would then claim that, since CN instances have counter-cases, they should not count as truths of the logic of connectives, identity and quantification. The asymmetry in information perceived between C30 and C31 also does not rely on (strict) logicality considerations. If we take the analyticity route, that C31 is not analytically contained in C30 is a fact independent of which expression is logical in C30 and C31. The warranty route is also untouched by logicality considerations. Thus, the attempted solutions so far seems to me either pseudo-solutions or partial solutions.

2.4 The context-sensitive account of the quantifiers

2.4.1 A methodological remark

The argument for the logical truth of CN instances assumed that there is an intended interpretation of the quantifiers, connectives and identity, which ought to be kept fixed. The intended interpretation was identified with the interpretation these expressions have as we are using them. This, in turn, justified the move from the fact that we can show from the perspective of our meta-theory that there are n things to the conclusion that 'There are n things' is true in its only acceptable interpretation, and thus logically true. The move would be illicit if the logical expressions as we are using them did not mean the same thing in all acceptable interpretations.

Yet, what is the interpretation of logical expressions as we are using them? A methodological remark at this point is useful if we are to give a clearer road-map of the discussion. Firstly, we are interpreting a language from the standpoint of the meta-theory and its linguistic resources. The meta-theory contains axioms that govern the behaviour of logical expressions. For example, here we might assume that the meta-theory contains classical first-order logic. Thus, provided that we ought to interpret the logical expressions homophonically, as we assume in the argument,

we already know something about how we ought to interpret the logical expressions: according to the axioms and rules of classical first-order logic.[14] This fact does not entail that such rules and axioms are valid, however: it only means that they hold when we talk in the meta-theory. Furthermore, other axioms of the meta-theory will offer other clues about the meaning of logical expressions. For example, if the meta-theory is Zermelo-Fraenkel, then we know that '$\forall$' ranges over an infinite domain.[15]

An additional clue about the 'intended meaning of logical expressions' comes from the intuitive relation between the logical expressions in the formal language and their counterparts in natural language. This relation is indirectly shown by the paraphrase we give to sentences in the formal language. For example, we take '$\forall$' to roughly mean what 'everything' in English means, and '$\wedge$' to roughly mean what 'and' means. I say 'roughly' because we are not making a strict identification, but more like a guiding heuristics. For example, some claim that conjunction in English semantically conveys a temporal order. We might judge 'They fell in love and they got married' true but 'They got married and they fell in love' false, because the second gets the order of events wrong. We simply dismiss this chronological aspect of the meaning of conjunction in our formal language. Yet, the identification is often pivotal to our understanding of the formal language: it would be hard (or maybe impossible?) to understand what '$\forall$' means without thinking about it roughly as 'everything'.

2.4.2 The proposal

The existential quantifier in the object-language should roughly be taken to mean 'something' in English. A very common idea is that 'something' is context-sensitive. An expression is context-sensitive when an aspect of its meaning cannot be captured through linguistic means alone, but must be supplied by some contextual, non-linguistic parameter. Indexical expressions are the paradigmatic examples: I find a

[14] This is not trivial: sometimes we do not want the axioms of the meta-theory to apply to the expressions of the object-language. For example, if we aim to give a semantics for intuitionistic logic in a classical meta-theory, we assume that negation ought not to be read homophonically, and we do not expect negation in the object-language to validate excluded middle, even though it does in the meta-theory.

[15] Problems will arise in trying to interpret quantifiers homophonically in standard model-theory, because quantification in any model is restricted to a set, yet in the meta-theory we seem to be quantifying over all sets. A similar problem arises when we try to define the 'intended model' of the language of set theory in standard model theory. These issues however are only tangential to this paper, since the discussion at hand is not about standard model-theory. We could very well spell out the interpretational account using higher-order resources, where quantifiers can be interpreted homophonically (Rayo and Uzquiano 1999, Rayo and Williamson 2003).

piece of paper on the ground, it says: 'I am Greek'. I cannot tell what the content of what is written is, unless I am told who wrote the sentence.

Is 'something' context-sensitive? It is, if 'everything' is. Even though the discussion on the semantics of quantifiers in English is admittedly quite nuanced, a great deal of literature backs up this claim. I am at a dinner party and – sadly – we ran out of wine. When I say: 'Every bottle is empty', I do not mean to say that every bottle in the universe is empty, just the ones we bought for the dinner party. If I then go out to the shop to buy more, and I say: 'Every bottle was sold out', I am now talking about the bottles in the shop, not the empty ones we just drank. Different contexts, different meanings, hence context-sensitivity.[16]

The claim I wish to defend is that '$\forall$' and '$\exists$' in first-order logic ought to be taken as context-sensitive, as well, roughly like 'everything' and 'something' in English.[17] This is *prima facie* natural, given the intuitive semantic identification of '$\forall$' with its English counterpart, and the likely context-sensitivity of the latter. Suppose we had a language with the syntactic expressions 'H' and 'Y', which we paraphrase as 'here' and 'you'. Imagine how odd it would be to paraphrase them in such a way, while refusing to treat 'Y' and 'H' as indexical expressions.

If quantifiers in English are not context-sensitive, my proposal admittedly loses much of its grip. However, it might still be interesting given its high effectiveness in dealing with the problems we have been discussing, and its adherence with common practice. Logicians simply assume that different models have different domains: this practice is something that a theory of validity should try to make sense of. Thus, *modulo* the assumption that validity is purely interpretational and the fact that models ought to be (only) interpretations, a theory that can make sense of models as

[16]For a discussion, see Stanley and Gendler Szabó 2000. There are ways to counter the idea that 'something' is context-sensitive. According to the pragmatic approach, 'something' is always absolutely general, but what one conveys at a pragmatic level might not be what is said on a semantic level. According to a grammatical solution, on the other hand, quantified sentences are often elliptical: there is an unarticulated syntactic portion of the sentence that is not uttered, which restricts the domain. Both these solutions are discussed and countered by Stanley and Gendler Szabó (2000). Finally, another theory worth mentioning is speech-act pluralism, according to which an utterance of a quantified sentence does not express any specific proposition: there are many, equally good candidates for what it is said, and in all of them the quantifier is not context-sensitive (Cappelen and Lepore 2008).

[17]For my theory to work, I do not need to be more specific. One can take '$\forall$' to either be indexical, or non-indexical but still context-sensitive. Arguably, some wording I use in the text will have to be changed, if one takes this second option, because it won't be true that 'different things' are said in different contexts; rather, the same thing is said, but it happens to be true in some context, false in some other. For a discussion, see MacFarlane 2009.

interpretations with different cardinality ought to be preferred to one that cannot, *ceteris paribus*.

The resulting theory is still interpretational, because it is impossible to interpret context-sensitive expressions without specifying a context. Truth is still interpretational: it is not that the formula is true 'in a case' – a context. Rather, only in a context does a sentence mean something which can be said to be true or not. In the example above, 'I am Greek' means different things in the mouth of different people. We can assess the sentence only if we are contextually given a subject. If the subject of the context is Pericles the sentence is true, if it is Caesar it is false. Context-sensitivity shows that an interpretation can be given in different ways: sometimes it is only linguistic, sometimes it is partly linguistically, partly contextually determined. It is not that the sentence is true in an interpretation and context; rather, the context is part of the interpretation.

Albeit context-sensitive, quantifiers can still be treated as logical expressions.[18] In this regard, my proposal is superior to Etchemendy's first and second suggestion, and to Quine's. For in all those theories we deny that some obvious candidates of logicality are logical, in direct contrast with most of the literature on the subject. Also, by not relying on a shift in the extension of 'logical expression', my theory also works against the objection that CN instances are not truths of quantification, connectives and identity, and not just against the objection that CN instances are not logical truths *simpliciter*, if there are any.

Usually, the meaning of a logical expression ought to be kept constant across all interpretations. This is correct but misleading with context-sensitive expressions, because only the *linguistic* part of their meaning ought to be kept fixed. In Kaplanian terms, we save the character, not the content: we keep constant the linguistic rule that fixes the content, not the content itself. For example, in a logic for the first person 'I', the character of 'I' is the way we fix its referent when given a context; the content of 'I', given a context, is the subject of the context (Kaplan 1989). It would be foolish to assume that, when treated as a logical expression, 'I' ought to keep its content fixed: that it ought to refer to the same person across all interpretations. Rather, we let the referent vary arbitrarily, yet 'I' is still treated as logical because we do not change its character: in every interpretation its content is always the subject of the context. In quite the same manner, what must be kept fixed across interpretations is only the

[18] We simply need to treat them as loosely logical for the second half of the problem of contingency and asymmetry to arise. Of course, '$\forall$' is one of the safest candidates for strict-logicality, if there is such a thing. The discussion on strict logicality for context-sensitive expressions is rather scarce. An exception is Woods 2017.

character, not the content of '$\forall$'. Its character is given by the recursive clause for truth of quantified formulae under a context. The content of '$\forall$' in a context is given by its character, paired with a condition of relevance provided by the context. The character is fixed: in all interpretations '$\forall$' ranges over everything relevant to the context. Yet, its content is not fixed: in a context where dogs are all that is relevant, the content of '$\forall$' is that of 'every dog', while in a context where only numbers are salient, it is that of 'every number'.

Contrary to Etchemendy's first proposal, that in no interpretation 'everything' has the content of 'some dog' is also explained: we must keep the character fixed. Contrary to his second proposal, that in an interpretation 'Something is F' has the content of 'Something is U and F' is also explained: 'U' is what is relevant in the context of that interpretation. However, we are not forced to say that 'something' is interpreted as 'some U', violating its logicality: rather, its content coincides with 'some U' under this interpretation, but the two still differ in meaning, in the same way that 'I' and the name 'Caesar' do in a context where Caesar is the subject.

We also do not need to assume that the admissible reinterpretations of 'thing' are to be decided by how we formalised 'thing'. Rather, the reinterpretation of the range of the quantifiers is done at the meta-theoretic level, via the recursive definition of truth of a quantified formula in an interpretation. In this way, since the relevance conditions are not part of the syntactic structure of the sentence, if in a context '$x = x$' is the relevance condition, and in another it is 'Fx', we are not forced to say that we have reinterpreted '$x = x$' as 'Fx', violating the logicality of $x = x$.

The fallacy in the argument for the validity of CN is diagnosed as an illicit switch in context: the fact that 'There are n things' is true in the meta-theory for any n only shows that CN instances are true when interpreted under the context of the meta-theory, not under any context.[19] The context-sensitive approach is able to frame the contingency problem in an interpretational theory. Hanson's counterexample can be accommodated by setting 'being the Washington Monument or the White House' as the relevance condition for a context. The reasoning, however, is quite different: we are not imagining a possible situation where there are only two things, and ask what would be true 'in it'. Rather, we claim that, even after a linguistic interpretation is

[19]Here I am not assuming that there is only one possible context of the meta-theory, for that will depend on one's stance on how to interpret the meta-theory itself. For example, suppose that the meta-theory is ZFC. If there is such a thing as *the* intended interpretation of ZFC, then that is the only possible context of the meta-theory. Yet, if interpretations of ZFC are indefinitely extensible, there is more than one context we can be in, as we speak in ZFC-ese. In this case, in whatever context we happen to end up in, we are assured by the axiom of infinity that there are infinitely many things, so there are going to be n things for any n, as well.

specified, we still do not know the content of 'There are at least 30 things', so we cannot say whether it is true or not. For that we need a context; when we consider what the sentence means in a context where all that is salient are the Washington Monument and the White House, we realise that, so interpreted, the sentence is simply false.

The context-sensitive approach is also able to validate the intuition that CN instances are (metaphysically) contingent, since its instances are not as necessary as any other theorem of mathematics. Rather, in the context of a mathematical theory 'There are n things' expresses something whose content is the same as a necessary truth of mathematics. Yet, in the context of me talking about what is in my closet, 'There are n things' expresses something contingent, because there could have been nothing in my closet. However, regardless of what instances of CN mean in a context, there is always a strict entailment between C31 and C30.[20]

One can also make sense of the asymmetry in information between C30 and C31: to interpret these sentences, one needs a context; regardless of the context given, if C31 is true, so-and-so interpreted, so is C30. By shifting context we change the information these sentences convey: in a context where dogs are all that is salient, C30 conveys that there are 30 dogs and C31 that there are 31 dogs; yet, in any case, the piece of information conveyed by C31 contains that conveyed by C30. If we frame asymmetry in terms of analytic containment, we would say that, in every context, C31 guarantees the truth of C30 in virtue of what the terms mean in that context (their content), even though C30 and C31 say different things in different contexts. If we frame it in terms of warranty, we would say that we are warranted to infer C30 from C31 in every context, because of what the sentences mean in that context; however, the reverse does not hold, nor are we guaranteed to infer C30 from scratch, regardless of what it means in a context, for we do not know whether 30 things are always relevant.

2.5 Ramifications and objections

2.5.1 Meta-context and generality

The context-sensitive theory of interpretations is very neutral with respect to issues about interpretations and generality. An interpretational theory states that validity is truth under all interpretations. If context-sensitive expressions are present in the

[20]In two-dimensional semantics, one would say that C31 implies C30 on the diagonal; i.e. the diagonal intension of C31 is a subset of the diagonal intension of C30.

language, then a theory of interpretations must include a theory of contexts. The theory is otherwise silent on how a theory of interpretations ought to be spelled out. The standard way to define an interpretation is in model theory, where an interpretation is a specific set. However, some think that model-theory is inadequate, for they claim that many interpretations that are not set-sized cannot be accounted for. More strikingly, the 'intended' interpretation of the language we are using in the meta-theory is missing, for it is not set-sized: after all, there is no set of all sets. To fix these issues, a plural or higher-order theory of interpretations has been proposed (Rayo and Williamson 2003), as well as an axiomatic/substitutional approach (Halbach 2018). Again, the interpretational theory of validity is compatible with all such options.

An interpretational theory is also neutral with respect to how general contexts can be. Proponents of absolute generality claim that quantifiers can range over absolutely everything, without any restriction (Williamson 2003). Sceptics claim that quantification is meaningful only given a domain of discourse. This latter view is quite attuned with a contextual view of quantification: to even spell out an interpretation of the quantifiers, a domain must be provided in some way: a context can serve this purpose. However, in no way the context-sensitive theory forces us to follow the sceptics. Absolute generality is compatible with the kind of contextualism implicit in a context-sensitive account of quantification: it only implies that the most general context exists, not that the most general context is the only context.

Arguably, if there is a most general context, then this ought to be the context of the meta-theory. For a general theory of contexts can be spelled out only from the perspective of an absolutely general context. If not, it is only a theory of the contexts specifiable from the context one is in, and this is not what the theory intends to be (Williamson 2003). Due to set-theoretic paradoxes, absolute generality cannot be framed in standard model-theory. For the domain of a model is always a set, and due to Cantorian reasons, a set cannot contain all of its subsets, because it has more subsets than elements. Again, plural, higher-order or axiomatic approaches can be employed to model absolute quantification and the most general context.[21]

Someone who does not believe in absolute generality will not be swayed by the purported need for a most general context. If there is no such thing, then what one can conclude is only that the meta-theoretic context ought to be general enough for the theory of validity to be adequate. What 'enough' amounts to will depend on the purpose at hand. A reasonable, minimal requirement is the adequacy of a suitably chosen deductive calculus with the semantic system. In order to show completeness,

[21]For a comprehensive discussion of absolute generality, see Rayo and Uzquiano 2006.

often set theory is employed; yet, arithmetical counterexamples have been proven to be enough for completeness to go through (Hilbert and Bernays 1939). So, it is fair to say that the meta-context ought to be at least as general as arithmetic for the theory to be minimally adequate: for it to be able to prove a completeness theorem. Other than that, the theory does not need to say anything.

Suppose that a friend of absolute generality adds to the language a quantifier which is explicitly absolutely general. Arguably, the resulting expression will not be context-sensitive, for in any context it 'points back' at the most general. Suppose we substitute such quantifier to 'there are' in C30, obtaining the resulting sentence:

A30 There (absolutely) are at least 30 things.

$A30$ is a truth of the logic of the absolutely general quantifier iff it is true under any linguistic interpretation and context, iff it is true under all linguistic interpretations paired with the general context, iff it is true under the most general context. It is true, so it is a truth of such a logic. If there are strict logical expressions and the absolutely general quantifier is one of them, then $A30$ is a logical truth *simpliciter*.

It is not obvious that A30 is subject to the same issues C30 had. It does not mean that there are 30 things, it means that there absolutely are 30 things. Suppose we had an 'arithmetical quantifier' which means 'There are (in arithmetic)', and suppose we treated it as a logical expression. It seems to me that 'There are (in arithmetic) 30 things' should count as logically true.

A representationalist will complain, for A30 is still false 'in' a story where my pen and my notebook are all there is. In this case, an interpretationalist should bite the bullet. When said in a context where the socks in my closet are all that is relevant, 'There are at least 30 things' says something contingent: there could have been nothing in the closet. On the other hand, A30 is as contingent as mathematics, and thus cannot have been false in any possible complete situation. Incomplete situations falsify excluded-middle, and there is no good reason to take them as relevant counterexamples to invalidities. The asymmetry problem also does not apply, for we do not need a context to pin down what A30 means: it always means that there absolutely are 30 things. I can always prove that there are 30 things in the most general context, so I am always warranted to infer the truth of A30, under any interpretation. Regardless of the context given, I am warranted to infer 'from scratch' that there are n things in the general context, for any n; thus, I am also warranted to infer any variation of $C1$ with absolute quantification from any other, as well.[22]

[22]For a similar discussion about an (explicitly) absolutely general quantifier, see Williamson 1999.

2.5.2 The Socrates problem

A context-sensitive account of quantification seems incompatible with universal instantiation. Adapting an example from Gauker 1997, suppose that in a context c all that is relevant are some wooden figurines on a table. Under c, 'Everything is made of wood' is true even though 'Socrates is made of wood' is false. The usual solution in model-theory is to reinterpret 'Socrates' and all other expressions inside the range of the quantifiers. As Etchemendy suggests, this holistic fitting procedure makes sense in representational terms (1990, 66). For if the domain of quantification is something like a picture that describes what exists in a story, then all the expressions of the language ought to be reinterpreted so that they 'squeeze' into the picture. Yet, this move is hard to justify in our interpretational semantics. One will need to achieve the same result by systematically banning all interpretations where, say, a name refers to something outside of the range of the quantifiers, or where a predicate is given an extension that outstrips it. As Etchemendy points out, these 'cross-term restrictions' are hard to make sense of, since the interpretations banned are perfectly legitimate, if not for making some rules or axioms invalid, which puts the cart before the horse.

In a context-sensitive, interpretational theory, no cross-term restrictions should be adopted. Take English: under a context where everything relevant is in on the table, we are not reinterpreting the whole language so that everything fits on the table: we are still considering English, where 'Socrates' still means Socrates, even though he is not here. Of course, maybe we cannot actually make utterances about him in that context because, just by evoking him, we would make him relevant, shifting context.[23] Yet, at a semantic level, we can still ask what do sentences about Socrates mean, and whether they are true or not. As far as I can see, there are two main routes one can take. The first claims that Socrates is human regardless of what is relevant to a context. The second states that the truth of a formula about Socrates might change, if Socrates is not relevant to the context. My preference is for the first option, but I am open to both.

2.5.2.1 Free logic route

Suppose we take the first route. How can a sentence in a context fail to be true or false, in the sense of truth and falsity we are focusing on? A confusion might easily arise when we ask whether 'Socrates is human' is true, under an interpretation that

[23] I am oversimplifying things here to make a point: things are not that simple. For example, when I say 'I have everything in my backpack' I do not mean to imply that I am inside the backpack, nor that my backpack is inside itself (I thank an anonymous referee for bringing this up).

specifies a context. For we are not asking whether 'Socrates is human' is true 'in' the context. We are not presenting a story about what is relevant to the context, and asking whether 'Socrates is human' is true 'in that story'. Rather, we are asking what 'Socrates is human' means, when presented with an interpretation that specifies such a context, and whether what it says is true, so-and-so interpreted. Following this line of thought, the proponent of the first route might make the following argument: assuming compositionality, the meaning of a sentence is determined by the meaning of its constituents. The meaning of 'Socrates is human' is determined by the meaning of 'Socrates' and 'human'. Since 'Socrates' does not shift reference across contexts – it always refers to Socrates – the change in truth-value must be attributed to 'human', so 'human' must be context-sensitive. In fact, we are forced to assume that every predicate is context-sensitive in this way: even 'Socrates is Socrates' is not true when Socrates is not relevant, so identity must be context-sensitive in some sense, as well. The proponent of the first route rejects this conclusion: the most sensible solution is to say that 'Socrates is human' is not context-sensitive, because none of its components are.[24] I find a piece of paper on the ground, it says: 'Socrates is human'. *Modulo* the assumption that the author intended to write in English, I do not need to know the context in which the piece of paper was written to understand what it means. Under any context, it means that Socrates is human, which is something true regardless. The resulting theory of validity fits a free logic, where existential generalisation and universal instantiation both fail: since Socrates is human, 'Socrates is human' is true, regardless of the context at hand, even though 'Something is human' might be false, given what it means under a context. Existential generalisation can be substituted with its free-logic friendly version: $\ulcorner(\phi(x) \wedge \exists y\, y = x) \rightarrow \exists z \phi[z/x]\urcorner$. That we need to abandon universal instantiation and existential generalisation is a *malus*, given their *prima facie* validity. Yet, the proponents of the first route might choose to bite the bullet. They might claim that this is holistically the best solution, given that we wish to save the interpretational account, that context-sensitivity is a very natural way to do that, and that such rules fail to preserve truth in all interpretations, when we take contexts to be an interpretational component.

[24] We are disregarding the tense in the verb, because we are dealing with an extensional logic. What happens to the interpretational account in modal and temporal languages is beyond the scope of this paper.

2.5.2.2 Kleene and the classical logic route

Suppose we take the second route instead: when Socrates is not relevant to a context, formulae about him can change truth-value. The semantic theory in the background postulates a pervasive context-sensitivity of truth, which is not triggered only by the presence of quantification. What kind of notion of validity we extrapolate from this semantic theory will depend on how we treat the truth-value of formulae about irrelevant things, under an interpretation that specifies a context. Say that Socrates is not relevant to the context: one option is to evaluate all formulae about him uniformly false. Yet, it seems to me that if 'Socrates is human' was false under said context, then 'Socrates is not human' ought to be true but it is not, because again the formula is about Socrates. A more sophisticated option is to say that any formula about Socrates is neither true nor false, if Socrates is not relevant. A multi-valued logic must be employed. If validity is truth-preservation under all interpretations, then we obtain something like Kleene validity. This is the route taken by Gauker (1997), under the name 'Kaplanesque three-valued free logic'. In such a logic no tautology will be valid, since any object is irrelevant to some context, so any formula about that object will be neither true nor false, when interpreted under that context.[25] As explained by Gauker, the rule of universal instantiation is invalid, because the fact that everything relevant to a context satisfies a condition does not imply that Socrates does, as Socrates might not be relevant. On the other hand, if Socrates does satisfy some condition under some context, then he must be relevant, so something relevant satisfies that condition: the rule of existential generalisation is valid (1997, 196). This means that the rule of contraposition fails, *Modulo* the equivalence between $\ulcorner\exists x\phi\urcorner$ and $\ulcorner\neg\forall x\neg\phi\urcorner$. Since no tautology is logically true, the deduction theorem also fails: $\ulcorner\phi\urcorner$ implies $\ulcorner\phi\urcorner$, yet $\ulcorner\phi \rightarrow \phi\urcorner$ is not logically true. Similarly, proofs by contradictions are invalid: $\ulcorner\phi \wedge \neg\phi\urcorner$ implies any formula and its negation, since under no interpretation it is true; yet, $\ulcorner\neg(\phi \wedge \neg\phi)\urcorner$ is not logically true.

What should we make of such a massive departure from classical logical reasoning? Gauker (1997) takes truth and falsity in a context to be pragmatic notions: the truth and falsity of a sentence in a context are its assertability and deniability, respectively, in the context. Interpreted as such, the departure is reasonable: if Socrates is not

[25]This can be proven by induction on the complexity of formulae in a standard three-valued semantics of the like defined by Gauker (1997). Some atomic formulae might not be about any object in particular, like 'it's cold'. These sentences can be formalised as sentence letters. Since they are not about anything, some might think that they are true in any context if true, at all. This fact does not change the results of the theory, because 'It rains' can arguably be reinterpreted as any other atomic formula.

something relevant to the context at hand, then anything that can be said about him is unutterable, for if uttered it would cause the context to shift. It is not deniable either, for to deny something we would have to utter its negation. Gauker is not interested in what I call 'interpretational theory of validity', and his system has independent motivations, and cannot be applied here if not by changing the subject. For now we are not talking anymore about the notion of truth we have been interested in: that which applies to a sentence exactly when what it says is the case. Now, we are talking about speech acts, and norms of communication in a context. Contexts in the interpretational theory of validity were introduced to serve one purpose only: they are devices to interpret formulae, by providing a range for the quantifiers and maybe to limit the extension of other context-sensitive predicates. They are part of a pragmatic theory as much as variable assignments are. At no point did we mention by what sort of norms a speaker should abide, nor how should we use sentences in contexts. The job of a semantic theory ends with a definition of truth of a sentence, given an interpretation that explains what the sentence means. A context-sensitive sentence does not mean something that can be said to be true or false unless we specify a context: that is all contexts are there to do.

In the scope of the interpretational account we are interested in, we need to deny that truth and falsity in a context are pragmatic notions, and defend a purely semantic use of contexts. If we stick to the second route, we are also insisting that it is not only quantifiers that affect the truth of a sentence in a context: truth is highly context-sensitive, and even atomic formulae might change truth-value based on what is relevant to the context at hand. Bivalence fails, but failure of bivalence does not necessarily lead to Kleene logic, or in general to non-classical logic (Rumfitt 2015).

Here is one way to recapture classical logic in the scope of a context-sensitive, interpretational theory of validity: we can distinguish between a strong and a weak counterexample to the validity of a formula. A strong counterexample is given by an interpretation under which the formula is false, a weak counterexample is given by an interpretation where the formula is just not true. If we accept all counterexamples with no filters, we obtain Kleene validity. Yet, under a more stringent filter, only strong counterexamples matter. There might be a sense in which 'Trump is wearing a wig' and 'Trump is not wearing a wig' are neither true nor false under a context where all that is relevant is what is happening in London, because Trump is not in London. So, 'Trump is wearing a wig or not' is not true under this interpretation, strictly speaking, but this failure seems shallow. Again, there is something odd in pointing at what happens in London and claiming that we have found a cogent counterexample

to excluded middle. Under this stronger filter for counterexamples, we obtain the following notion of validity: $\ulcorner\phi\urcorner$ logically follows from Δ exactly when, under all acceptable reinterpretations, if each member of Δ is true, $\ulcorner\phi\urcorner$ is not false. It is easily checked that the corresponding notion of validity is classical, because every valuation where the conclusion is not given a classical truth-value is discarded, so only classical valuations matter. For example, universal instantiation and existential generalisation are valid, because at most they are not true when we instantiate with or generalise from irrelevant things.[26] I do not wish to defend this second notion of validity against the first. I only wish to show how versatile and powerful the context-sensitive theory of validity is. We have different solutions to the Socrates problem, and one of them recaptures classical logic: no such routes are available in the strategies present in the literature, which I discussed in section 3.

2.5.3 Pragmatic vs semantic uses of contexts

A possible complaint against the context-sensitive, interpretational account of validity is that, contrary to what we mantained, by adding a context in the interpretational theory validity ends up being a pragmatic notion. One example is Hanson who, talking about Kaplan's work, takes indexes of contexts (speaker, time etc) to be 'pragmatic elements', and Kaplan's definition of logical consequence as truth in all models and contexts to be 'a kind of pragmatic universality'. For him, Kaplan's logical truths are sentences that 'cannot be uttered falsely' (2006, 438). As we have seen, Gauker (1997) fully embraces this conclusion: logical validity should be defined, roughly, as 'assertability in a context', which is a pragmatic notion. Yet, again, Gauker is not interested in the kind of project we are interested in.

Contexts can surely play a prominent pragmatic role. For example, they are paramount to explaining pragmatic implicatures. Their role, when pragmatic, is to fill the gap between what we say and what we communicate, given the norms of use of sentences in contexts. However, in an interpretational theory, contexts do not explain what is not said but implicitly communicated, but simply what is said, which is the job of a semantic theory. The job of semantics is to interpret a language and define truth; with context-sensitive languages, one simply cannot interpret them nor define truth without the notion of a context.

[26]Yet another notion of validity we can obtain is equivalent to Priest's logic of paradox (1979). We just need to read the third truth-value as true-false, and redefine validity so that the designated truth-values are true and true-false.

Theories of quantification in natural language offer a helpful, explanatory example of pragmatic versus semantic use of contexts. According to the pragmatic view discussed by Stanley & Szabó, 'Everything is packed' always means that everything whatsoever in the universe is packed: contexts do not play any semantic role in this view, and quantification is always unrestricted. Yet, what I communicate at a pragmatic level is not what I mean at a semantic level. Precisely because I blatantly lied when I said that everything is packed, I have violated the Gricean maxim of quality. Thus, helping themselves with the context at hand, 'the interpreter looks for some proposition that the speaker may reasonably be taken to have meant'; for example, that everything that I presently need is packed. According to the semantic view, on the contrary, contexts act at a semantic level in explaining what the sentence says to begin with: what proposition it expresses. They are part of the semantic interpretation, and are needed to spell out the truth-conditions of quantified sentences. According to this view, 'Everything is packed' might be true, even though not everything in the universe is packed; for, in a normal context, the sentence *means* that everything that I need is packed.

It is also incorrect to label logical validity in context-sensitive languages as impossibility to be uttered falsely under any interpretation, as Hanson himself half acknowledges, and as Kaplan himself makes very clear (1989, 522). Distinguishing between what can be said in context and what is true under an interpretation which specifies a context is pivotal to understanding how indexical expressions work. Many sentences that cannot be uttered falsely are still false, when interpreted in a context. Kaplan makes some convincing examples:

> 'By the time an agent finished uttering a very, very long true premise and began uttering the conclusion, the premise may have gone false. Thus even the most trivial of inferences, P therefore P, may appear invalid. Also, there are sentences which express a truth in certain contexts, but not if uttered. For example, "I say nothing" ' (1989, 584).[27]

Contexts, in a semantic theory, are non-linguistic ways to specify meaning. They are not the only examples. An open formula like 'x is human' cannot be interpreted on purely linguistic grounds. All that language can do is to specify the meaning of the predicate 'human'; yet x, linguistically, does not mean anything: it is only a place-holder. A variable assignment is needed to provide a referent and complete

[27]I am not arguing that Kaplan had in mind exactly what I have in mind when I talk about truth under a context, for this is debatable. I am only saying that his and mine interpretations are not Hanson's.

the content of 'x is human': the meaning of an open formula must be supplied by some non-linguistic parameter. In this sense, contexts are not introducing something completely new, and it is normal to assume that there is continuity between a theory of validity applied to a context-insensitive language and a context-sensitive language.

2.6 Conclusion

It is common to assume that the purely interpretational theory of validity overgenerates because statements about the cardinality of the universe incorrectly turn out to be logically true. I spelled out different ways of framing the problem. By distinguishing between strict and loose logical expressions, I argued that the solutions in the literature either miss the mark or solve only one way of framing the issue. I then discussed the underlying assumptions of the cardinality objection. I claim that the objection fails if we take quantifiers to be context-sensitive. I showed how versatile the context-sensitive proposal is, and how it can make space for many notions of validity, one of which is classical. I then defended my theory against the widespread distrust of context-sensitive expressions. A context-sensitive solution is natural given the likely context-sensitivity of quantifiers in natural language, and allows us to solve the cardinality problem in a purely interpretational theory of validity, without the need to rely on additional, modal notions.

Chapter 3

Substitutional Validity for Modal Logic

In a substitutional framework, validity is preservation of truth under all substitutions of the non-logical vocabulary. I develop a semantic theory where $\Box$ is interpreted as a way to reflect substitutional validity, or extensions of it: $\Box$ is preservation of truth under all (or some) substitutions. I show how to prove soundness and completeness for common modal calculi using this definition.

3.1 Philosophical context of this work

3.1.1 Simple truth vs truth in a model

According to the substitutional account of logical consequence, an argument is logically valid if it preserves truth under all uniform substitutions of the non-logical vocabulary. The substitutional account has a longstanding pedigree.[1] Quine is probably the main modern proponent and the target of most of the contemporary discussion, especially in his 'Philosophy of Logic' . Quine did not have in mind any specific formal theory for his account, and the discussion on his proposal on logical validity is complicated and entangled with sophisticated historical points. Many have attacked him (Hinman et al. 1968, Boolos 1975, Etchemendy 1990, Hanson 1997), and some have tried to defend him, or at least to soothe the critics (Berlinski and Gallin 1969, McKeon 2004, Eder 2016).

It is common to view the substitutional account as an alternative to the model theoretic account, and yet when it comes to making the former precise, people have

[1]For a discussion, see Halbach 2018, note 2. The substitutional account arguably traces back at least to Buridan.

often used model theory. Arguably, Berlinski and Gallin (1969) developed one of the best versions of the substitutional account, which however is framed in model theory. First, they argue that Quine's definition of logical truth is incomplete 'since "truth" must be relative to an *interpreted* language'. Then, they consider one specific model – they call it a 'Quine-model' – and define a sentence to be logically true when true *in that model*, under all substitutions. Whilst their account answers many questions others have raised, it neglects the fact that a substitutional account does not need the scaffolding of model theory to work; it is really a genuine alternative on its own. In their system, on the other hand, the substitutional account is not very different from the model theoretic: all the semantic machinery is done exactly in the same way, what changes is only the definition of validity.

What is the point of developing an alternative to model theory to frame substitutional validity? In model theory we define the concept of truth in a model. However, the substitutional definition does not need the concept of truth in a model to work, but only a concept of simple truth: validity is preservation of truth – simple truth – under all substitutions of the non-logical vocabulary. It follows that all models are superfluous, except maybe for the intended model where truth in a model collapses into truth *simpliciter*. However, such an intended model is often nowhere to be found. For example, there is no model of ZFC where truth in it is equivalent to truth *simpliciter* (i.e. truth that satisfies the Tarskian axioms in (Tarski 1933)). All models have a set-sized domain, and a set-sized extension of membership.[2] Yet, the intended extension of membership is not set-sized, nor is it the intended domain of quantification of the meta-theory, where we quantify over all models and thus over all sets. Consequently, for such languages truth in a model is never simple truth. We require truth for a language to satisfy a suitable T-schema, maybe restricted to truth-free formulae. Yet, no model satisfies a T-schema even for atomic formulae of the language of ZFC; there is no model where, for any x and y, membership holds between them in the model exactly when x is in y. We also expect truth to give a homophonic reading of some expressions, like the quantifiers, yet there is no model where ⌜Everything is ϕ⌝ is true in the model iff *everything* satisfies ϕ in the model, because quantification is always restricted to a set.

3.1.2 Advantages of the axiomatic approach

If we wish to apply the substitutional account to the language of set-theory, or to any language that shares a similar level of generality and strength, the concept of truth

[2]By 'set-sized' I mean the size of something that can be collected in a set.

in a model looks superfluous, and even unsuitable to capture the intuitive meaning of validity in substitutional terms. If all we need for a substitutional account are substitutions and simple truth, it seems reasonable to just build a theory of truth for the object-language directly, rather than hoping that the model theoretic machinery will provide us with a suitable substitute. Recently, Halbach followed this idea (2018, 2020). He builds a truth-theory for ZFC using a primitive notion of simple satisfaction; validity is preservation of satisfaction under all substitutions and assignments of values to the variables: no concept of truth in a model is needed.

The axiomatic approach has many advantages over the model theoretic account. Firstly, there is a substitution – the identity function – that provides a homophonic translation of all formulae, and under which truth satisfies the compositional axioms for connectives and quantifiers, and a T-schema for all formulae of set-theory. There exists no model that matches the homophonic substitution, however, nor is there a model equivalent to any substitution which maps membership to some formula that fails to define a set, like '$x \notin x$'. In contrast, for any model there exists a substitution where truth in that model is equivalent to (simple) truth under that substitution (Halbach 2020).[3]

A second advantage is the ability to model absolute generality. Many argued that quantifiers can often be absolutely general, especially when we do logic and we state logical truths; when we say that everything is self-identical, we really mean that everything whatsoever is self-identical, 'without restriction'. (Williamson 2003).[4] Many have struggled to make sense of the restriction to a set in the interpretation of quantifiers, which is essential in model theory to ensure that truth in a model is definable in set-theory. In the axiomatic approach, quantifiers are naturally taken to be absolutely general given the unrestricted Tarskian clause for the satisfaction of quantified formulae; 'Everything is ϕ' is satisfied exactly when everything satisfies ϕ, without restrictions. The notion of a domain has no place in the substitutional account.

Many have tried the higher order route to provide interpretations where there are no restrictions to a set-sized domain, often appealing to plural resources (Boolos 1985, Rayo and Uzquiano 1999, Rayo and Williamson 2003, Rayo 2006). For the strategy to be successful one needs to bear a lot of ideological baggage: Quinean-style skepticism

[3]Quine thinks this cannot be the case due to a variation of what he calls 'Grelling's paradox'. He considers the set of all sentences of the object-language that do not satisfy themselves. In a type-free system this set presents no problem: the satisfaction predicate is in the object-language, the paradox is stopped because a T-schema is not provable for sentences like 'x does not satisfy x'.

[4]For a contemporary and in-depth discussion on absolute generality, see Rayo and Uzquiano 2006.

about higher order languages needs to be set aside; higher order quantification needs to be *sui generis*, irreducible to quantification over set-like, singular objects. Under the axiomatic approach, there is no need to appeal to higher order resources to model absolute generality: everything can be done in a first order language; the higher order, ideological baggage is not needed. Also, in the axiomatic approach we are able to provide a universal theory of validity that applies to possibly any first order language, even the meta-language. This is not possible using higher order resources, where the object-language needs to be of strictly lower order than the meta-language.

Lastly, in the substitutional, axiomatic approach one can show that logical truths are true and that validity preserves truth simply by looking at the definitions. For example, since logical truths are true under all substitutions, they are true under the homophonic: i.e. they are true. The situation is quite different in model theory, where there is no intended model, so no model where truth in it really means simple truth. Hence, the relation between validity and truth is far from obvious: one will need to appeal to reflection principles to achieve similar results.[5]

3.1.3 Intensional languages

All the substitutional accounts out there have been focusing on first order logic. As far as I know, no one has tried to extend the account to intensional languages. Quine surely was not interested in this endeavour, for he was skeptical of modal logic to begin with. The substitutional tradition thus far has followed him in his disregard for intensional languages. Yet, much of the success of model theory resides in its adaptability to any language, and in particular in its ability to interpret intensional languages through the possible world machinery. If the substitutional account cannot work for intensional languages at all, it would be less interesting compared to the model theoretic.

I propose a substitutional account of validity for modal languages where modality is interpreted as a way of talking about what is (substitutionally) valid. If what we are interested in is not this kind of modality, but something else – maybe we are working on the logic of knowledge, or on the logic of duty and moral permissibility – then the substitutional account is not really an alternative to the Kripke-style semantics, where we have more flexibility on what we take the possible worlds of the theory to be:

[5]Again, I point to Halbach 2018 and Halbach 2020 for a discussion. Reflection principles roughly show that any formula of set-theory is satisfiable if what it says holds in the universe of sets. So if it is true in all models, it holds (i.e. it is true). They have been proved for ZFC by Levy et al. (1960) and Montague (1961). However, reflection principles are not provable in higher order set-theory (Shapiro 1987).

epistemic possibilities, moral possibilities etc. Yet, when they are compatible, the two accounts are genuine alternatives, and I would argue that the substitutional account has many advantages over the usual Kripkean account, spelled out in set-theory.

Similar reasons favouring a substitutional approach carry over to intensional languages. The lack of an intended interpretation puts into jeopardy the relation between validity and truth: we cannot show that logical truths are true, nor that validity is truth-preserving just by looking at the definitions. Kripke semantics is spelled out in ZFC where there are no modal operators. It follows that reflection principles are of no help. It cannot be proven that $\ulcorner\Box\phi\urcorner$ holds only if it is satisfiable, and that therefore if $\ulcorner\Box\phi\urcorner$ is true in all models, it is true. So, as far as we know, it might be that $\ulcorner\Box\phi\urcorner$ is true in all models but false, or false in all models but true.

Suppose that $\Box$ means 'It is valid that'. This is roughly the meaning it will take in the substitutional system. If we try to model 'It is valid that' in Kripke semantics, however, we run into problems. According to Kripke semantics, validity is preservation of truth under all Kripke models. So, points in a Kripke model now represent Kripke models. A first problem is that there are at least set-many models of a given language, but only set-sized many points in any Kripke model.[6] So, in no model are we really getting all the interpretations, and it is not clear whether there is a model that is intended for $\Box$. Sometimes we can get around the issue: if the logic of 'It is valid that' is canonical for Kripke semantics, for example, then there is a canonical Kripke model where something is true in it exactly when it is valid. Arguably 'It is valid that' in interpretational terms is S5, which is canonical. So, $\Box$ interpreted in its canonical model will likely receive its adequate extension. However, if we decided to change the reading of $\Box$ as truth in some given subset of the Kripke models, it is not clear at all that any Kripke model will provide the correct reading of $\Box$: we need to rely on the meta-logical properties of the semantic system and the logic generated by that particular subset of Kripke models, which might not be there.

Other problems lurk around. The link that should be there between $\Box$ and truth is missing. It should be clear from the definition of $\Box$ as logical validity that $\ulcorner\Box\phi\urcorner$ implies the truth of ϕ, since validity ought to be truth preserving. Of course, we can designate a point in the Kripke model that plays the role of the intended interpretation, but, for the same reasons discussed above, truth at the designated point cannot be truth. Incidentally, this shows how Kripke semantics is also unable to model $\Box$ as 'It is substitutionally valid that': in no Kripke model are we able to mimic the homophonic interpretation.

[6] By 'set-many' I mean 'as many as the sets', i.e. the cardinality of the universe of sets.

The membership relation under the intended interpretation has the extension it actually has, and the quantifiers range over what they actually range; since all points of any Kripke model receive a set-sized domain, the actual meanings of the quantifiers and membership cannot be matched by any of them. Thus they are impossible according to the system, because they are not represented by any 'possible world' of any Kripke model. So, even though the schema $\ulcorner\Box\phi \rightarrow \phi\urcorner$ might be valid, it cannot be shown that the truth of $\ulcorner\Box\phi\urcorner$ implies the truth of ϕ, even though it should be obvious from the definitions. As before, reflection principles are of no help, here, since ZFC is not a modal language: we cannot obtain a reflection principle for a formula that does not belong to ZFC. Other obvious links between validity, $\Box$ and truth are missing. It should be obvious from the definition of validity that if ϕ is logically true then its necessitation is true, and that if ψ follows from ϕ then the strict implication between ϕ and ψ is true. None of this can be shown in standard Kripke semantics.

3.1.4 A proposal

Our approach will be based on the following, reductive analysis of modality: the truth of $\ulcorner\Box\phi\urcorner$ is reduced to (possibly weaker versions of) logical truth. The broadest notion of necessity is logical validity: $\ulcorner\Box\phi\urcorner$ is true exactly when ϕ is logically true. To be logically true, according to the substitutional account, is to be true under all substitutions of the non-logical vocabulary. Thus, $\ulcorner\Box\phi\urcorner$ is true exactly when ϕ is true under all substitutions of the non-logical vocabulary.[7] In virtue of this definition, we can read $\Box$ roughly as 'It is valid that'.

Valid: Where $\Box$ means 'It is valid that', $\ulcorner\Box\phi\urcorner$ is true iff ϕ is true under all substitutions.

In the resulting system, the links highlighted above between truth, validity and necessity are met: if ϕ is logically true then it is true under all substitutions. In particular, under the identity substitution, so ϕ is true. Also, because the truth of ϕ is preserved under all substitutions, $\ulcorner\Box\phi\urcorner$ is true as well. If ψ follows from ϕ then the implication from ϕ to ψ is satisfied, because it is true under the identity substitution. Also, if ψ follows from ϕ then under all substitutions if ϕ is true ψ is true, which means that the strict implication from ϕ to ψ is true. Finally, if $\ulcorner\Box\phi\urcorner$ is true then it is true under all substitutions, in particular under the identity substitution, so ϕ is true.

[7] If ϕ is a formula and not a sentence, it is simply true iff satisfied under all assignments, and thus $\ulcorner\Box\phi\urcorner$ is true iff true under all assignments and substitutions.

My account is reductive: the language of necessity can be interpreted in a non-modal language. Contrary to Kripke semantics, I do not wish to add a concept of truth in a world, for the same reasons I do not want a concept of truth in a model: substitutional validity is preservation of simple truth under all substitutions. It is neither preservation of truth in a model nor preservation of truth in a world: the possible world machinery is a superfluous structure, from a substitutional point of view. All that is needed to define validity for a modal language is a primitive concept of truth and a definition of substitution. Those are the same 'tools' we use to define validity in first order logic, no additional tools are needed.

It would be wrong and unintended to expect that this definition of $\Box$ ought to work in general, given any suitable use of modal logic. $\Box$ in our system does not talk about truth in different possible worlds, or different instants, or across different agents in different epistemic alternatives. Our semantics is not neutral: it presupposes a specific understanding of $\Box$ as a way of expressing (substitutional) logical validity, or suitable expansions of it. If this understanding is unfeasible, the semantics is unfeasible. The aim of this paper is to show that there is an interesting way of framing a substitutional understanding of validity for intensional languages. It is not our goal to extend the theory to any reading of $\Box$. Since we interpret $\Box$ as a way of expressing substitutional validity – that very same notion of validity we are discussing in first order, extensional logic – it is not too unreasonable to take $\Box$ as a logical constant. Validity for this modal logic is therefore a natural extension of validity for first order logic: it is preservation of truth under all substitutions of the non-logical vocabulary.[8]

Things are different for Kripke semantics, where $\Box$ can be interpreted pretty much as anything: the nature of the points in a structure is not specified. Yet, I feel like the extreme adaptability of Kripke semantics to different logics often obscures the differences between talk of validity in different logics. It is not obvious at all that the same notion of validity is at play across first order logic, epistemic logic, logic of action, temporal logic, and so on. Whilst in first order logic validity is usually preservation of truth under all interpretations; in epistemic logic to define validity we quantify over epistemic possibilities, or other 'representational' objects; to define validity in a logic of actions we quantify over agents; to define validity in a temporal logic we quantify over instants. So, even though we are using the same name, it looks to me like it is

[8]Of course the natural theory is where $\Box$ is just logical validity. In theories that interpret $\Box$ as some extravagant expansion of validity we get a far less natural reading. We study them to see how far the system can get us to.

not of the same 'validity' that we are talking about in these different settings.[9] If our theorising about validity did not realise the semantic shift and expected to apply the same concept of validity to first order logic and epistemic logic, it would be problem, not a virtue. Thus, we should not expect the substitutional account to work in an epistemic logic: a substitutional understanding of validity applied to epistemic logic does not make much sense. The substitutional account works only when it makes sense to apply it.

Different truth theories can specify different interpretations of the modal operators, by restricting the range of substitutions relevant to the truth of $\ulcorner\Box\phi\urcorner$, with the limit cases being the trivial and universal restriction. This nicely tracks the fact that, usually, if something is valid then its necessitation is true, even though the reverse direction might not hold. Given any element i of the powerset of the substitution functions, there is a truth-theory where the following holds:

Valid (extension): $\ulcorner\Box\phi\urcorner$ is true exactly when ϕ is true under all substitutions in i.

Thus, in effect I obtain (countably) many different theories of modal truth, that reflect different extensions of 'It is (substitutionally) valid'. In each of these systems the only reading of $\Box$ available is the intended: it is truth under all substitutions of the relevant chosen subset; this is contrary to Kripke's semantics, where we needed to rely on the logic generated having specific properties, like being canonical.

It is an open question whether, for some given calculus, there is a truth-theory where validity coincides with provability in the calculus. In such a case, I call the truth-theory 'canonical' for the modal calculus. In the rest of the paper, I show how to develop a canonical theory when possible, and I prove soundness and completeness for the most common (normal) modal logics with constant domain semantics. I then discuss variable domain semantics, as well as the limits and scope of the present proposal.

3.1.5 Why I don't choose a non-reductive account

In an axiomatic, substitutional approach we define truth axiomatically, as a primitive notion. We then read logical expressions homophonically in the definition of truth.

[9]There is quite a bit of debate on whether the same concept of validity applies to first order logic and modal logic or not. Zalta (1988) argues for the first view: validity in modal logic is preservation of truth under all interpretations, and thus some logical truths might be contingent. Etchemendy (2008), too, argues for the first view, but concludes that validity cannot be preservation of truth under all interpretations, because it makes little sense to apply the framework to a vast array of modal logics. I argue for the second view: it is not the same notion of validity at play, here. See Hanson 2006 for a sympathetic point.

We do not try to reduce talk of connectives and quantifiers to talk in a language with no connectives or quantifiers. One might wonder why I do not try the same approach to modal operators. After all, such a move looks like the natural one to do in a substitutional account; also, in this way we would not need to impose any specific interpretation of $\Box$.

There are, however, many technical obstacles to this strategy. First of all, whilst theories of truth for extensional languages are well established, their analogue for intensional languages are in no similar state of development. The axiomatic approach might try to read necessity homophonically. To do so, one usually stipulates commutativity of $\Box$ and satisfaction axiomatically, treating $\Box$ pretty much like negation. Yet, whether this axiom is feasible does depend on what $\Box$ is taken to mean, so it is not clear that we achieve the neutrality on the meaning of $\Box$ we expected. In fact, many doubted this axiom under fairly common readings of $\Box$: even though 'Socrates is necessarily human' is true, it is not clear that necessarily 'Socrates is human' is true, since the sentence might have meant something else completely (Harman 1972, Wallace 1975).[10] Commutativity would not be sound if we read $\Box$ as 'It is valid that', either: 'It is valid that it rains if it rains' is true, yet that 'It rains if it rains' is true is not valid.

Arguably, to work out a completeness theorem, one would need a T-schema for arbitrary modal formulae without satisfaction. Yet, to obtain such a T-schema one will need to strengthen all the axioms with a sufficiently strong modal operator, which must be at least S4. This would put a cap on the modal logics we can provide validity for: they would need to be at least as strong as S4.

3.2 Technical preliminaries

3.2.1 Language

I call L_{ZFC} the language of ZFC. In L_{ZFC} the only relation symbols are '$\in$' and '$=$'.[11] The only connectives and quantifiers are: '$\forall$', '$\neg$', '$\wedge$'. The other common quantifiers and connectives are to be defined as meta-abbreviations. I call L_{Sat} the result of

[10] Some suggested different ways of getting out of the problem, for example by fixing a particular interpretation of the language (Peacocke 1978, Gupta 1978). However, Davies (1978) doubts even this solution because, even if we relativised truth to a particular language (and thus to a particular interpretation), since languages themselves are contingent objects, in some world they would hold no relation to the sentence.

[11] Halbach (2018) builds the theory from ZFC, as well. Substitutional validity for first order logic can be defined using Peano arithmetic as base. I expect that similar results hold for modal logic.

adding a binary satisfaction predicate 'Sat' to L_{ZFC}, and updating the definition of atomic formulae, adding $\ulcorner \mathrm{Sat}(v, w) \urcorner$. L_{Sat} will be our meta-language. The language we are interpreting, on the other hand, is $L_{\Box}$, which is the result of adding a modal operator $\Box$ to L_{Sat}, and extending the definition of well-formed formula, accordingly, adding $\ulcorner \Box\phi \urcorner$ as well-formed. I use $\Diamond\phi$ as an abbreviation for $\neg\Box\neg\phi$.

To recap:

- L_{ZFC}: the language of ZFC;
- L_{Sat}: the result of adding Sat to the syntax of L_{ZFC} and updating the definition of well-formed formula accordingly (the language has no $\Box$). It is our meta-language;
- $L_{\Box}$: the result of adding $\Box$ to the syntax of L_{Sat} and updating the definition of well-formed formula accordingly. It is our object-language.

As we will see, L_{Sat} and not $L_{\Box}$ is our meta-language because we do not need to use $\Box$, but only to mention it, pretty much like in a Kripke semantics where the meta-language is always non-modal.

I code all formulae of $L_{\Box}$ in a natural way in set-theory. There are no function symbols in L_{ZFC}, but functions can be represented in ZFC through suitable formulae. 'Form' is the recursively definable condition of being (the code of) a formula of $L_{\Box}$. I use Greek letters ϕ, ψ etc as meta-variables for things that satisfy 'Form'. For readability, I do not distinguish between a formula and its code. That is, where ψ is some condition and $\ulcorner \phi \urcorner$ the code of ϕ, to say that ϕ is ψ I do not write '$\psi(\ulcorner \phi \urcorner)$' but just '$\psi(\phi)$'.

I do need some coding notation when I need to show the structure of the formula: in that case I either use semi-quotes $\ulcorner \urcorner$ or coding functions. I read $\ulcorner xyz \urcorner$ as the result of conjoining the code of x with the code of y with the code of z. A dot under a symbol represents its correspondent coding function: $\dot{\neg}$ is a function from the code of ϕ to the code of its negation. Thus, the negation of ϕ can be written either as $\ulcorner \neg\phi \urcorner$ or as $\dot{\neg}\phi$. Similarly, $\dot{\in}$ is a function from the (code of) variables v and w to $\ulcorner w \in v \urcorner$ or – which is the same thing – $w \dot{\in} v$. I occasionally abuse notation, using coding functions for connectives that do not belong to the language: $\phi \dot{\rightarrow} \psi$ is an abbreviation for $\dot{\neg}(\phi \dot{\wedge} \dot{\neg}\psi)$. When I need to talk about the numeral of the code of ϕ, I use $\overline{\phi}$ (again, this can be defined in set-theory): this is useful when we need to 'double-quote' something. For example, if I wish to say that if x is true then the sentence that expresses that x is true is true. This latter fact is (more informally) expressed by saying that the string

obtained by writing down the numeral of x and then the code of 'Is true' expresses a true sentence.

To save space I sometimes write $\forall x, y\phi$ instead of $\forall x \forall y \phi$. For functions, often instead of $f(g(x))$ I write $fg(x)$. $|\phi, v_1/w_1, .., v_n/w_n|$ is the result of uniformly substituting v_1 for w_1, .., v_n for w_n in ϕ, and it is recursively definable. ϕ_{at} is a meta-variable for (the code of) atomic formulae (i.e. $\dot{\text{Sat}}(v_n, v_m)$, $v_n \dot{\in} v_m$ or $v_n \dot{=} v_m$).

3.2.2 Substitution function

A substitution function $\mathcal{I}$ that reinterprets the non-logical expressions is defined recursively on formulae of $L_\Box$. I do not write $\mathcal{I}(\ulcorner \phi \urcorner)$, much less $\psi(\ulcorner \mathcal{I}(\ulcorner \phi \urcorner) \urcorner)$, because it makes the text confusing and aesthetically unpleasant.

Substitution functions ought to treat logical expressions homophonically. In the following, I consider logical expressions $\neg$, $\wedge$, $\forall$ and $\Box$. I take a Quinean stance on identity: it is not logical. This can be changed if we wished to, and the proofs arranged to make space for the logicality of identity. A substitution function can treat quantifiers in two different ways: it can relativise them to some condition or not. I call the first relativised substitution, the second unrelativised. If it adds a condition, one might say that quantifiers are not treated as logical constants, after all. A similar objection holds for model theory, where in different models quantifiers range over different domains. I discuss quantification in section 6 in more detail. For now I assume that substitutions are unrelativised.

I do not allow a substitution function to add modal operators, otherwise the system is inconsistent. I return to this issue in section 2.4. Each substitution respects the modal depth of formulae. The modal depth *md* can be defined recursively:

$$\begin{aligned}
&\text{md}(\phi_{at}) = 0 \\
&\text{md}(\dot{\neg}\phi) = \text{md}(\phi) \\
&\text{md}(\phi \dot{\wedge} \psi) = \max(\text{md}(\phi), \text{md}(\psi)) \\
&\text{md}(\forall v \phi) = \text{md}(\phi) \\
&\text{md}(\dot{\Box}\phi) = \text{md}(\phi) + 1
\end{aligned}$$

I call a substitution $\mathcal{I}$ 'modally bounded' whenever the following holds:

$$\forall \phi (\text{md}(\phi) = \text{md}(\mathcal{I}(\phi)))$$

I stipulate that every $\mathcal{I}$ is modally bounded. We will see later that this requirement ensures relative consistency and substitution of identicals.

Usually a substitution function will rename free variables. Variable renaming is natural, because a renaming of variables looks like a correct substitution of the formula. Also, it is used to avoid unintended variable bindings. Yet, we cannot let substitutions rename free variables, because variable renaming disrupts the relation between modality and quantification. Consider $\exists x \Box Fx$. We would like to say that the sentence is true if some thing y satisfies every modally relevant substitution of Fx. To do so, we need to talk about that specific y across different substitutions, and yet it might be that the variable x gets substituted arbitrarily, making us 'lose track' of the thing we were considering through the existential quantifier when we started to evaluate the formula. There is a lack of coordination between the way we rename variables and the unravelling of the formula via the compositional axioms of truth. This lack of coordination also makes substitution of identicals unsound.

For the above reasons, I do not let substitutions rename free variables. To avoid unintended variable bindings, substitutions must uniformly rename only bound variables. I define a formula in 'signed form' as follows:

Definition 1 (Signed form)**.** *A formula ϕ is in signed form if and only if every free variable in ϕ is even, and every bound variable in ϕ is odd.*

I stipulate that $\mathcal{I}$ maps '$\in$', 'Sat' and '=' to three formulae which are all in signed form. I stipulate that in each of these formulae we ordered every free and bound variable: I use α_1 for the first first free variable, α_2 for the second, etc, and β_1 for the first bound variable, β_2 for the second etc. I show how to define $\mathcal{I}$ for the case $\ulcorner v_n \in v_m \urcorner$. Similar definitions hold for $\ulcorner v_n = v_m \urcorner$ and $\ulcorner \mathrm{Sat}(v_n, v_m) \urcorner$. First, where β_1, .., β_s are the bound variables in $\mathcal{I}(\in)$, and k is the maximum between the indices in β_1, .., β_s, define the following transformation τ:

$$\tau(r, \mathcal{I}(\in)) \coloneqq |\mathcal{I}(\in), v_{2(k+r+1)+1}/\beta_1, .., v_{2(k+r+s)+1}/\beta_s|$$

Now, define $\mathcal{I}(v_n \dot{\in} v_m)$ as follows:

$$\mathcal{I}(v_n \dot{\in} v_m) \coloneqq \begin{cases} |\tau(n, \mathcal{I}(\in)), v_n/\alpha_1, v_m/\alpha_2| & \text{if } \max(n,m) = n \\ |\tau(m, \mathcal{I}(\in)), v_m/\alpha_2, v_n/\alpha_1| & \text{if } \max(n,m) = m \end{cases}$$

$\mathcal{I}$ always chooses a bigger odd subscript for bound variables, so the free variables it substitutes do not get bound, and the old (even) free variables α_1, .. ,α_n do not get bound, either. It then substitutes the variable with the biggest subscript first; otherwise, where $\mathcal{I}(\in) = \phi(\alpha_1, \alpha_2)$, $\mathcal{I}(\alpha_2 \dot{\in} \alpha_3) = |\phi(\alpha_1, \alpha_2), \alpha_2/\alpha_1, \alpha_3/\alpha_2|$, which is the unintended $\phi(\alpha_2, \alpha_2)$.

We extend the definition of $\mathcal{I}$ recursively to all formulae of $L_\Box$. By construction, where k is the maximum of all variables in $\mathcal{I}(\psi)$, $u = 2k + 1$.

$$\mathcal{I}(\phi) \coloneqq \begin{cases} \dot{\neg}\mathcal{I}(\psi) & \text{if } \phi \text{ is } \dot{\neg}\psi \\ \mathcal{I}(\psi)\dot{\wedge}\mathcal{I}(\chi) & \text{if } \phi \text{ is } \psi\dot{\wedge}\chi \\ \dot{\forall}u|\mathcal{I}(\psi), u/v| & \text{if } \phi \text{ is } \dot{\forall}v\psi \\ \dot{\Box}\mathcal{I}(\psi) & \text{if } \phi \text{ is } \dot{\Box}\psi. \end{cases}$$

In the proofs below I occasionally omit variables renaming because it makes the formula unreadable, hoping that the reader trusts us with the fact that variable renaming works and we can avoid unintended variable bindings.

3.2.3 Satisfaction

I do not use a concept of simple truth, but rather a concept of satisfaction under an assignment. Simple truth is truth under all assignments. An assignment is a function from the domain of variables: I use Asg for variable assignments, and Gothic letters $\mathfrak{a}$, $\mathfrak{b}$ etc as meta-variables for things that satisfy Asg. By '$\mathfrak{a}^x_v$' I mean the result of changing the value of v to x in $\mathfrak{a}$, keeping everything else the same:

$$\mathfrak{a}^v_x(u) = \begin{cases} x & \text{if } u = v \\ \mathfrak{a}(u) & \text{otherwise.} \end{cases}$$

In the following I discuss different truth-theories, which differ on how they interpret modal formulae. However, all the truth-theories I consider share the following Tarskian axioms, for all (codes of) formulae of $L_\Box$:

$$\forall v, w, \mathfrak{a}(\mathrm{Sat}(v\dot{\in}w, \mathfrak{a}) \leftrightarrow \mathfrak{a}(v) \in \mathfrak{a}(w)) \tag{SM1}$$

$$\forall v, w, \mathfrak{a}(\mathrm{Sat}(v\dot{=}w, \mathfrak{a}) \leftrightarrow \mathfrak{a}(v) = \mathfrak{a}(w)) \tag{SM2}$$

$$\forall \phi, \mathfrak{a}\big(\mathrm{Sat}(\dot{\neg}\phi, \mathfrak{a}) \leftrightarrow \neg\mathrm{Sat}(\phi, \mathfrak{a})\big) \tag{SM3}$$

$$\forall \phi, \psi, \mathfrak{a}\big(\mathrm{Sat}(\phi\dot{\wedge}\psi, \mathfrak{a}) \leftrightarrow (\mathrm{Sat}(\phi, \mathfrak{a}) \wedge \mathrm{Sat}(\psi, \mathfrak{a}))\big) \tag{SM4}$$

$$\forall \mathfrak{a}, \phi(\mathrm{Sat}(\dot{\forall}v\phi, \mathfrak{a}) \leftrightarrow \forall y\mathrm{Sat}(\phi, \mathfrak{a}^y_v)) \tag{SM5}$$

$$\forall v, w, \mathfrak{a}, \phi_{at}\big(\mathfrak{a}(v) = \mathfrak{a}(w) \rightarrow \mathrm{Sat}(\phi_{at} \leftrightarrow |\phi_{at}, v/w|, \mathfrak{a})\big) \tag{SM6}$$

$$\forall \mathfrak{a}, v, \phi\Big(\phi \in L_{\mathsf{ZFC}} \rightarrow \big(\mathrm{Sat}(\dot{\mathrm{Sat}}(\overline{\phi}, v), \mathfrak{a}) \leftrightarrow \mathrm{Sat}(\phi, \mathfrak{a}(v))\big)\Big) \tag{SM7}$$

$$\forall x, y(\neg(\mathrm{Form}(x) \vee \mathrm{Asg}(y)) \rightarrow \neg\mathrm{Sat}(x, y)) \tag{SM8}$$

By extending ZFC with axioms SM1-SM8 and extending the axioms and rules of ZFC to L_{Sat}, we obtain the theory SM. The language of SM is L_{Sat}: it is a non-modal language. The compositional axioms of satisfaction for connectives and quantifiers hold also for modal formulae.

Axioms SM1-SM5 provide a weak type-free theory of satisfaction: they are but the extension to the modal language of the theory Halbach (2018) labels S. Since we lack a T-schema for formulae with the satisfaction predicate, the scope of Sat is opaque: SM6 ensures that substitution of identicals holds regardless, for all atomic formulae. SM6 for ϕ_{at} = $v \dot{\in} u$ or $v \dot{=} u$ is redundant, as it can be proved via SM1 and SM2. We will see how the axiom is enough to ensure that substitution of identicals holds inside the scope of the modal operator, as well.

We can prove a T-schema for formulae of L_{ZFC}:

Lemma 1. *For all formulae ϕ of L_{ZFC} with free variables x_1, .., x_n, the following holds:*

$$\mathsf{SM} \vdash \forall \mathfrak{a}\big(\mathrm{Sat}(\phi(x_1, .., x_n), \mathfrak{a}) \leftrightarrow \phi(\mathfrak{a}(x_1), .., \mathfrak{a}(x_n))\big)$$

Proof. The proof is by induction on the complexity of ϕ using axioms SM1-SM5. □

SM7 allows us to iterate satisfaction twice, if we start from a formula of set-theory. If iteration was extended to formulae with the satisfaction predicate, the resulting theory would be inconsistent *modulo* SM3 (see lemma 14.3 in Halbach 2014, 146). However, iteration restricted to satisfaction-free formulae is harmless. Iteration will prove useful for the adequacy theorem.

The axioms of SM are debatable and can be changed if we wish to. With some modifications, what we prove can probably be done in another truth-theory of similar strength. If we choose a truth-theory like Kripke-Feferman[12], the resulting theory of validity will probably be non-classical and paracomplete. A classic, typed Tarskian satisfaction may work too, even though the adequacy proof will have to be changed.[13] In a typed system, however, logical consequence cannot be defined for all first order languages, but only for weaker object-languages. We lose self-applicability: validity cannot be defined in a language for that language.

3.2.4 Modal truth theories

We can provide different readings of $\Box$ through different truth theories, which result from adding specific axioms that cover the clause for $\dot{\Box}\phi$. The simplest equates its satisfaction to the satisfaction of ϕ under all substitutions:

$$\forall \mathfrak{a}, \phi\big(\mathrm{Sat}(\dot{\Box}\phi, \mathfrak{a}) \leftrightarrow \forall \mathcal{I}\ \mathrm{Sat}(\mathcal{I}(\phi), \mathfrak{a})\big)$$

[12] See Feferman (1991). For a discussion, see Halbach (2014).

[13] I suppose that one will need either to add some syntax to the language, thus strengthening the substitution functions, or have two typed truth-predicates, and work in a meta-meta-language to mimic axiom SM7.

We call a formula 'simply true' when satisfied under all assignments. Via such axiom, $\dot{\Box}\phi$ is simply true exactly when it is simply true under all substitutions: when it is logically true.[14] Different readings of modality can be given through different modifications of the above truth-clause. In general, for any collection of substitutions, there is a truth-theory that makes it the set of modally relevant substitutions. Where $\mathfrak{I}$ is the set of substitution functions defined as above over $L_{\Box}$, for any element i of $\wp\mathfrak{I}$, there is a theory of truth $\mathsf{SM}+\mathsf{SMi}$, which is the result of adding to SM the following axiom:

$$\forall \mathfrak{a}, \phi\big(\mathrm{Sat}(\dot{\Box}\phi, \mathfrak{a}) \leftrightarrow \forall \mathcal{I}(\mathcal{I} \in i \rightarrow \mathrm{Sat}(\mathcal{I}(\phi), \mathfrak{a}))\big) \tag{SMi}$$

There are countably many $\mathcal{I}$, so there are uncountably many different readings of $\Box$ available. The first one I considered equated i with the class of $\mathcal{I}$. It is important to point out that, even though we have many ways of interpreting $\Box$, when we specify validity we always need to choose one first, and then work in that truth-theory only: we never define validity by quantifying over different ways of interpreting $\Box$. If we did, 'truth' would not be simple truth when we define validity, but rather truth under different interpretations of the $\Box$ – in so doing we would betray the main assumption behind the substitutional account.

If substitutions are not modally bounded the system might be inconsistent. For example, consider $i = \{\mathcal{I} : \mathcal{I}(F) = \dot{\neg}\Box F\}$, where F is some designated atomic formula of $L_{\Box}$. Via $\mathsf{SM}+\mathsf{SMi}$, $\mathrm{Sat}(\dot{\Box}F, \mathfrak{a})$ iff $\mathrm{Sat}(\mathcal{I}(F), \mathfrak{a})$, for any $\mathcal{I}$ in i. That is, iff for any $\mathcal{I}$ such that $\mathcal{I}(F) = \dot{\neg}\Box F$, $\mathrm{Sat}(\mathcal{I}(F), \mathfrak{a})$; that is, iff $\neg\mathrm{Sat}(\dot{\Box}F, \mathfrak{a})$. Contradiction. However, when substitutions are modally bounded, we have reasons to believe that the system is consistent. I call S the restriction of the axioms of Sat to the non-modal language (L_{Sat}). S is a very weak type-free theory, and it is what is true at level 0 of modal depth. We prove the following lemma:

Lemma 2. *Give any* $\mathsf{SM}+\mathsf{SMi}$*, given any level of modal depth* n*, what is true at level* n *in* $\mathsf{SM}+\mathsf{SMi}$ *is inconsistent only if what is true at level* 0 *is inconsistent.*

Proof. Take any SMi, and any ϕ, and say that $\mathrm{md}(\phi) = n$. Via SMi, $\mathrm{Sat}(\dot{\Box}\phi, \mathfrak{a})$ exactly when $\mathrm{Sat}(\mathcal{I}(\phi), \mathfrak{a})$, for any $\mathcal{I} \in i$. Since $\mathcal{I}$ is modally bounded, $\mathrm{md}(\phi) = \mathrm{md}(\mathcal{I}(\phi))$, and thus $\mathrm{md}(\mathcal{I}(\phi)) < \mathrm{md}(\dot{\Box}\phi)$.

Now, if $\mathrm{md}(\phi)$ is 0, S would be inconsistent, so the lemma holds. Assume that the lemma holds at any level strictly lower than n, with $n > 0$. Suppose that the lemma

[14] $\dot{\Box}\phi$ might be satisfied but not logically true, like for $\phi(v)\dot{\rightarrow}\phi(u)$, where $\mathfrak{a}(v) = \mathfrak{a}(u)$ (since substitutions do not rename free variables). To be true is not to be satisfied under some assignment: to be true is to be satisfied under all assignments.

does not hold for n. Then, for some ϕ of modal depth n, $\mathsf{SM}+\mathsf{SMi} \vdash \mathrm{Sat}(\phi \dot{\wedge} \dot{\neg}\phi, x)$, Via repeated applications of the compositional axioms for Sat, in virtue of the definition of well-formed formula and modal depth, we end up evaluating some sub-formula of ϕ of form $\Box\psi$. Via SMi, the satisfaction of $\Box\psi$ is equivalent to the satisfaction of $\mathcal{I}(\psi)$, for all $\mathcal{I} \in i$. Since each $\mathcal{I}$ is modally bounded, $\mathrm{md}(\mathcal{I}(\psi)) < n$, so now we are evaluating formulae that, by induction hypothesis, are not inconsistent, so $\phi \wedge \neg\phi$ cannot be true at level n, contrary to our hypothesis. We conclude that the lemma holds for level n, too, and so for all n by strong induction on the modal depth. □

3.3 Substitutional logical consequence

Substitutional logical consequence is defined in the language of L_{Sat}. We are interpreting $L_{\Box}$. I denote substitutional logical consequence with the symbol '$\vDash_S$'.

Definition 2 (Substitutional Validity)**.** *Where Δ and ϕ are sets of formulae and a formula of $L_{\Box}$, respectively:*

$$\forall\Delta, \phi\Big(\Delta \vDash_S \phi \coloneqq \forall\mathfrak{a}\forall\mathcal{I}\big(\forall\psi(\psi \in \Delta \to \mathrm{Sat}(\mathcal{I}(\psi), \mathfrak{a})) \to \mathrm{Sat}(\mathcal{I}(\phi), \mathfrak{a})\big)\Big)$$

For the proofs in the next sections, it is useful to highlight some properties of logical validity. A quite obvious but useful lemma is the following:

Lemma 3. *Given any* $\mathsf{SM}+\mathsf{SMi}$*:*

$$\mathsf{SM}+\mathsf{SMi} \vdash \forall\phi\psi(\phi) \to \forall\phi, \mathcal{I}\psi(\mathcal{I}(\phi))$$

$\mathcal{I}(\phi)$ is always a formula. Therefore, if some condition ψ holds of all formulae, it also holds of all their substitutions.

Theorem 1. *Given any* $\mathsf{SM}+\mathsf{SMi}$*, where FOL is first order logic without identity, and ϕ and Δ are meta-variables for formulae and sets of formulae of $L_{\Box}$, the following holds:*

$$\mathsf{SM}+\mathsf{SMi} \vdash \forall\phi, \Delta(\Delta \vdash_{FOL} \phi \to \Delta \vDash_S \phi)$$

Proof. The proof is by induction on the length of proofs: the axioms hold for all formulae and thus for all substitutions, via lemma 3. In particular, we can safely prove soundness for universal instantiation and generalisation even for *de re* modal formulae, since $\mathcal{I}$ does not change free variables.[15] □

[15]Here is what I mean. Suppose each $\mathcal{I}$ in i did rename variables via some recursive mapping $'$, in particular $v'_n = v_n$, $v'_m = v_n$ and $v'_k \neq v_n$. Suppose further that $\mathcal{I}(x \dot{\in} y) = x' \dot{=} y'$ and that $\mathfrak{a}(v_n) = n$. Then $\mathrm{Sat}(\dot{\forall} v_n \dot{\Box} v_n \dot{\in} v_m, \mathfrak{a})$: by the axioms for satisfaction, $\forall y, \mathcal{I}(\mathcal{I} \in i \to \mathrm{Sat}(\mathcal{I}(v_n \dot{\in} v_m), \mathfrak{a}^y_{v_n}))$ and, by construction of i, $\forall y \mathrm{Sat}(v_n \dot{=} v_n, \mathfrak{a}^y_{v_n})$. Yet, $\mathrm{Sat}(|\dot{\Box} v_n \dot{\in} v_m, v_k/v_m|, \mathfrak{a})$ fails, because it is equivalent to $\mathrm{Sat}(v_n \dot{=} v'_k, \mathfrak{a})$, with $v_n \neq v'_k$. Universal instantiation fails!

Granted that ϕ is logically necessary exactly when it is logically true, as expected, where i is the class of all $\mathcal{I}$, necessity is logical necessity, and $\vDash_S \phi$ is equivalent to $\forall \mathfrak{a} \mathrm{Sat}(\dot{\Box}\phi, \mathfrak{a})$: the logical truth of a formula is equivalent to the truth of its logical necessitation. For arguments, we can prove that, if ϕ follows from a conjunction $\psi_1, .., \psi_n$, then the strict conditional from $\psi_1 \wedge .. \wedge \psi_n$ to ϕ is satisfied. We cannot show that being logically true is equivalent to being logically necessary, because there is no T-schema for modal sentences, nor in general for all sentences of L_{Sat}, on pain of contradiction. For similar reasons, in any truth-theory we consider we cannot show that logical truths are necessary, but only that their necessitation is true.

Substitution of identicals is satisfied. The result relies on substitutions being modally bounded.

Lemma 4. *Given any* $\mathsf{SM} + \mathsf{SMi}$*:*

$$\mathsf{SM} + \mathsf{SMi} \vdash \forall v, w, \mathfrak{a}, \phi\big(\mathfrak{a}(v) = \mathfrak{a}(w) \rightarrow (\mathrm{Sat}(\phi, \mathfrak{a}) \leftrightarrow \mathrm{Sat}(|\phi, v/w|, \mathfrak{a}))\big)$$

Proof. The proof is by induction on the modal depth of formulae. When the modal depth is 0, we prove the lemma for atomic formulae via SM6, and for the other non-modal formulae via SM3-5 (via induction on the complexity of ϕ). Assume that the lemma holds for all ψ whose md is less than m. Then whenever ϕ has $md = m$, we have added a $\Box$ to some subformula χ of ϕ. Assume $\mathrm{Sat}(\dot{\Box}\chi, \mathfrak{a})$ but $\neg\mathrm{Sat}(|\dot{\Box}\chi, v/u|, \mathfrak{a})$, even though $\mathfrak{a}(v) = \mathfrak{a}(u)$. Then for any $\mathcal{H}$ in i, $\mathrm{Sat}(\mathcal{H}(\chi), \mathfrak{a})$. Since by construction, $\mathcal{H}$ does not rename variables, $|\mathcal{H}(\chi), v/u| = \mathcal{H}(|\chi, v/u|)$. So, for some $\mathcal{H}$ in i, $\neg\mathrm{Sat}(\mathcal{H}(|\chi, v/u|), \mathfrak{a})$. By construction of $\mathcal{H}$, $\mathcal{H}$ is modally bounded, so for any ψ $\mathcal{H}(\psi)$ has the same modal depth of ψ. We can therefore apply the induction hypothesis, and conclude that for any $\mathcal{H}$ in i, $\mathrm{Sat}(\mathcal{H}(\chi), \mathfrak{a}) \leftrightarrow \mathrm{Sat}(\mathcal{H}(|\chi, v/u|), \mathfrak{a})$, which contradicts our assumption that $\neg\mathrm{Sat}(\mathcal{H}(|\chi, v/u|), \mathfrak{a})$ for some $\mathcal{H}$ in i. Since the lemma holds for $\dot{\Box}\chi$, we can show that it holds for ϕ, via the definition of modal depth and well-formed formula. □

Lemma 5. *Given any* $\mathsf{SM} + \mathsf{SMi}$*,*

$$\mathsf{SM} + \mathsf{SMi} \vdash \forall \phi(\vDash_S \phi \rightarrow \forall \mathfrak{a} \mathrm{Sat}(\phi, \mathfrak{a}))$$

Proof. If $\vDash_S \phi$ then ϕ is true under all $\mathfrak{a}$ and $\mathcal{I}$. In particular, it is true under the identity function, so $\mathrm{Sat}(\phi, \mathfrak{a})$, for any $\mathfrak{a}$. □

Lemma 6 (Additivity)**.** *For any* $\mathcal{I}$ *and* $\mathcal{I}'$*, there exists a* $\mathcal{H}^{\mathcal{I}+\mathcal{I}'}$ *such that, for all* ϕ*,* $\mathcal{I}'\mathcal{I}(\phi) = \mathcal{H}^{\mathcal{I}+\mathcal{I}'}(\phi)$

Proof. I construct the relevant $\mathcal{H}$ as follows:

$$\mathcal{H}(\in) = \mathcal{I}'\mathcal{I}(\in)$$
$$\mathcal{H}(\mathrm{Sat}) = \mathcal{I}'\mathcal{I}(\mathrm{Sat})$$
$$\mathcal{H}(=) = \mathcal{I}'\mathcal{I}(=)$$

The proof is a simple induction on the complexity of formulae.[16] □

We can read the definition of logical truth and necessity through an accessibility relation between formulae. A formula ψ is *logically accessible* to ϕ if ψ is a substitution of ϕ. So, ϕ is logically true exactly when all the formulae logically accessible to it are true. Say that i is the set of modally relevant substitutions: ψ is *modally accessible* to ϕ if there is a substitution function in i from ϕ to ψ. $\dot{\Box}\phi$ is true iff every modally accessible ψ from ϕ is true. The process can be iterated: $\dot{\Box}\dot{\Box}\phi$ is true iff every ψ modally accessible from ϕ is such that every χ modally accessible from ψ is true.

Definition 3 (Accessibility relations)**.**

$$\mathcal{L}(\phi, \psi) \coloneqq \exists\mathcal{I}\mathcal{I}(\phi) = \psi.$$
$$\mathcal{N}(\phi, \psi) \coloneqq \exists\mathcal{I}(\mathcal{I} \in i \wedge \mathcal{I}(\phi) = \psi).$$

Lemma 7. *Given any* $\mathsf{SM} + \mathsf{SMi}$ *the following holds:*

(L8.1) $\forall\phi(\vDash_S \phi \leftrightarrow \forall\psi, \mathfrak{a}(\mathcal{L}(\phi, \psi) \rightarrow \mathrm{Sat}(\psi, \mathfrak{a})))$
(L8.2) $\forall\phi\forall\mathfrak{a}\big(\mathrm{Sat}(\dot{\Box}\phi, \mathfrak{a}) \leftrightarrow \forall\psi(\mathcal{N}(\phi, \psi) \rightarrow \mathrm{Sat}(\psi, \mathfrak{a}))\big)$

Proof. The proofs are trivial given the definitions. □

3.3.1 Logical necessity

$\mathcal{L}$ is clearly reflexive and transitive, as lemma 6 and 5 show, respectively. It is not symmetric, though: an atomic formula might be swapped with a complex one, making us not able to either 'go back' to the original formula, nor to anything equivalent to the original formula. It follows that axiom B fails, and so does axiom 5, where i

[16] The reader might worry that lemma 6 relies on substitutions being unrelativised. However, this is not the case: if we allow them to add a condition δ to the quantifiers, we just need to set $\delta^{\mathcal{H}} = \delta^{\mathcal{I}} \dot{\wedge} \delta^{\mathcal{I}'}$. For the induction step in the quantified case, we use the fact that $(\phi \wedge \psi) \rightarrow \chi$ is equivalent to $\phi \rightarrow (\psi \rightarrow \chi)$ under satisfaction, via axioms SM3-SM4 and the fact that $\phi \rightarrow \psi$ iff $\neg(\phi \wedge \neg\psi)$.

is the set of all substitution functions. For example, assume that $\mathrm{Sat}(v \dot{\in} u, \mathfrak{a})$. We show that $\neg\mathrm{Sat}(\dot{\Box}\dot{\Diamond} v \dot{\in} u, \mathfrak{a})$. Assume the contrary, then, using the axioms for truth we obtain: $\forall\mathcal{H}\exists\mathcal{H}'\mathrm{Sat}(\mathcal{H}'\mathcal{H}(v \dot{\in} u), \mathfrak{a})$. Yet, consider $\mathcal{H}(v \dot{\in} u) = v \dot{\in} u \dot{\wedge} \dot{\neg} v \dot{\in} u$. Any $\mathcal{H}'$ is such that $\neg\mathrm{Sat}(\mathcal{H}'\mathcal{H}(v \dot{\in} u), \mathfrak{a})$, contrary to our assumption.

In a well-known paper, Burgess (1999) argues that the correct modality for logical necessity is S5. Our $\Box$, however, when interpreted through logical accessibility, is not S5. Is this a problem? Burgess takes the logical $\Box$ to be read as 'It is true by logical form alone'. No wonder he disagrees with us, for he has a different definition of $\Box$ in mind: our definition states that $\Box\phi$ is true when all correct substitutions of ϕ are true. The two definitions come apart. His argument at page 84-5 for an axiom of S5 ($\neg\Box\phi \rightarrow \Box\neg\Box\phi$) makes this shift apparent: he is (rightly) assuming that, if ψ has the same logical form as ϕ, and χ has the same logical form as ϕ, then ψ and χ have the same logical form, as well. This is false with our accessibility relation: some complex ϕ is a correct substitution of an atomic Fx; Gx is also a correct substitution of Fx; yet Gx is not a correct substitution of ϕ.

Burgess also states that, if $\dot{\Box}\phi$ is false, then some ψ of the same form of ϕ must be false, which implies the following rule: if all ψ of the same form of ϕ are true, then $\dot{\Box}\phi$ is true. Even though he does not state it, the right to left direction is obvious: if $\dot{\Box}\phi$ is true (i.e. ϕ is true by logical form alone), then any ψ of the same logical form as ϕ are true. So, according to Burgess, $\dot{\Box}\phi$ is true exactly when all ψ of the same form are true. Yet, our logical accessibility behaves differently. Since a correct substitution of an atomic formula can be a very complex formula with radically different logical form, the two accounts are different, and mixing them together only brings confusion. I suppose that we can recover the notion Burgess has in mind if we consider correct substitutions that do not change the complexity of the formula substituted. Then one can check that the generated modal logic is S5.

3.4 Soundness

In this section I discuss soundness. I first consider a modal logic where Barcan and Converse are provable. That is, the modal logics I consider for now have the following axioms (of which some might be redundant in some of them):

BF $\forall v\Box\phi \rightarrow \Box\forall v\phi$

CBF $\Box\forall v\phi \rightarrow \forall v\Box\phi$

3.4.1 Soundness for K

K is first order logic plus BF, CBF and the following:

Nec $\vdash_K \phi \Rightarrow \vdash_K \Box\phi$

K $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$

K is sound with any theory of truth we consider, for the following holds:

Lemma 8. *Given any* $\mathsf{SM} + \mathsf{SMi}$*:*

$$\mathsf{SM} + \mathsf{SMi} \vdash \forall\phi(\vDash_S \phi \to \vDash_S \dot{\Box}\phi)$$

Proof. Given any $\mathsf{SM} + \mathsf{SMi}$:

$$\begin{aligned}
\vDash_S \phi \quad &\leftrightarrow \forall\mathcal{I}, \mathfrak{a}\, \mathrm{Sat}(\mathcal{I}(\phi), \mathfrak{a}) \\
&\to \forall\mathcal{I}, \mathcal{I}', \mathfrak{a}\, \mathrm{Sat}(\mathcal{H}^{\mathcal{I}+\mathcal{I}'}(\phi), \mathfrak{a}) \\
&\to \forall\mathcal{I}, \mathcal{I}', \mathfrak{a}\, \mathrm{Sat}(\mathcal{I}\mathcal{I}'(\phi), \mathfrak{a}) \\
&\to \forall\mathcal{I}, \mathcal{I}', \mathfrak{a}\, (\mathcal{I} \in i \to \mathrm{Sat}(\mathcal{I}\mathcal{I}'(\phi), \mathfrak{a})) \\
&\to \forall\mathcal{I}', \mathfrak{a}\, \mathrm{Sat}(\dot{\Box}\mathcal{I}'(\phi), \mathfrak{a}) \\
&\to \forall\mathcal{I}', \mathfrak{a}\, \mathrm{Sat}(\mathcal{I}'(\dot{\Box}\phi), \mathfrak{a}) \\
&\to \vDash_S \dot{\Box}\phi
\end{aligned}$$

□

Lemma 9. *Given any* $\mathsf{SM} + \mathsf{SMi}$*:* $\mathsf{SM} + \mathsf{SMi} \vdash \forall\phi, \psi(\vDash_S \dot{\Box}(\phi \dot{\to} \psi) \dot{\to} (\dot{\Box}\phi \dot{\to} \dot{\Box}\psi))$

Proof. Via the axioms for truth, $\mathrm{Sat}(\dot{\Box}(\phi \dot{\to} \psi), \mathfrak{a})$ is equivalent to $\forall\mathcal{I}(\mathcal{I} \in i \to (\mathrm{Sat}(\mathcal{I}(\phi), \mathfrak{a}) \to \mathrm{Sat}(\mathcal{I}(\psi), \mathfrak{a})))$, and $\mathrm{Sat}(\dot{\Box}\phi, \mathfrak{a})$ to $\forall\mathcal{I}(\mathcal{I} \in i \to \mathrm{Sat}(\mathcal{I}(\phi), \mathfrak{a}))$. The proof easily follows by logic. □

Lemma 10. *Given any* $\mathsf{SM} + \mathsf{SMi}$*:* $\mathsf{SM} + \mathsf{SMi} \vdash \forall\phi \vDash_S \dot{\Box}\dot{\forall} v\phi \dot{\leftrightarrow} \dot{\forall} v\dot{\Box}\phi$

Proof. Note that $|\mathcal{I}(\phi), u/v| = \mathcal{I}(|\phi, u/v|)$ where u and v are free. We show the equivalence as follows, where $\mathcal{H}$ are the substitutions in i:

$$\begin{aligned}
&\mathrm{Sat}(\dot{\Box}\dot{\forall} v\phi, \mathfrak{a}) && \mathrm{Sat}(\dot{\forall} v\dot{\Box}\phi, \mathfrak{a}) \\
\leftrightarrow &\forall\mathcal{H}\mathrm{Sat}(\mathcal{H}(\dot{\forall} v\phi), \mathfrak{a}) && \leftrightarrow \mathrm{Sat}(\dot{\forall} u\dot{\Box}|\phi, u/v|, \mathfrak{a}) \\
\leftrightarrow &\forall\mathcal{H}\mathrm{Sat}(\dot{\forall} u\mathcal{H}(|\phi, u/v|), \mathfrak{a}) && \leftrightarrow \forall y\mathrm{Sat}(\dot{\Box}|\phi, u/v|, \mathfrak{a}^y_u) \\
\leftrightarrow &\forall\mathcal{H}\forall y\mathrm{Sat}(\mathcal{H}(|\phi, u/v|), \mathfrak{a}^y_u) && \leftrightarrow \forall\mathcal{H}\forall y\mathrm{Sat}(\mathcal{H}(|\phi, u/v|), \mathfrak{a}^y_u)
\end{aligned}$$

The last two lines are the same. □

Theorem 2. *K is sound in any* $\mathsf{SM} + \mathsf{SMi}$

Proof. The proof is by induction on the length of proofs of K. By theorem 1, lemma 9 and 10, each proof of length 0 is valid. The induction step follows by theorem 1 and lemmas 8, 9 and 10. □

Since no matter which i we consider K is sound, the minimal logic for which we can define validity in the present substitutional system is K. The system is unable to define validity adequately for non-normal modal logics, exactly like (standard) Kripke semantics.

3.4.2 Soundness for other normal modal calculi

I give some examples of relations between features of $\mathcal{N}$ and modal axioms, as well as a more general way of checking soundness.

M $\Box\phi \rightarrow \phi$

B $\phi \rightarrow \Box\Diamond\phi$

4 $\Box\phi \rightarrow \Box\Box\phi$

5 $\Diamond\phi \rightarrow \Box\Diamond\phi$

First and foremost, one can check that there is a nice correspondence between features of $\mathcal{N}$ and features of the accessibility relation in a Kripke semantics:

Lemma 11 (Conditions on $\mathcal{N}$)**.**

(L12.1) When $\mathcal{N}$ is reflexive, M is valid;

(L12.2) When $\mathcal{N}$ is symmetric, B is valid;

(L12.3) When $\mathcal{N}$ is transitive, 4 is valid;

(L12.4) When $\mathcal{N}$ is Euclidean, 5 is valid.

Proof. I prove lemma 11.1-2. Similar proofs work for 11.3-4.

Suppose $\mathrm{Sat}(\dot{\Box}\phi, \mathfrak{a})$. By lemma 7.2, $\forall\psi(\mathcal{N}(\phi,\psi) \rightarrow \mathrm{Sat}(\psi, \mathfrak{a}))$. $\mathcal{N}$ is reflexive, so $\mathrm{Sat}(\phi, \mathfrak{a})$. ϕ and $\mathfrak{a}$ were arbitrary, so I can generalise to any ϕ and $\mathfrak{a}$ and, via lemma 3, to any $\mathcal{I}$ and $\mathfrak{a}$.

For lemma 11.2, assume that $\mathrm{Sat}(\phi, \mathfrak{a})$ and yet $\neg\mathrm{Sat}(\dot{\Box}\dot{\Diamond}\phi, \mathfrak{a})$. Where $\mathcal{H}$ and $\mathcal{I}$ are all in i, via the axioms for truth, this second formula is equivalent to the following: $\exists\mathcal{H}\forall\mathcal{I}\neg\mathrm{Sat}(\mathcal{I}\mathcal{H}(\phi), \mathfrak{a})$. By definition of $\mathcal{N}$, trivially $\mathcal{N}(\phi, \mathcal{H}(\phi))$. $\mathcal{N}$ is

symmetric, and since $\mathcal{N}(\phi, \mathcal{H}(\phi))$, $\mathcal{N}(\mathcal{H}(\phi), \phi)$, meaning that there is a $\mathcal{I}$ such that $\mathcal{I}\mathcal{H}(\phi) = \phi$. But $\mathrm{Sat}(\phi, \mathfrak{a})$, so $\mathrm{Sat}(\mathcal{I}\mathcal{H}(\phi), \mathfrak{a})$, contrary to our assumption that, for all $\mathcal{I}$, $\neg\mathrm{Sat}(\mathcal{I}\mathcal{H}(\phi), \mathfrak{a})$. □

The problem is that these requirements are far too strict and don't say much. In most of the cases we will consider, axiom T is sound even though $\mathcal{N}$ is not reflexive, and similarly for the other axioms and features of $\mathcal{N}$. We offer now a stronger resolution strategy to check soundness.

Definition 4 (Path)**.**

- $\phi \approx \psi : \forall\mathfrak{a}\mathrm{Sat}(\phi \dot{\leftrightarrow} \psi, \mathfrak{a})$*;*
- *There is a direct path from ϕ to ψ: $\mathcal{N}(\phi, \psi)$;*
- *There is an indirect path from ϕ to ψ: $\exists\chi(\mathcal{N}(\phi, \chi) \wedge \chi \approx \psi)$*
- *Path-reflexivity: for any ϕ, there is always a path from ϕ to ϕ;*
- *Path-symmetry: for any ϕ and ψ, if there is a direct path from ϕ to ψ, there is an indirect path from ψ to ϕ;*
- *Path-transitivity: for any ϕ, ψ, χ, if there is a direct path from ϕ to ψ and ψ to χ, then there is an indirect path from ϕ to χ;*
- *Path-Euclideanicity: for any ϕ, ψ, χ, if there is a direct path from ϕ to ψ and ϕ to χ, then there is an indirect path from ψ to χ.*

Clearly, since anything is self-equivalent, if there is a direct path between x and y there is also an indirect one. We offer the following lemma:

Lemma 12.

(L13.1) M is valid if path-reflexivity holds;

(L13.2) B is valid if path-symmetry holds;

(L13.3) 4 is valid if path-transitivity holds;

(L13.4) 5 is valid if path-Euclideanicity holds.

Proof. The proof works as in lemma 11, noting that an indirect path is enough. For example, about M if there is always a path to ϕ from ϕ, then there must be a $\mathcal{H}$ in i where $\mathcal{H}(\phi) \approx \phi$, so $\mathrm{Sat}(\dot{\Box}\phi, \mathfrak{a})$ implies $\mathrm{Sat}(\mathcal{H}(\phi), \mathfrak{a})$, so $\mathrm{Sat}(\phi, \mathfrak{a})$. Similar reasoning works for the other conditions. □

System T is K plus axiom M. System S4 is K plus 4 and M. S5 is K plus M plus 5.

Theorem 3 (Soundness for T, S4, S5)**.**

(T3.1) Calculus T is sound for validity in any SM+SMi *where path-reflexivity holds;*

(T3.2) Calculus S4 is sound for validity in any SM + SMi *where path-reflexivity and path-transitivity holds;*

(T3.3) Calculus S5 is sound for validity in any SM + SMi *where path-reflexivity and path-Euclideanicity holds.*

Proof. The proofs are by induction on the length of proofs of T, S4 and S5. □

3.5 Completeness

In this section I explore some completeness results. Completeness in modal logic is a notoriously complex matter: some techniques work for some logics but not for others and, in any case, there are continuum many Kripke incomplete normal modal logics.[17] For now I focus on constant domain semantics.

3.5.1 Canonical model and adequacy

For canonical modal logics, we can construct a canonical truth-theory. A truth-theory X is canonical for a modal calculus Y whenever Y is adequate for validity defined in X. That is, when X proves the following:

$$\forall\phi(Y \vdash \phi \leftrightarrow \vDash_S \phi)$$

I call a choice i of $\mathcal{I}$ canonical for Y if the canonical truth-theory for Y is SM + SMi. I use a Henkin-style completeness proof for the calculus K, by specifying the canonical i Similar proofs can be run for other canonical modal calculi. To prove the adequacy theorem, we need to prove the following 'substitutional truth-lemma':

Lemma 13 (Substitutional Truth-lemma)**.** *There is a truth-theory* SM + SMξ *where the following holds, where* Γ *is a meta-variable for Henkin extensions of K:*

$$\exists\mathfrak{c}\forall\Gamma\exists\mathcal{I}\forall\phi\big(\phi \in \Gamma \leftrightarrow \mathrm{Sat}(\mathcal{I}(\phi), \mathfrak{c})\big)$$

[17] See for example theorem 6.22 in Chagrov and Zakharyaschev 1997, 173 and Miyazaki 2007.

The rest of the section contains the proof of the lemma, and finally the adequacy theorem that follows from it.

The canonical Kripke model is usually built by taking as points of the model all complete and consistent extensions of the modal calculus with certain properties, which we may call 'Henkin extensions'. It is then shown that, with a suitable choice of accessibility relation, membership to the extension is equivalent to truth at that point. I do something similar: the points in my system are substitutions, and the accessibility relation is $\mathcal{N}$.

A first issue is that the cardinality of the canonical model for K is the continuum, yet we have only denumerably many substitution functions. We have two options: one is to add uncountably many variables. Alternatively, we can appeal to a theorem similar to the downward Löwenheim-Skolem:

Theorem 4. *Given any consistent extension of K, for any of its characteristic Kripke model there is an equivalent countable model (i.e. a model with at most countably many worlds).*[18]

Proof. For a proof, see Chagrov and Zakharyaschev 1997, theorem 6.29. □

By theorem 4, there is a countable equivalent to the uncountable canonical model: its countable filtration; it follows that our denumerable substitutions are enough.

A Henkin extension of K for the language of $L_\Box$ is a set of sentences Γ with the following properties:

$$\forall\phi(\mathsf{K} \vdash \phi \to \phi \in \Gamma) \qquad \text{(K-conservativity)}$$

$$\forall\phi(\dot{\neg}\phi \in \Gamma \to \phi \notin \Gamma) \qquad \text{(Consistency)}$$

$$\forall\phi(\phi \notin \Gamma \to \dot{\neg}\phi \in \Gamma) \qquad \text{(Maximality)}$$

$$\forall\phi, v\exists v'(\ulcorner \exists v\phi \to |\phi, v'/v| \urcorner \in \Gamma) \qquad \text{(Henkin witness)}$$

Where v and v' are meta-variables for variables. The Henkin witness property (Henkin property, for short) is usually achieved by adding constants to the language; here we use variables as witnesses, as in Halbach 2020. We also use variables as proxies for points of the Kripke model. Because of this two-fold use, and because substitutions cannot switch free variables, variables are to be handled carefully to avoid variable clashes. Out of notational convenience and to make the proofs more readable, in the following I assume that variables in the language have a subscript and a superscript,

[18] The theorem is from Makinson 1966. Surprisingly, it does not hold for frames, see for example Chagrov and Zakharyaschev 1997, 189.

the latter being either 0 or 1. It is routine logic homework to show that everything that can be done using superscripts and subscripts can be done using only subscripts: we can always get rid of one superscript $m > 0$ in the following way: first we set $v_n^{m-1} = v_{2n+1}^{m-1}$, and then we set $v_n^m = v_{2n}^{m-1}$. By reiterated applications of the process, we end up with only one superscript: 0, and then we just set $v_n^0 = v_n$. When we translate v_n^1 as an even variable and v_n^0 as an odd variable, in the new notation to be in signed form means that every free variable has superscript 1 and every bound variable has superscript 0.

Given a consistent set of formulae, we build Henkin extensions in a somewhat unorthodox way: first, in each formula we get rid of the superscript 1 in the way just described: now each variable of each formula has superscript 0. Then, we use variables with superscript 1 as witnesses, which are guaranteed to be fresh by construction. We then again get rid of the superscript 1: we are left with formulae whose variables have all superscript 0. There is no loss of generality: as just shown, superscripts do not add any expressive strength to the language. All formulae of the Henkin extensions have variables with superscript 0 (i.e. odd subscripts, if we had no superscripts).

We assume we have well-ordered the countable subset of Henkin extensions we need to match the countable filtration of the canonical model in a suitable way. Given such a well-order, define $g(\Gamma) = n$ iff Γ is the n-th Henkin extension. We use the following two variable assignments $\mathfrak{c}$ and $\mathfrak{b}$:

$$\mathfrak{c}(v_n^0) = n \text{ for all } n \in \omega \qquad \mathfrak{b}(v_{g(\Gamma)}^1) = \Gamma \text{ for all } \Gamma$$
$$\mathfrak{c}(v_3^1) = \mathfrak{b}$$

The values not listed are set to 0 and do not matter.

For any Γ, to define its characteristic substitution, called $\mathcal{I}^\Gamma$, we only need to set the substitution of the atomic formulae. We do so for '$\in$', similar definitions hold for 'Sat' and '='. Henkin extensions have only formulae with variables with superscript 0, by construction. Thus, our aim is to show that $\mathrm{Sat}(\mathcal{I}^\Gamma(v_n^0 \dot{\in} v_m^0), \mathfrak{c}) \leftrightarrow v_n^0 \dot{\in} v_m^0 \in \Gamma$. We define the following recursive function f:

$$f(x) = \begin{cases} x & \text{if } x \in \omega \\ 0 & \text{otherwise.} \end{cases}$$

$\mathcal{I}^\Gamma(\in)$ is the following formula, where $g(\Gamma)$ is k:

$$\forall x, y, t, z\Big((x = f(v_1^1) \wedge y = f(v_2^1) \wedge t = \ulcorner v_x^0 \in v_y^0 \urcorner \wedge z = \bar{t} \dot{\in} v_k^1) \to \mathrm{Sat}(z, v_3^1)\Big)$$

I abbreviate it to:

$$\forall x\big(x = \overline{v_{f(v_1^1)}^0 \dot{\in} v_{f(v_2^1)}^0} \dot{\in} v_{g(\Gamma)}^1 \to \mathrm{Sat}(x, v_3^1)\big)$$

$\mathcal{I}^\Gamma(\in)$ is in signed form (we assume x, y, z and t have superscript 0). The formula has only v_1^1, v_2^1 and v_3^1 free. All these variables have superscript 1, so they will not be inadvertently substituted when we construct $\mathcal{I}^\Gamma(\phi)$, where ϕ is in some Henkin extension, since Henkin extensions have only variables with superscript 0. Via the definition of substitutions, if we ignore the renaming of the bound variables, we obtain the following:

$$\mathcal{I}^\Gamma(v_n^0 \dot{\in} v_m^0) = \forall x(x = \overline{v^0_{f(v_n^0)} \dot{\in} v^0_{f(v_m^0)}} \dot{\in} v^1_{g(\Gamma)} \to \mathrm{Sat}(x, v_3^1))$$

Under $\mathfrak{c}$, v_3^1 is $\mathfrak{b}$ and v_n is n; under $\mathfrak{b}$, $v^1_{g(\Gamma)}$ is Γ, so the formula is saying that the formula that says that $\ulcorner v_n^0 \in v_m^0 \urcorner$ is in Γ is true. Since the formula talks about a formula of set-theory, axiom SM7 applies, so if it is true that such a formula is true, then the formula is true. The equivalence between $\mathcal{I}^\Gamma$ and Γ is proven below (every line is equivalent to the next):

$$\begin{array}{ll}
\mathrm{Sat}(\mathcal{I}^\Gamma(v_n^0 \dot{\in} v_m^0), \mathfrak{c}) & \\
\forall x(x = \overline{v^0_{f(\mathfrak{c}(v_n^0))} \dot{\in} v^0_{f(\mathfrak{c}(v_m^0))}} \dot{\in} v^1_{g(\Gamma)} \to \mathrm{Sat}(\mathrm{Sat}(v, v_3^1), \mathfrak{c}_v^x)) & \text{lm.1, df. } \mathcal{I}^\Gamma \\
\forall x(x = \overline{v_n^0 \dot{\in} v_m^0} \dot{\in} v^1_{g(\Gamma)} \to \mathrm{Sat}(\mathrm{Sat}(v, v_3^1), \mathfrak{c}_v^x)) & \text{df. } f, \mathfrak{c} \\
\mathrm{Sat}(\overline{v_n^0 \in v_m^0} \dot{\in} v^1_{g(\Gamma)}, \mathfrak{c}(v_3^1)) & \text{SM7} \\
\ulcorner v_n^0 \in v_m^0 \urcorner \in \mathfrak{b}(v^1_{g(\Gamma)}) & \text{df. } \mathfrak{c}, \text{lm.1} \\
\ulcorner v_n^0 \in v_m^0 \urcorner \in \Gamma & \text{def. } \mathfrak{b}
\end{array}$$

This takes care of the induction base. We now proceed to the induction step. The cases for negation and conjunction are trivial. The case for the quantifier is handled as in Halbach 2020, 325. For the modal case, first we need to define the set ξ of modally relevant $\mathcal{I}$ that is canonical for K. The canonical accessibility relation C between Henkin extensions is defined as follows:

$$C(x, y) \coloneqq \forall \phi(\dot{\Box}\phi \in x \to \phi \in y)$$

We now need to define ξ via C, so that, when we apply a modally relevant substitution to a $\mathcal{I}^\Gamma(\phi)$, we 'end up' in a formula that is equivalent to some $\mathcal{I}^\Omega(\phi)$ such that $C(\Gamma, \Omega)$ (and, vice-versa, whenever $C(\Gamma, \Omega)$, there is a modally relevant substitution that makes us jump from $\mathcal{I}^\Gamma$ to $\mathcal{I}^\Omega$).[19] Here are the details:

[19] A little informal explanation of how I came up with this definition. If we could switch free variables, this would be easy, but substitutions cannot switch free variables. So, instead, I chose to 'double quote' the formula, and then change the clause for $\mathrm{Sat}(v, u)$, so that instead of v being the formula that says that x is in Γ, it is the formula that says that x is in Ω. I could have added a name

Definition 5 (Canonical condition)**.** *$\mathcal{I} \in \xi$ if and only if it satisfies the following condition: there is a Γ and Ω such that $C(\Gamma, \Omega)$ and, for all ϕ_{at}, $\mathcal{I}$ is as follows:*

$$\mathcal{I}(\phi_{at}) = \begin{cases} \phi_{at} & \text{if } \phi_{at} \neq \ulcorner \mathrm{Sat}(v, u) \urcorner \\ \ulcorner \mathrm{Sat}(|v, v^1_{g(\Omega)}/v^1_{g(\Gamma)}|, u) \urcorner & \text{otherwise} \end{cases}$$

We make use of the following observation:

Lemma 14 (Composition lemma)**.** *If a substitution function $\mathcal{H}$ is in ξ, then there are some Γ and Ω such that $C(\Gamma, \Omega)$ and, for any ϕ, truth under $\mathcal{HI}^\Gamma$ is equivalent to truth under $\mathcal{I}^\Omega$. That is:*

$$\forall \Gamma, \Omega \Big(C(\Gamma, \Omega) \leftrightarrow \exists \mathcal{H} \big(\mathcal{H} \in \xi \wedge \forall \phi, \mathfrak{a} (\mathrm{Sat}(\mathcal{HI}^\Gamma(\phi), \mathfrak{a}) \leftrightarrow \mathrm{Sat}(\mathcal{I}^\Omega(\phi), \mathfrak{a}))\big)\Big)$$

Proof. The right to left direction follows from the definition of ξ. For the other direction, by construction of ξ, the relevant $\mathcal{H}$ is clearly the one that swaps $\dot{\mathrm{Sat}}(v, u)$ with $\dot{\mathrm{Sat}}(|v, v^1_{g(\Omega)}/v^1_{g(\Gamma)}|, u)$. That truth under $\mathcal{HI}^\Gamma$ is equivalent to truth under $\mathcal{I}^\Omega$ is shown by induction on the complexity of ϕ. We prove the base, using $v^r_n \dot{\in} v^s_m$ as an example. First, in $\mathcal{I}^\Gamma(\in)$ there is only one occurrence of Sat. So, the following holds:

$$\mathcal{HI}^\Gamma(v^r_n \dot{\in} v^s_m) = \forall x (x = \overline{v^0_{f(v^r_n)} \dot{\in} v^0_{f(v^s_m)}} \dot{\in} v^1_{g(\Gamma)} \rightarrow \mathrm{Sat}(|x, v^1_{g(\Omega)}/v^1_{g(\Gamma)}|, v^1_3))$$

The proof continues as follows (every line is equivalent to the next):

$$\begin{array}{l}
\mathrm{Sat}(\mathcal{HI}^\Gamma(v^r_n \dot{\in} v^s_m), \mathfrak{a}) \\
\mathrm{Sat}(\dot{\forall} x (x \dot{=} \overline{v^0_{f(v^r_n)} \dot{\in} v^0_{f(v^s_m)}} \dot{\in} v^1_{g(\Gamma)} \dot{\rightarrow} \dot{\mathrm{Sat}}(|x, v^1_{g(\Omega)}/v^1_{g(\Gamma)}|, v^1_3)), \mathfrak{a}) \\
\forall x (x = \overline{v^0_{f(\mathfrak{a}(v^r_n))} \dot{\in} v^0_{f(\mathfrak{a}(v^s_m))}} \dot{\in} v^1_{g(\Gamma)} \rightarrow \mathrm{Sat}(\dot{\mathrm{Sat}}(|v, v^1_{g(\Omega)}/v^1_{g(\Gamma)}|, v^1_3), \mathfrak{a}^x_v)) \\
\forall x (x = \overline{v^0_{f(\mathfrak{a}(v^r_n))} \dot{\in} v^0_{f(\mathfrak{a}(v^s_m))}} \dot{\in} v^1_{g(\Gamma)} \rightarrow \mathrm{Sat}(|x, v^1_{g(\Omega)}/v^1_{g(\Gamma)}|, \mathfrak{a}(v^1_3))) \\
\ulcorner v^0_{f(\mathfrak{a}(v^r_n))} \in v^0_{f(\mathfrak{a}(v^s_m))} \urcorner \in \mathfrak{a}(v^1_3)(v^1_{g(\Omega)}) \\
\forall x (x = \overline{v^0_{f(\mathfrak{a}(v^r_n))} \dot{\in} v^0_{f(\mathfrak{a}(v^s_m))}} \dot{\in} v^1_{g(\Omega)} \rightarrow \mathrm{Sat}(x, \mathfrak{a}(v^1_3))) \\
\mathrm{Sat}(\mathcal{I}^\Omega(v^r_n \dot{\in} v^s_m), \mathfrak{a})
\end{array}$$

In the step from line 3 to line 4 we are applying axiom SM7, since the formula belongs to L_{ZFC}. The step from line 4 to 5 is proven as follows: $\mathfrak{a}(v^1_3)$ must be an assignment; if it was not then $\forall x \neg \mathrm{Sat}(x, \mathfrak{a}(v^1_3))$ via SM8, in particular when

for every Henkin extension, and then let the substitutions swap names, yet I do not like this solution for two reasons: first, it makes the system rely on syntactic resources too much, which is one of the main criticisms against the substitutional account (Etchemendy 1990, Hinman et al. 1968). Second, if we wished to take identity as a logical expression, we would probably want $a = b \rightarrow \Box a = b$ to be logically true, which it would not be if names can be swapped.

$x = \overline{v^0_{f(\mathfrak{a}(v^r_n))} \dot{\in} v^0_{f(\mathfrak{a}(v^s_m))}} \dot{\in} v^1_{g(\Gamma)}$, so line 4 would not hold, contrary to our assumption. Since $\mathfrak{a}(v^1_3)$ is an assignment, and x is a formula of set theory, we can apply lemma 1, and proceed to line 5.

The induction step follows by definition of $\mathcal{I}$ and the compositional axioms of Sat. □

The lemma shows that truth under $\mathcal{H}\mathcal{I}^\Gamma(\phi)$ is equivalent to truth under $\mathcal{I}^\Omega(\phi)$. For this reason, we call $\mathcal{H}$ the composition from $\mathcal{I}^\Omega$ to $\mathcal{I}^\Gamma$, written: $\mathcal{H}^{\mathcal{I}^\Gamma \mapsto \mathcal{I}^\Omega}$.

We are now ready to prove in the usual way the induction step of the substitutional truth lemma for the modal case. First, suppose $\dot{\Box}\phi \in \Gamma$ and yet $\neg\mathrm{Sat}(\mathcal{I}^\Gamma(\dot{\Box}\phi), \mathfrak{c})$. Via commutation of $\mathcal{I}$ with $\Box$, $\neg\mathrm{Sat}(\dot{\Box}\mathcal{I}^\Gamma(\phi), \mathfrak{a})$. Via the axioms for truth, $\exists\mathcal{H}(\mathcal{H} \in \xi \wedge \neg\mathrm{Sat}(\mathcal{H}\mathcal{I}^\Gamma(\phi), \mathfrak{c}))$. Via construction of ξ, $\exists\mathcal{H}^{\mathcal{I}^\Gamma \mapsto \mathcal{I}^\Omega}(C(\Gamma, \Omega) \wedge \neg\mathrm{Sat}(\mathcal{H}^{\mathcal{I}^\Gamma \mapsto \mathcal{I}^\Omega}\mathcal{I}^\Gamma(\phi), \mathfrak{c}))$. Via the composition lemma, $\neg\mathrm{Sat}(\mathcal{I}^\Omega(\phi), \mathfrak{c})$. By induction hypothesis, $\phi \notin \Omega$. By construction of C, since $\dot{\Box}\phi \in \Gamma$, $\phi \in \Omega$. Contradiction.

For the other direction, assume that $\dot{\Box}\phi \notin \Gamma$ and yet $\mathrm{Sat}(\mathcal{I}^\Gamma(\dot{\Box}\phi), \mathfrak{c})$. We claim that the set $S = \{\psi : \dot{\Box}\psi \in \Gamma\} \cup \{\dot{\neg}\phi\}$ is consistent. Suppose it was not: then a finite subset $\psi_1, .., \psi_n$ in Γ is such that $(\psi_1 \wedge .. \wedge \psi_n) \to \bot$. By logic, $(\psi_1 \wedge .. \wedge \psi_n) \to \phi$. But Γ is closed under K rules, so by necessitation and K: $(\Box\psi_1 \wedge .. \wedge \Box\psi_n) \to \Box\phi$. Thus, $\dot{\Box}\phi \in \Gamma$, contrary to our assumption. Since S is consistent, by Lindenbaum's lemma we can extend it to a maximal and consistent set, to which we can add the Henkin Witness property via the renaming of variables explained above. We obtain a Henkin extension Ω where, by construction, $C(\Gamma, \Omega)$, and $\phi \notin \Omega$. We construct the substitution that maps $\mathrm{Sat}(v, u)$ to $\mathrm{Sat}(|v, v^1_{g(\Omega)}/v^1_{g(\Gamma)}|, u)$, and reads the other atomic formulae homophonically. By construction of ξ, the substitution is in ξ: it is $\mathcal{H}^{\mathcal{I}^\Gamma \mapsto \mathcal{I}^\Omega}$. By assumption, $\mathrm{Sat}(\dot{\Box}\mathcal{I}^\Gamma(\phi), \mathfrak{c})$, so $\mathrm{Sat}(\mathcal{I}\mathcal{I}^\Gamma(\phi), \mathfrak{c})$, for any $\mathcal{I} \in \xi$, in particular $\mathrm{Sat}(\mathcal{H}^{\mathcal{I}^\Gamma \mapsto \mathcal{I}^\Omega}\mathcal{I}^\Gamma(\phi), \mathfrak{c})$. Via the composition lemma $\mathrm{Sat}(\mathcal{I}^\Omega(\phi), \mathfrak{c})$, and $\phi \in \Omega$, by induction hypothesis, contrary to our assumption. This concludes the induction and the proof of the substitutional truth-lemma.

Theorem 5 (K-adequacy)**.** *There is a canonical truth-theory for K. That is, a truth-theory* $\mathsf{SM} + \mathsf{SMi}$ *such that:*

$$\mathsf{SM} + \mathsf{SMi} \vdash \forall\phi(K \vdash \phi \leftrightarrow \vDash_S \phi)$$

Proof. The theory in question is of course $\mathsf{SM} + \mathsf{SM}\xi$. Theorem 2 ensures soundness. For the right to left direction, suppose that ϕ is not provable in K. Then $\dot{\neg}\phi$ is in some Henkin extension of K. Via lemma 13, $\neg\mathrm{Sat}(\mathcal{I}^\Gamma(\phi), \mathfrak{c})$, and thus $\nvDash_S \phi$. □

The proofs used in this section can be generalised to the usual, familiar canonical modal logics like S4, S5, B etc. By construction, ξ will ensure the relevant features for soundness (that is, path-reflexivity, path-transitivity etc). We show this for T.

The canonical condition for T is reflexive, so it follows that there is, for any Γ, a $\mathcal{H}$ in ξ such that $\mathcal{H}(\mathrm{Sat}(v,u) = \mathrm{Sat}(|v, v^1_{g(\Gamma)}/v^1_{g(\Gamma)}, u)$, leaving the other atomic formulae the same. We show that, for any ϕ, $\phi \approx \mathcal{H}(\phi)$. The base is trivial for $v \in u$ and $v = u$. For $\mathrm{Sat}(v,u)$, we reason as follows. $|x, y/y| = x$ and $\mathrm{Sat}(|x, y/y'|, z)$ is actually a short for $\forall t(t = |x, y/y'| \rightarrow \mathrm{Sat}(t, z))$.

$$\begin{aligned}
\mathrm{Sat}(\mathrm{Sat}(v,u), \mathfrak{a}) &\leftrightarrow \forall x(x = \mathfrak{a}(v) \rightarrow \mathrm{Sat}(\mathrm{Sat}(v_n, u), \mathfrak{a}^x_{v_n})) \\
&\leftrightarrow \forall x(x = |\mathfrak{a}(v), v^1_{g(\Gamma)}/v^1_{g(\Gamma)}| \rightarrow \mathrm{Sat}(\mathrm{Sat}(v_n, u), \mathfrak{a}^x_{v_n})) \\
&\leftrightarrow \mathrm{Sat}(\forall v_n(v_n \doteq |v, v^1_{g(\Gamma)}/v^1_{g(\Gamma)}| \dot{\rightarrow} \mathrm{Sat}(v_n, u)), \mathfrak{a}) \\
&\leftrightarrow \mathrm{Sat}(\mathcal{I}^\Gamma(\mathrm{Sat}(v,u)), \mathfrak{a})
\end{aligned}$$

The induction step follows because $\mathcal{I}$ always reads connectives and quantifiers homophonically. Thus, since $\mathcal{H}(\phi) \approx \phi$, there is always a path from ϕ to ϕ, and the system is path-reflexive.

The adequacy proof cannot be extended to modal logics for which there is no canonical model, even when such logics can be shown to be Kripke complete via aternative methods. For example, the logic of provability GL is K4 plus Löb's axiom $\Box(\Box\phi \rightarrow \phi) \rightarrow \phi$. GL is not canonical (and not strongly complete) but admits selective filtration, thus it has the finite model property, being characterised by the class of finite strict partial orders (Chagrov and Zakharyaschev 1997, 150).[20] Filtrations do not work in the substitutional framework, because each truth-theory and choice of i can copy at most one filtration, for some specific ϕ. Of course, if we wanted to we could change the definition of validity, by quantifying over different truth-theories, which differ in the choice of i. An argument would be valid iff substitutionally valid in all truth-theories. Then, we could pick as relevant all and only the modal theories that correspond to each filtration, and adequacy would follow. However, this choice would betray the substitutional understanding of validity: validity is preservation of truth – simple truth – under all substitutions. It is not preservation of different models of truth under all substitutions. The correct philosophical conclusion is simply that substitutional semantics is not as strong as Kripke semantics.

[20] For some other examples, see Chagrov and Zakharyaschev 1997, 5.5.

3.6 Quantification

3.6.1 Barcan and converse

In the semantic systems we have been working on so far, the following formulae are valid:

BF $\forall x \Box \phi \rightarrow \Box \forall x \phi$

CBF $\Box \forall x \phi \rightarrow \forall x \Box \phi$

These are the famous Barcan Formula and its converse. BF is equivalent to $\Diamond \exists x(\phi) \rightarrow \exists x(\Diamond \phi)$; CBF to $\exists x(\Diamond \phi) \rightarrow \Diamond \exists x(\phi)$. They are valid in the semantics discussed so far, because $\mathcal{I}$ totally commutes with '$\forall$' (and with '$\exists$', because it also commutes with negation). In Kripke constant domain semantics, this is the case, as well. In variable domain semantics it is not. Barcan is falsified if the domain can expand, by going from world to world, following R, and the converse of Barcan is falsified by shrinking models. Constant domains are not the only way for Barcan and converse to be true: CBF is valid in every Kripke frame in which $R(w, v)$ implies $D_w \subseteq D_v$, where D_w is the domain of things that exist in w. BF is valid only in frames in which $R(w, v)$ implies $D_v \subseteq D_w$. If B is valid then one can always 'go back' to the original world, and the two requirements collapse into one. Thus, one can prove that the validity of CBF plus B implies constancy: $R(w, v) \rightarrow D_w = D_v$. Similarly, the validity of BF plus B implies the validity of CBF.[21]

In a substitutional theory of validity we can falsify BF and CBF by restricting the quantifiers to some condition. We change the definition of $\mathcal{I}$, by extending the clause for quantified formulae to functions that satisfy the following recursive clause, where δ is some designated formula:

$$\mathcal{I}(\dot{\forall} v \phi) \coloneqq \dot{\forall} u(|\delta, u/\alpha_1| \dot{\rightarrow} |\mathcal{I}(\phi), u/v|)$$

We call these substitution functions 'relativised', while the others are called 'unrelativised'. We assume that δ is in signed form, and the usual care needs to be applied to avoid unintended variable bindings. Restricting quantifiers to a condition δ is optional, so the new system is a proper extension of the old one: we are considering both unrelativised and relativised substitutions. As expected in variable domain semantics for modal logics, we obtain a free and 'inclusive' logic: free because existential generalisation and universal instantiation might fail, since a term might denote something

[21] For a more detailed discussion, see Cresswell and Hughes 1996.

outside of the relevant condition for the quantifiers; inclusive because $\dot{\exists} x \top$ is not a logical truth, as the condition δ might be paradoxical.[22]

In the new semantics we can falsify BF and CBF. For simplicity, we focus on truth theories whose sets of modally relevant $\mathcal{I}$ are unrelativised: they do not add any conditions to the quantifiers. In the system, BF is valid in $\mathsf{SM}+\mathsf{SMi}$ iff true under all assignments and substitutions. A substitution $\mathcal{I}$ can now swap any instance of $\dot{\forall} v \dot{\Box} \phi \dot{\rightarrow} \dot{\Box} \dot{\forall} v \phi$ with $\dot{\forall} v (\delta \dot{\rightarrow} \dot{\Box} \psi) \dot{\rightarrow} \dot{\Box} \dot{\forall} v (\delta \dot{\rightarrow} \psi)$, where $\psi = \mathcal{I}(\phi)$. Given the axioms for truth, the formula is satisfied under $\mathfrak{a}$ when $\mathrm{Sat}(\dot{\forall} v (\delta \dot{\rightarrow} \dot{\Box} \psi), \mathfrak{a}) \rightarrow \mathrm{Sat}(\dot{\Box} \dot{\forall} v (\delta \dot{\rightarrow} \psi), \mathfrak{a})$. We unravel the two formulae side by side. We ignore variable renaming for readability; $\mathcal{H}$ are substitutions in i:

$$
\begin{array}{ll}
\mathrm{Sat}(\dot{\forall} v (\delta \dot{\rightarrow} \dot{\Box} \psi), \mathfrak{a}) & \mathrm{Sat}(\dot{\Box} \dot{\forall} v (\delta \dot{\rightarrow} \psi), \mathfrak{a}) \\
\forall x (\mathrm{Sat}(\delta, \mathfrak{a}^x_v) \rightarrow \mathrm{Sat}(\dot{\Box} \psi, \mathfrak{a}^x_v)) & \forall \mathcal{H} \mathrm{Sat}(\mathcal{H}(\dot{\forall} v (\delta \dot{\rightarrow} \psi), \mathfrak{a})) \\
\forall x (\mathrm{Sat}(\delta, \mathfrak{a}^x_v) \rightarrow \forall \mathcal{H} \mathrm{Sat}(\mathcal{H}(\psi), \mathfrak{a}^x_v)) & \forall \mathcal{H} \mathrm{Sat}(\dot{\forall} v (\mathcal{H}(\delta) \dot{\rightarrow} \mathcal{H}(\psi)), \mathfrak{a}) \\
\forall \mathcal{H}, x (\mathrm{Sat}(\delta, \mathfrak{a}^x_v) \rightarrow \mathrm{Sat}(\mathcal{H}(\psi), \mathfrak{a}^x_v)) & \\
\forall \mathcal{H}, x \, \mathrm{Sat}(\delta \dot{\rightarrow} \mathcal{H}(\psi), \mathfrak{a}^x_v) & \\
\forall \mathcal{H} \mathrm{Sat}(\dot{\forall} v (\delta \dot{\rightarrow} \mathcal{H}(\psi)), \mathfrak{a}) &
\end{array}
$$

We end up with the following conditional:

$$\forall \mathcal{H} \mathrm{Sat}(\dot{\forall} v (\delta \dot{\rightarrow} \mathcal{H}(\psi)), \mathfrak{a}) \rightarrow \forall \mathcal{H} \mathrm{Sat}(\dot{\forall} v (\mathcal{H}(\delta) \dot{\rightarrow} \mathcal{H}(\psi)), \mathfrak{a})$$

We can easily falsify this formula, for example by setting $\psi = \bot$, $\delta = v \dot{\in} v$ and considering a truth-theory where all $\mathcal{I}$ in i are such that $\mathcal{I}(v \dot{\in} v) = \dot{\neg} v \dot{\in} v$, for any v. Then $\forall x (x \in x \rightarrow \bot)$ is true because $x \notin x$, but $\forall x (\neg x \in x \rightarrow \bot)$ is false. Similarly, there are many ways to falsify CBF. If CBF is valid, then the converse of the conditional above holds. To falsify the condition, just keep the same example as above except now $\delta = \dot{\neg} v \dot{\in} v$. We will see below in which systems with relativised substitutions are BF and CBF valid. Admittedly, the details are not pretty.

In a substitutional theory, even though BF is not valid, it is definitely true under the homophonic reading, regardless of the $\mathsf{SM}+\mathsf{SMi}$ we are considering. For what it means is something obvious: that if everything satisfies all the relevant substitutions of a condition P, then 'Everything satisfies P' is satisfied under all relevant substitutions. This is almost trivial, given the clause for satisfaction of quantified formulae,

[22]For a discussion on free logic, see Bencivenga 1986. For a discussion on quantification in the substitutional theory of validity see Halbach 2018.

and the homophonic reading of the quantifiers. In fact, the truth of BF in its intended reading, given axiom $\mathsf{SM} + \mathsf{SMi}$, is almost a paraphrase of the compositional axiom for the truth of quantified formulae. We should welcome this result. The intended reading of '$\forall$' in the present semantics is meant to be absolutely general. Since quantifiers in their intended reading are unrestricted, all a substitution function can do is to restrict them to some condition. So, if absolutely everything satisfies some condition, how can there be something else we overlooked that does not? It is no surprise that BF comes out true. Even if quantifiers are not absolutely general, BF is still true under the reductive reading of $\Box$ in the present system. The most general quantification we can achieve inside any meta-theory is that of the meta-theory, of course. Otherwise, I could prove in the meta-theory that there is something we cannot quantify over in the meta-theory, which is contradictory.

3.6.2 Adequacy in a variable domain semantics

In this section I provide adequacy theorems for modal logics with BF, CBF or neither, based on K. I call FK the result of adding axiom K and the necessitation rule to a free logic F. I call $\mathsf{FK} + \mathsf{BF}$ the result of adding BF to FK. I call $\mathsf{FK} + \mathsf{CBF}$ the result of adding CBF to FK.

3.6.3 Adequacy for calculi without BF or CBF

I assume we have defined a Kripke model $\mathfrak{M} = \langle W, D, R, V, Q\rangle$ for the language of $L_\Box$ in the usual way, where W is the set of worlds, D the domain, R the accessibility relation, V the valuation function and Q a function from W to the power set of D. I assume we defined inductively a concept of truth in a world w of $\mathfrak{M}$, in set theory, in the usual way. $\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \phi$ means that ϕ is true in a Kripke model $\mathfrak{M}$ at w under $\mathfrak{a}$.

$$\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \dot{\mathrm{S}}\mathrm{at}(v, u) \leftrightarrow \langle \mathfrak{a}(v), \mathfrak{a}(u), w\rangle \in V(\mathrm{Sat})$$
$$\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash v \dot{\in} u \leftrightarrow \langle \mathfrak{a}(v), \mathfrak{a}(u), w\rangle \in V(\in)$$
$$\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash v \dot{=} u \leftrightarrow \langle \mathfrak{a}(v), \mathfrak{a}(u), w\rangle \in V(=)$$
$$\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \dot{\neg}\phi \leftrightarrow \neg\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \phi$$
$$\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \psi \dot{\wedge} \phi \leftrightarrow \langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \psi \wedge \langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \phi$$
$$\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \dot{\forall} v\phi \leftrightarrow \forall y(y \in Q(w) \rightarrow \langle \mathfrak{M}, w, \mathfrak{a}^y_v\rangle \vDash \phi)$$
$$\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \dot{\Box}\phi \leftrightarrow \forall v(R(w, v) \rightarrow \langle \mathfrak{M}, v, \mathfrak{a}\rangle \vDash \phi)$$

We say that ϕ is true in $\mathfrak{M}$ (written '$\mathfrak{M} \vDash \phi$') iff $\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \phi$ for all $\mathfrak{a}$ and w of $\mathfrak{M}$. We say that ϕ is valid in a frame iff true in all $\mathfrak{M}$.

In the following we again help ourselves with superscripts 1 and 0 to make the notation uniform across the article. We define π as the result of getting rid of superscript 1:

$$\begin{aligned}\pi(v^0_n) &= v^0_{2n+1}\\ \pi(v^1_n) &= v^0_{2n}\\ \pi(\phi) &= \text{the result of uniformly applying } \pi \text{ to each variable in } \phi.\end{aligned}$$

The following lemma identifies a canonical truth-theory, and ensures completeness of FK:

Lemma 15. *For any countable Kripke model $\mathfrak{M}$ of* FK *there is a truth-theory* SM + SMξ *where the following holds, where w are worlds of $\mathfrak{M}$ and $\mathcal{H}$ substitutions in ξ:*

$$\forall \mathfrak{a}\exists \mathfrak{c}\forall w\exists \mathcal{H}\forall \phi\big(\langle \mathfrak{M}, w, \mathfrak{a}\rangle \vDash \pi(\phi) \leftrightarrow \mathrm{Sat}(\mathcal{H}\pi(\phi), \mathfrak{c})\big)$$

The rest of the section is the proof of lemma 15, and the adequacy theorem that follows from it.

For any ϕ, $\pi(\phi)$ has only variables with superscript 0. Consider any countable Kripke model $\mathfrak{M} = \langle W, D, R, V, Q\rangle$. $\mathfrak{M}$ is countable (i.e. W is countable), so we can match the elements of W with the natural numbers. We order the elements of the set $W \cup \{V, Q\}$. $g(x) = n$ iff x is the x-th thing in the order. We make use of the following assignments:

$$\begin{aligned}\mathfrak{c}(v^1_3) &= \mathfrak{b} & \qquad \mathfrak{b}(v^1_{g(x)}) &= x\\ \mathfrak{c}(v^0_n) &= n & \qquad \mathfrak{b}(v^0_n) &= \mathfrak{a}(v^0_n)\end{aligned}$$

The values not specified do not matter and can be set to 0. Given a w of $\mathfrak{M}$, we build its characteristic $\mathcal{I}^w$ inductively from the substitution for atomic formulae and the condition for the quantifiers. We consider $\ulcorner v_n \in v_m \urcorner$ as an example for atomic formulae. We use the same recursive function f as above. $\mathcal{I}^w(\in)$ is the following formula, where $g(w) = k$ and $g(V) = r$

$$\forall x, y, z\big((x = f(v^1_1) \wedge y = f(v^1_2) \wedge z = \overline{\langle v^0_x, v^0_y, v^1_k\rangle \in v^1_r(\in)}) \rightarrow \mathrm{Sat}(z, v^1_3)\big)$$

The formula has only v^1_1, v^1_2 and v^1_3 free. It does not matter if – say – $g(w)$ is 1 or 2 or 3, since technically $v^1_{g(w)}$ is not occurring in the formula, but rather what occurs is the numeral of (the code of) a formula where that variable occurs. We abbreviate the formula to:

$$\forall x\big(x = \overline{\langle v^0_{f(v^1_1)}, v^0_{f(v^1_2)}, v^1_{g(w)}\rangle \dot{\in} v^1_{g(V)}(\in)} \rightarrow \mathrm{Sat}(x, v^1_3)\big)$$

The designated condition δ of $\mathcal{I}^w$ for quantifiers is the following, where $g(w) = k$ and $g(Q) = s$:

$$\forall x, y((x = f(v_1^1) \wedge y = \overline{v_x^0 \dot{\in} v_s^1(v_k^1)}) \to \mathrm{Sat}(y, v_3^1))$$

I abbreviate it to:

$$\forall x(x = \overline{v^0_{f(v_1^1)} \dot{\in} v^1_{g(Q)}(v^1_{g(w)})} \to \mathrm{Sat}(x, v_3^1))$$

I now prove the base of the lemma, using $v_n^0 \in v_m^0$ as an example (every line is equivalent to the next):

$$\begin{aligned}
&\langle \mathfrak{M}, w, \mathfrak{a} \rangle \vDash v_n^0 \dot{\in} v_m^0 \\
&\langle \mathfrak{a}(v_n^0), \mathfrak{a}(v_m^0), w \rangle \in V(\in) \\
&\langle \mathfrak{b}(v_n^0), \mathfrak{b}(v_m^0), \mathfrak{b}(v^1_{g(w)}) \rangle \in \mathfrak{b}(v^1_{g(V)})(\in) \\
&\mathrm{Sat}(\langle v_n^0, v_m^0, v^1_{g(w)} \rangle \dot{\in} v^1_{g(V)}(\in), \mathfrak{b}) \\
&\mathrm{Sat}(\langle v^0_{f(\mathfrak{c}(v_n^0))}, v^0_{f(\mathfrak{c}(v_m^0))}, v^1_{g(w)} \rangle \dot{\in} v^1_{g(V)}(\in), \mathfrak{c}(v_3^1)) \\
&\forall x(x = \overline{\langle v^0_{f(\mathfrak{c}(v_n^0))}, v^0_{f(\mathfrak{c}(v_m^0))}, v^1_{g(w)} \rangle \dot{\in} v^1_{g(V)}(\in)} \to \mathrm{Sat}(\dot{\mathrm{S}}\mathrm{at}(v, v_3^1), \mathfrak{c}_v^x)) \\
&\mathrm{Sat}(\mathcal{I}^w(v_n^0 \dot{\in} v_m^0), \mathfrak{c})
\end{aligned}$$

For the induction step of lemma 15, the axioms for satisfaction trivially take care of negation and conjunction. The quantified case is handled as follows. First, we prove the following: $\mathfrak{a}(v_n^0) \in Q(w) \leftrightarrow \mathrm{Sat}(\delta(v_n^0), \mathfrak{c})$:

$$\begin{aligned}
&\mathrm{Sat}(\delta(v_n^0), \mathfrak{c}) \\
&\mathrm{Sat}(\dot{\forall} v(v \dot{=} \overline{v^0_{f(v_n^0)} \dot{\in} v^1_{g(Q)}(v^1_{g(w)})} \dot{\to} \dot{\mathrm{S}}\mathrm{at}(v, v_3^1)), \mathfrak{c}) \\
&\forall x(x = \overline{v_n^0 \dot{\in} v^1_{g(Q)}(v^1_{g(w)})} \to \mathrm{Sat}(\dot{\mathrm{S}}\mathrm{at}(v, v_3^1), \mathfrak{c}_v^x)) \\
&\mathrm{Sat}(v_n^0 \dot{\in} v^1_{g(Q)}(v^1_{g(w)}), \mathfrak{b}) \\
&\mathfrak{a}(v_n^0) \in Q(w)
\end{aligned}$$

Then, the proof continues as follows:

$$\begin{aligned}
&\langle \mathfrak{M}, w, \mathfrak{a} \rangle \vDash \dot{\forall} v_n^0 \phi \\
&\forall y(y \in Q(w) \to \langle \mathfrak{M}, w, \mathfrak{a}^y_{v_n^0} \rangle \vDash \phi) \\
&\forall y(\mathfrak{a}^y_{v_n^0}(v_n^0) \in Q(w) \to \langle \mathfrak{M}, w, \mathfrak{a}^y_{v_n^0} \rangle \vDash \phi) \\
&\forall y(\mathrm{Sat}(\delta(v_n^0), \mathfrak{c}^y_{v_n^0}) \to \mathrm{Sat}(\mathcal{I}^w(\phi), \mathfrak{c}^y_{v_n^0})) \\
&\mathrm{Sat}(\forall u|\delta, u/v_n^0| \to |\phi, u/v_n^0|, \mathfrak{c}) \\
&\mathrm{Sat}(\mathcal{I}^w(\forall v_n^0 \phi), \mathfrak{c})
\end{aligned}$$

For the modal case, we set ξ exactly as above, if not that now we focus on the R of the Kripke model at hand, and not on a relation defined on Henkin extensions of the language:

Definition 6 (ξ for $\mathfrak{M}$)**.** *Given a countable $\mathfrak{M}$ and its R, the correspondent ξ is defined as follows: $\mathcal{I}$ is in ξ if and only there is a w and w' in W of $\mathfrak{M}$ such that $R(w, w')$ and, for all ϕ_{at}, $\mathcal{I}$ is as follows:*

$$\mathcal{I}(\phi_{at}) = \begin{cases} \phi_{at} & \text{if } \phi_{at} \neq \ulcorner \mathrm{Sat}(v, u) \urcorner \\ \ulcorner \mathrm{Sat}(|v, v^1_{g(w')}/v^1_{g(w)}|, u) \urcorner & \text{otherwise} \end{cases}$$

We stipulate that modal substitutions are unrelativised. By construction of ξ and via the definition of $\mathcal{I}^w$, it is easy to see that a result parallel to lemma 14 holds: any $\mathcal{H}$ is in ξ exactly when there are two worlds w, v in $\mathfrak{M}$, $R(w, v)$ and $\mathcal{H}$ is a composition from $\mathcal{I}^w$ to $\mathcal{I}^v$: i.e. for any ϕ, $\mathrm{Sat}(\mathcal{H}\mathcal{I}^w(\phi), \mathfrak{c})$ is equivalent to $\mathrm{Sat}(\mathcal{I}^v(\phi), \mathfrak{c})$. The only new part of the proof is for δ: the proof works in basically the same way as the atomic cases, by exploiting the fact that the only difference between δ of $\mathcal{I}^w$ and δ of $\mathcal{I}^v$ is the occurrence of $v^1_{g(v)}$ instead of $v^1_{g(w)}$ in the numeral of $v^1_{f(v^1_1)} \dot{\in} v^1_{g(v)}$.

The modal case for the induction for lemma 15 proceeds as follows: $\langle \mathfrak{M}, w, \mathfrak{a} \rangle \vDash \dot{\Box}\phi$ iff for all v such that $R(w, v)$, $\langle \mathfrak{M}, v, \mathfrak{a} \rangle \vDash \phi$. Via induction hypothesis, for all v such that $R(w, v)$, $\mathrm{Sat}(\mathcal{I}^v(\phi), \mathfrak{c})$. Via the result just explained, this is so iff $\forall \mathcal{H}(\mathcal{H} \in \xi \rightarrow \mathrm{Sat}(\mathcal{H}\mathcal{I}^w(\phi), \mathfrak{c}))$, iff $\mathrm{Sat}(\dot{\Box}\mathcal{I}^w(\phi), \mathfrak{c})$ iff $\mathrm{Sat}(\mathcal{I}^w(\Box\phi), \mathfrak{c})$. This concludes the induction and the proof of lemma 15.

Theorem 6. *There is a canonical truth-theory for FK.*

Proof. Pick the Kripke canonical model of FK. By theorem 4, there is an equivalent countable model: its 'countable filtration' $\mathfrak{M}$. A canonical truth-theory for FK is the characteristic truth-theory modelled from R of $\mathfrak{M}$ via definition 5. Assume $\mathsf{FK} \nvdash \phi$. Then, in some world w of the countable filtration, under some $\mathfrak{a}$, $\langle \mathfrak{M}, w, \mathfrak{a} \rangle \nvDash \phi$. Construct $\mathfrak{a}'$ as follows: $\mathfrak{a}'(\pi(v^s_n)) = \mathfrak{a}(v^s_n)$. Clearly, $\langle \mathfrak{M}, w, \mathfrak{a} \rangle \vDash \phi$ exactly when $\langle \mathfrak{M}, w, \mathfrak{a}' \rangle \vDash \pi(\phi)$, so $\langle \mathfrak{M}, w, \mathfrak{a}' \rangle \nvDash \pi(\phi)$. By lemma 15, $\neg\mathrm{Sat}(\mathcal{I}^w(\pi(\phi)), \mathfrak{c})$, therefore $\pi(\phi)$ is not substitutionally valid, and thus ϕ is not substitutionally valid either, since validity is preserved under substitutions of free variables.

For soundness, via the axioms for truth and the definition of validity we can show that free logic is sound for $\vDash_S$. Since lemma 8 and 9 still hold, we can prove soundness for FK by induction on the length of proofs. □

Theorem 6 can be extended to other familiar canonical extensions of FK without BF nor CBF. We can always apply lemma 15 to their canonical model. Soundness will still hold by construction of ξ, as before.

3.6.4 Adequacy for logic with BF or CBF

When it comes to modal logics with BF or CBF, things get more complicated. The condition δ I used before will not work in general, since its meaning depends on the variable assignment at hand. So whilst we might get completeness using the construction above, we lose soundness, because that same substitution falsifies both BF or CBF under other variable assignments. To ensure this does not happen, we need to 'surgically remove' any combination of variable assignments and substitutions that will falsify either BF or CBF. I provide the following lemma:

Lemma 16. *In any truth-theory* $\mathsf{SM} + \mathsf{SMi}$*, given a* $\mathcal{I}$ *and an* $\mathfrak{a}$*, BF is true if the following holds:*

NoExp $\forall \mathcal{H}(\mathcal{H} \in i \rightarrow \mathrm{Sat}(\dot{\forall} x(|(\delta^{\mathcal{H}} \dot{\wedge} \mathcal{H}(\delta^{\mathcal{I}})) \dot{\rightarrow} \delta^{\mathcal{I}}, x/\alpha_1|), \mathfrak{a}))$

CBF is true if the following holds:

NoShrink $\forall \mathcal{H}(\mathcal{H} \in i \rightarrow \mathrm{Sat}(\dot{\forall} x(|\delta^{\mathcal{I}} \dot{\rightarrow} (\delta^{\mathcal{H}} \dot{\wedge} \mathcal{H}(\delta^{\mathcal{I}})), x/\alpha_1|), \mathfrak{a}))$

Proof. I ignore variable renaming for readability. I show the lemma for BF, for CBF a very similar proof works. Similarly to how it was shown above, we can show here that, for any ϕ, where $\mathcal{I}(\phi) = \psi$ the satisfaction of $\forall v \Box \phi \rightarrow \Box \forall x \phi$ under $\mathcal{I}$ and $\mathfrak{a}$ is equivalent to the following conditional, where $\mathcal{H}$ are substitutions in i:

$$\forall \mathcal{H} \mathrm{Sat}(\dot{\forall} v(\delta^{\mathcal{I}} \dot{\rightarrow} \mathcal{H}(\psi)), \mathfrak{a}) \rightarrow \forall \mathcal{H} \mathrm{Sat}(\dot{\forall} v((\delta^{\mathcal{H}} \dot{\wedge} \mathcal{H}(\delta^{\mathcal{I}})) \dot{\rightarrow} \mathcal{H}(\psi)), \mathfrak{a})$$

Since NoExp holds, for any x $\mathrm{Sat}(\delta^{\mathcal{H}} \dot{\wedge} \mathcal{H}(\delta^{\mathcal{I}}), \mathfrak{a}_v^x)$ implies $\mathrm{Sat}(\delta^{\mathcal{I}}, \mathfrak{a}_v^x)$. Assume the conditional is false. Then (1) for all $\mathcal{H}$, $\forall v(\delta^{\mathcal{I}} \rightarrow \mathcal{H}(\psi))$ is true but (2) for some $\mathcal{H}$, $\forall v((\delta^{\mathcal{H}} \wedge \mathcal{H}(\delta^{\mathcal{I}})) \rightarrow \mathcal{H}(\psi))$ is not. From (2), for some $\mathcal{H}$ and x, $\delta^{\mathcal{H}} \wedge \mathcal{H}(\delta^{\mathcal{I}})$ is true under $\mathfrak{a}_v^x$ but $\mathcal{H}(\psi)$ is false, under $\mathfrak{a}_v^x$. From (1), since $\mathcal{H}(\psi)$ is false under $\mathfrak{a}_v^x$, it must be that $\delta^{\mathcal{I}}$ is also false under $\mathfrak{a}_x^v$. But $\delta^{\mathcal{H}} \wedge \mathcal{H}(\delta^{\mathcal{I}})$ is true under $\mathfrak{a}_v^x$, and by NoExp it implies that $\delta^{\mathcal{I}}$ is true under $\mathfrak{a}_v^x$. Contradiction.

□

Theorem 7. $\mathsf{FK} + \mathsf{BF}$ *is adequate for some truth-theory* $\mathsf{SM} + \mathsf{SMi}$*, where validity is restricted to tuples of* $\mathcal{I}$ *and* $\mathfrak{a}$ *that satisfy NoExp.* $\mathsf{K} + \mathsf{CBF}$ *is adequate for some truth-theory* $\mathsf{SM} + \mathsf{SMi}$ *where validity is restricted to tuple of* $\mathcal{I}$ *and* $\mathfrak{a}$ *that satisfy NoShrink.*

Proof. I use the construction of lemma 15. We have to check that each pair $\langle \mathcal{I}^w, \mathfrak{c} \rangle$ satisfies NoExp and is not excluded in the definition of validity. Since $\mathcal{H}$ are unrelativised, this amounts to $\forall \mathcal{H} \mathrm{Sat}(\dot{\forall} v(\mathcal{H}(\delta^{\mathcal{I}}) \dot{\rightarrow} \delta^{\mathcal{I}}), \mathfrak{c})$. I use the fact that any Kripke model for BF is such that, for any x, $x \in Q(v) \rightarrow x \in Q(w)$, if $R(w, v)$. As shown above, for any w, $\mathrm{Sat}(\delta^{\mathcal{I}^w}(v), \mathfrak{c})$ is equivalent to $\mathfrak{a}(v) \in Q(w)$, and, given any $\mathcal{H}$, $\mathcal{H}$ must be the composition from $\mathcal{I}^w$ to $\mathcal{I}^v$, where $R(w, v)$ for some w and v. $\mathrm{Sat}(\mathcal{H}(\delta^{\mathcal{I}^w}), \mathfrak{c})$ is equivalent to $\mathfrak{a}(v) \in Q(v)$, so:

$$\forall x(\mathfrak{a}^x_{v_n}(v_n) \in Q(v) \rightarrow \mathfrak{a}^x_{v_n}(v_n) \in Q(w))$$
$$\forall x(\mathrm{Sat}(\mathcal{H}(\delta^{\mathcal{I}^w}), \mathfrak{c}^x_{v_n}) \rightarrow \mathrm{Sat}(\delta^{\mathcal{I}^w}, \mathfrak{c}^x_{v_n}))$$
$$\mathrm{Sat}(\dot{\forall} v_n(\mathcal{H}(\delta^{\mathcal{I}^w}) \dot{\rightarrow} \delta^{\mathcal{I}^w}), \mathfrak{c})$$

So, any $\langle \mathcal{I}^w, \mathfrak{c} \rangle$ satisfies NoExp. A similar reasoning holds for CBF and NoShrink. □

As before, I can generalise the theorem to other familiar extensions of FK + BF and FK + CBF, by relying on path-conditions. However, I did have to change the definition of validity quite *ad hoc*, to achieve this result. A constant domain semantics is far more natural for a substitutional account. For then there is no need to explain why we are not treating quantifiers as fully logical, by forcing into the syntax a condition δ, nor do we need to explain why we are disregarding tuples of substitutions and assignments, which seem perfectly legitimate if not for making BF or CBF false. Of course, naturalness is not something we are interested in when we are trying to prove adequacy theorems for some arbitrary modal logic. Yet, such considerations are cogent if we step back and ask what is the most natural modal logic, given our initial assumptions: a substitutional account of validity, an understanding of modality as a way of expressing substitutional validity, and the belief that quantifiers are logical constants, to be read homophonically – given these assumptions, the answer is constant domain semantics.

Chapter 4

Should Theories of Validity Self-apply?

A theory of validity self-applies when it applies to its own language. A theory is general for a language if it is able to capture all the interpretations of that language. Recently, some philosophers argued that self-applicability is inconsistent with generality, and proposed a typed, higher order theory of validity. I defend the idea that self-applicability is often pivotal to validity. I argue that a popular Russellian objection against the compatibility of self-applicability and generality can be resisted. However, a Cantorian argument still stands. I conclude that the issue at hand needs to be re-evaluated: validity is likely a technical notion, which should be assessed against our theoretical needs. Sometimes we simply need validity to be self-applicable.

4.1 Self-applicability and generality

It is common to define logical validity in terms of truth and interpretations: a formula ϕ logically follows from a set of premises exactly when, under all suitable interpretations of the language, if all the premises are true so and so interpreted, so is ϕ. Albeit common, the 'interpretational' definition does have its critics.[1] In the following, I assume that preservation of truth under an interpretation is necessary for logical validity.

[1]Etchemendy (1990, 2008) notably argued against a purely interpretational account. To answer Etchemendy's worries, some have claimed that we should add a 'representational' element; validity is also preservation of truth in all alternative ways the world might be or might have been (Hanson 1997, Shapiro 1998). Still, most will not dispense with interpretations altogether.

A necessary task for analysing validity is considering the concept of interpretation and truth under an interpretation. The scope of what a theory of validity can cover is determined by the class of languages for which truth under an interpretation is defined. Sometimes a theory of validity is self-applicable: it is able to interpret its own language. More specifically, a theory T of validity in L is self-applicable when truth under an interpretation applies to all formulae of L.[2]

Self-applicability is common: the text-book theory of validity is model theory, which is usually spelled out in Zermelo-Fraenkel set-theory with choice (ZFC). It is a theory of validity for all first level languages, in particular for the language of set-theory. Some philosophers however argued against self-applicability. The main works that come to mind are Rayo and Uzquiano 1999, Rayo and Williamson 2003 and Rayo 2006, among others. I call them 'higher orderists'. Technically, these philosophers propose a typed theory of validity, where the Tarskian hierarchy of truth is generated by the orders of quantification. The meta-theory they use is usually a higher level language (Rayo and Uzquiano 1999) or a 'higher plural' language (Rayo 2006). I use the term 'order' to refer to the order of quantification, whether this is plural quantification or quantification in predicate position; I use 'order' to talk about theories, too. On the other hand, I use the term 'level' for the syntactic level of variables, formulae, languages etc. I talk only about languages of finite level, even though what I say can probably be extended to languages of infinite level.

In all such systems, when we interpret a language of level n, we need order $n+1$ resources. By order 'n+1' resources I mean that we need to implicitly or explicitly define a concept of order $n+1$. Implicitly, by introducing axiomatically a notion of satisfaction of a formula under an interpretation and assignment; explicitly, via comprehension at order $n+1$. The higher orderist's system is typed: satisfaction does not apply to formulae of level $n+1$, so it does not apply to formulae with the satisfaction predicate – satisfaction under an interpretation is always beyond what the system is able to interpret. The system therefore is not self-applicable. It is also not universally applicable, since there is no unique concept of validity that applies to languages of any level. Rather, validity for second level languages is a third order concept, validity for third level languages is a fourth order concept, and so on.

[2]By that I mean that the definition of truth under an interpretation is meant to be defined for all formulae of L – truth will commute over the connectives, etc. There are many ways of abusing this concept, but hopefully it is clear enough.

4.1.1 Standard set-based semantics is not general

Higher orderists take issue with standard model-theoretic characterisations of logical validity. They claim that it is not 'general' and therefore not conceptually adequate. They call 'general' a theory of validity for a language L which is able to capture all the acceptable interpretations of L. What counts as 'acceptable interpretation' does not look like something that can be formally defined, rather it is left it to our pre-formal intuitions. Higher orderists are happy to identify 'acceptable interpretation' with their definition of interpretation in higher order systems.

Generality seems like a good *desideratum* for a theory of validity: if validity is truth under all interpretations of the language, and in our theory we fail to quantify over some interpretations that we think are perfectly acceptable, then the theory is not conceptually adequate. Also, it is not clear whether we can trust the results of the theory, since the theory might miss some counterexamples to invalidities. Maybe some formula is true in all the interpretations the theory is able to capture, but false in some unreachable one. Reflection principles and squeezing arguments come to the rescue if we are dealing with first order logic (Kreisel 1967).Higher-order logic is incomplete under the standard semantics, and reflection principles are provable in second-order ZFC only for the first-order fragment of the language. To extend them to higher-order formulae they need to be added as new axioms (Shapiro 1987).[3]

That a theory is general is a philosophical thesis not a mathematical theorem, and often it is easier to show how a theory fails to be general, rather than how it is, in fact, general. Usually, one considers a theory A and a theory B, which is allegedly more general. From within theory B, we show that we are able to capture some interpretation that looks intuitively acceptable, but that we were not able to capture from the perspective of theory A. In this way, we *see* from the standpoint of theory B that theory A is not general.

Here is an example: we know we can build a theory of validity for first order logic in Peano Arithmetic (PA), by 'arithmetizing' model theory in a suitable way.[4] We also know that the resulting notion of validity is adequate: Hilbert and Bernays (1939) showed how first order logic is sound and complete for it. Yet, of course, from the standpoint of the usual set-based semantics, we can 'see' that there are plenty of interpretations of many first level languages which we are able to quantify over in set-based semantics but not in the arithmetic-based semantics. For example, there is

[3]See the article 'Higher Order Semantics and Non-standard Interpretations' in this volume for a discussion.

[4]For an overview of arithmetised model theory, see Hájek and Pudlák 2017.

no model of a first level language in the arithmetic-based semantics where the domain is the real numbers, or where the extension of a predicate is more than countable – all we can do is quantify over natural numbers, which are at most countable.[5]

Due to Tarskian reasons we cannot show that there is a model of the language of PA specifiable in PA that satisfies the Tarskian axioms for truth, i.e. the axioms in Tarski (1933): a T-schema for atomic sentences, compositional axioms for connectives and semi-compositional axioms for quantifiers. However, we we can construct such a model for the language of PA in set-based semantics. Finally, there are as many set-based interpretations of a first level language as there are sets, since any set can be the domain of an interpretation. However, there are at most countably many interpretations of a first level language specifiable in the arithmetic-based semantics, since an interpretation needs to be a natural number.

From the standpoint of set-theory one can 'see' that an arithmetic theory of logical validity is not general for the language of arithmetic, as well as for any other first level language. Similarly, higher orderists argue that the standard set-based semantics is not general, when judged from the perspective of a higher order theory of interpretations. In set-based semantics a predicate always gets assigned a set-sized extension, and the quantifiers always range only over members of a given set. Higher orderists insist that their theory has no such size-restrictions. Let us focus on a 'higher orderist' theory spelled out in the usual language of higher order logic. I call 'concepts' whatever we range over with second order quantification. Nothing I say relies on calling the values of higher order variables 'concepts': we could call them 'pluralities', if we wished to. Higher orderists insist on a non-reductive view on their higher order theory. Concepts are not sets, they argue. First order quantification in their meta-theory ranges over absolutely every object, sets included: it is absolutely general. In the higher orderist's semantics, an interpretation chooses a concept as the range of (first level) quantifiers, and a concept is the value of a predicate, as well. Usually, higher orderists will help themselves to a higher order theory with a fully unrestricted comprehension principle. For example, given any $\phi(x_1, .., x_n)$ without X free, the following is a provable schema:

$$\exists X \forall y_1, .., y_n (X(y_1, .., y_n) \leftrightarrow \phi(y_1, .., y_n)) \qquad \text{(COMP1)}$$

[5] I am not talking about models of PA but about models of a first level language that are definable inside PA. So, the existence of non-standard models of arithmetic is irrelevant, here, unless we assume that when we 'talk' in arithmetic, we are 'really talking' from the perspective of some uncountable non-standard model, and we assume that it makes sense to even make such a statement.

No such unrestricted comprehension principle is available in set-theory, of course. Any X is the value of a (first level) predicate under some assignment, even those obtained by comprehension through a formula that does not define a set, like 'x is a set' or 'x is not the empty set', etc. No such values for predicates are available in set-based semantics.

By comprehension, there is a concept that applies to anything that is self-identical: a universal concept that applies to every object whatsoever (not to every concept though, of course). So, there is a higher order interpretation of the language of ZFC where the domain is the universal concept and where quantifiers are 'absolutely general': they range over any object whatsoever.[6] Again, no such interpretation is available in set-based semantics.

By comprehension there is a concept of being a an element of, so there is also an interpretation where, for any x and y, x and y satisfy '$v \in u$' exactly when x is an element of y, and where quantifiers range exactly over all sets. Truth under this interpretation satisfies the Tarskian, recursive axioms for truth for the language of ZFC. No equivalent model is available in set-based semantics.

4.2 Self-applicability in trouble

These shortcomings of set-based semantics are not related to self-applicability. To show that some self-applicable theory cannot be general is not the same as showing that any self-applicable theory cannot be general. To argue for the latter claim, higher orderists bring up two main arguments: a Russellian and a Cantorian argument.

4.2.1 Russellian argument

The following objection originated from Williamson 2003, 426. It was then referenced in a number of articles and books, with some minor modifications (Rayo and Williamson 2003, Linnebo 2006, Florio 2014) and Studd 2019. The objection relies on the assumption that, to be general, a conceptually adequate theory of validity needs to be 'semantically open':

Semantic Openness Given a predicate P of the language we are interpreting, we should be able to interpret P by any open formula ϕ of the meta-language.

[6] For a defence of absolute generality, see Williamson 2003. For a contemporary discussion of unrestricted quantification and its critics, see Rayo and Uzquiano 2006. I suppose one should not take the existence of such an interpretation to show that absolute generality is possible; rather, it presupposes it.

Semantic openness is necessary for generality; any condition looks like an acceptable reinterpretation of P, so if the theory cannot interpret P as that condition, some interpretation is missing and the theory is not general.

In a theory of interpretations we usually define a concept of satisfaction in an interpretation and under an assignment: something is true under an interpretation if satisfied in that interpretation under all assignments. Here is some dummy notation I use: $\mathcal{I}$ is a meta-variable for interpretations. I often do not distinguish between a formula and its code for notational simplicity. However, I often use semi-quotes: I read $\ulcorner xyz \urcorner$ as the result of concatenating the code of x with the code of y with the code of z. $\ulcorner x \urcorner$ is the result of concatenating the code of x with nothing; i.e. it is the code of x. Where $x_1, .., x_n$ are the values of $v_1, .., v_n$ under an assignment a, to specify that $\ulcorner P(v_1, .., v_n) \urcorner$ is satisfied in $\mathcal{I}$ under a I say that $\ulcorner P(v_1, .., v_n) \urcorner$ is *true of* $x_1, ..x_n$ under $\mathcal{I}$. If so, I say that the predicate P is true of $x_1, .., x_n$ under $\mathcal{I}$. I assume we have a way to code n-tuples, and I denote the n-tuples of x_1, .., x_n by $\langle x_1, .., x_n \rangle$. I use $\mathrm{Tr}(x, \langle y_1, .., y_n \rangle, z)$ to say that x is true of $y_1, .., y_n$ under z. I abbreviate a string of quantifiers $\forall x_1 .. \forall x_n$ as $\forall x_1, .., x_n$. This notation is useful nonsense; it is just a way to talk about different theories of validity, which will mold Tr, $\mathcal{I}$ etc as they see fit.

The Russellian objection translates semantic openness in a more formal requirement via the following rule:

A1 If an n-ary predicate P means that ϕ under an interpretation $\mathcal{I}$ then, for any $x_1, .., x_n$, P is true of $x_1, .., x_n$ under $\mathcal{I}$ exactly when $\phi(x_1, .., x_n)$.

In dummy notation, if under $\mathcal{I}$ P means that ϕ, then the following holds:

$$\forall x_1, .., x_n (\mathrm{Tr}(\ulcorner P(v_1, .., v_n) \urcorner, \langle x_1, .., x_n \rangle, \mathcal{I}) \leftrightarrow \phi(x_1, .., x_n))$$

If 'To be human' means to be a cat under $\mathcal{I}$, then 'To be human' is true of x under $\mathcal{I}$ iff x is a cat. A1 is implicit in the argument; without it, we could not infer from the fact that P means that ϕ under $\mathcal{I}$ to the fact that P is true of x under $\mathcal{I}$ exactly when $\phi(x)$.

From semantic openness and A1, we obtain a schema. Since the theory is semantically open, we ought to be able to interpret P by any condition ϕ. If so, following A1, we can infer that, for any x, P is true of x under such an interpretation exactly when $\phi(x)$. The following schema holds, given any ϕ of the meta-theory which applies to things of the like of $\mathcal{I}$:

RUS There is an interpretation such that, for any x, P is true of x under $\mathcal{I}$ exactly when x is ϕ. In dummy notation:

$$\exists \mathcal{I} \forall x (\mathrm{Tr}(\ulcorner P(v) \urcorner, \langle x \rangle, \mathcal{I}) \leftrightarrow \phi(x))$$

However, the Russellian objection shows that RUS is inconsistent with self-applicability. Consider a predicate P of the language. This is a formula definable in the language of interpretations:

$$\gamma =_{df} \neg \mathrm{Tr}(\ulcorner P(v) \urcorner, \langle u \rangle, u)$$

By semantic openness, there must be an $\mathcal{I}$ where we interpret P by γ. By A1:

$$\forall x (\mathrm{Tr}(\ulcorner P(v) \urcorner, \langle \mathcal{I} \rangle, x) \leftrightarrow \neg \mathrm{Tr}(\ulcorner P(v) \urcorner, \langle x \rangle, x))$$

We obtain a contradiction by instantiating $\mathcal{I}$ for x. The argument is structurally like Russell's paradox, where RUS acts like a comprehension schema, hence my name 'Russellian objection'.

4.2.2 Size restrictions

Standard model theory fails RUS (unless we restrict formulae to specific parameters). In particular, no model satisfies RUS for the set-theoretic equivalent of the Russellian formula. The reason is simple: to explicitly define satisfaction under a model and assignment in set-theory, satisfaction must be defined only over assignments from a set: the domain. Any ϕ that fails to define a set will fail RUS because, regardless of set-based interpretation $\mathcal{I}$ we pick, some x that $\phi(x)$ will not be in the domain of the model, and so it will not be the case that $\mathrm{Tr}(\ulcorner P(v) \urcorner, \langle x \rangle, \mathcal{I})$. All we can do is have a T-schema for every x 'inside' the domain of the interpretation.

A1 is satisfiable only if 'To be true of v under w' is defined for everything. For otherwise, when we instantiate ϕ with $\ulcorner v = v \urcorner$, A1 will invariably fail as there is no set of all sets, so in any model M, for some x, '$\mathrm{Tr}(\ulcorner v = v \urcorner, \langle x \rangle, M)$' fails. In virtue of this reliance on unrestricted interpretations, the Russellian objection has been discussed in the literature in close connection with 'absolute generality'. Williamson (2003) introduces the paradox as a way of showing that a naive, first orderist view on absolute generality is paradoxical. Linnebo (2006) takes the paradox to be the strongest case against unrestricted quantification.

Yet, the paradox seems much more pervasive, and works even in contexts where absolute generality is irrelevant. For example, we can derive the paradox in the arithmetised version of model theory in PA, where we are only quantifying over the

natural numbers. The reasoning will be different, yet it leads to the same result: in PA we are only able to talk about conditions ϕ specifiable in the language of arithmetic. According to A1, P means that ϕ under $\mathcal{I}$ exactly when, for any natural number n, n satisfies P exactly when $\phi(n)$. The Russellian paradox follows as before. We do not need to assume absolute generality; rather, absolute generality looks like a special case of the paradox, where the meta-theory is absolutely general. Studd 2019, 81 makes a similar point. Studd ultimately thinks that the problem lies in the assumption that the meta-theory and the range of the quantifiers under the interpretation share the same universe. This is so in arithmetised model-theory, where models always range over all natural numbers, and so does the meta-theory. I agree that the minimal assumptions needed to get the paradox are A1, semantic openness, self-applicability plus the assumption that the language we interpret contains predicates of the same level of 'To be true under an interpretation'. However, I will argue against A1.

4.2.3 Typing

In typed-theories like the ones Rayo and Uzquiano (1999) and others defend, the paradox is stopped because satisfaction is typed. For example, say that we are interpreting a second level language: a language with variables of level two but no second level predicates (no predicates that apply to second level variables). Interpretations of second level languages are second order concepts: the values of second level variables. Therefore, satisfaction (which applies to interpretations) must be a second level predicate. Yet, such a predicate is not part of a second level language. So, the theory does not interpret its own language because it cannot interpret satisfaction, nor can satisfaction be the reinterpretation of any predicate the theory is able to interpret, because all such predicates are first level. If we were to add a second level predicate, higher orderists will insist that now the whole theory needs to be defined one order higher, so now satisfaction would be third level, and we can repeat the same argument.

Often it is said that higher orderists stop the Russellian objection by denying that interpretations are objects. One line of reasoning is that the paradox arises because quantifiers under an interpretation can range over interpretations; by making interpretations concepts, we deny this ability to 'introspect', which leads to the paradox. This line of reasoning suggests that the problem eventually lies in the relation between the range of quantifiers in the meta-theory and under an interpretation. This is incorrect, though. If the problem was that interpretations of a first level language cannot be objects otherwise they could 'talk about themselves', you would expect that

an interpretation of a second level language cannot be a second order concept. Yet, following Rayo and Uzquiano (1999), a higher order interpretation of the language of second order logic is a concept. In particular, there is an interpretation where the second order domain is any concept; quantifiers under such an interpretation have the same range of the quantifiers in the meta-theory, yet the language for which truth under an interpretation is defined is so weak that no paradox results from this; in particular, the language does not contain any second level predicate. In other words, the paradox is avoided because the theory is strictly typed. It does not matter what the nature of interpretations is – whether interpretations are first order objects or concepts – what matters is whether they can offer an interpretation of the satisfaction predicate. In a higher order framework, the paradox is stopped because we are not able to interpret satisfaction in the language of satisfaction: because validity is typed. The move from object to concept is just a tool to achieve typing.

The Russellian objection can be re-stated at every order of the hierarchy. For semantic openness is as plausible in its n-th order version as it is in its first order version. Suppose we are interpreting a language of level n with a predicate P that applies to n-th level variables. We can ask whether the system offers arbitrary interpretations of P, and it will fail to do so if P cannot be interpreted by any open formula $\phi(X^n)$, where X^n is n-th level. We can now reinstate an explicitly higher order Russellian objection, using the Russellian instance '$\ulcorner P(V^n)\urcorner$ is not true of X^n under X^n'. Again, higher order, typed-theories will avoid the paradox by typing: satisfaction under an interpretation for a language with predicates like P is to be defined at order $n+2$, so it is a predicate of level $n+2$, too high to be interpretable by the system.

4.2.4 Cantorian argument

A Cantorian argument against self-applicability is often discussed in the literature (Rayo 2006, Linnebo and Rayo 2012). Say that our aim is to develop a semantic theory for a language with variables up to level n. We can define in higher order logic what it means to be a surjective function from order n to order $n+1$. This can be done at order $n+2$ or already at order $n+1$ by coding the function as a relation between X^n and the extension of a concept X^{n+1}.[7] In both cases, we can show that there is no surjective function from order n to order $n+1$.

[7] I am more precise in the technical appendix, section six of this volume.

The upshot is that, from the standpoint of order $n+1$, we can show that there are strictly more concepts of order $n+1$ than concepts of order n (or objects, if $n = 1$). Higher orderists reason as follows: suppose our theory of validity is self-applicable; interpretations, then, must be concepts of some order or other: say they are of order n; the satisfaction predicate must correspond to a relation between formulae and concepts of order n. Via a higher order semantic theory, from the perspective of order $n+1$ one can show that there are at least as many interpretations of the satisfaction predicate as there are concepts of order $n+1$. By Cantorian reasoning, there are strictly more concepts of order $n+1$ than concepts of order n. Interpretations are concepts of order n, however, so some interpretations of the satisfaction predicate must be missing; the semantic theory is not general.

What is the conclusion? Higher orderists are trying to convince us that validity must be general, but that if it is general it is not self-applicable. However, is that enough to show that we must avoid self-applicability? In the next section, I argue that this is not the case. Often self-applicability is required of an adequate notion of validity, given the use we make of the notion of validity.

4.3 Arguments for self-applicability

4.3.1 Cross order universality and semantic luminosity

Here are some general reasons in favour of self-applicability. Self-applicability naturally follows from the desire to define a universal theory of validity. The relation between a theory of validity and its own language is like the relation between a list and a set of rules about how to write a list (which is itself written as a list): either it talks about itself, or it is missing some lists; ideally we are after a theory of how to write a list, not a theory of how to write some lists but not some others.

Another intuitive appeal of self-applicability comes from what I would call 'semantic luminosity'. If a theory of validity is a semantic theory, then its aim is to interpret and to make sense of a language. It would be nice, then, if we could interpret the language we are using when interpreting other languages, to gain semantic grasp on the very tools we are using. English is an example where this happens: from elementary school English speakers are taught English in English. They constantly gain semantic insight on the semantic rules of English, and by doing so they warrant their own use of English 'from the inside', so to speak.[8] If the system of validity

[8]I am not saying that only if we can fix first an interpretation of our own language can we understand it, because that gets things backward. The meta-language must be pre-interpreted if we

at hand is not self-applicable this form of semantic luminosity is impossible, and we need to pass on the job of shedding light on the meaningfulness of the meta-theory to some other semantic theory. We have no other choice, because the interpretations we define are not capable of making sense of what they are doing; they are blind to their own machinery. This other semantic theory in turn, unless it is self-applicable, will need again to rely on another semantic theory to shed light on what it is doing, and we never reach a point where what is used is also made sense of. The upside is that, whenever we ask whether we are making sense in the meta-theory, there likely will be a positive answer in some other, stronger theory. The downside is that, for any answer, there is a new dilemma, and one might get the feeling that, instead of answering the question, we are pushing it up at higher and higher orders.

4.3.2 Argument from truth

Self-applicability is needed to make sense of the way we talk about validity, for some seemingly meaningful pieces of reasoning about validity can be expressed only if logical validity is untyped. I call a theory of validity 'typed' if validity is typed. Validity is typed when truth under an interpretation is typed. Truth under an interpretation is typed, in turn, when it applies only to sentences that do not contain the predicate 'To be true under an interpretation' or something equivalent. A theory is type-free when it is not typed. A self-applicable theory of validity is always type-free, by definition.

We can adapt some illuminating examples from Kripke (1976) to show why a theory of logical validity ought to be untyped. Consider the following:

(K1) Each sentence Nixon uttered during the Supreme Court case against him is logically false;

(K2) What Nixon said during the Supreme Court case is contradictory.

Admittedly, even though Nixon lied a few times during the trial, he did not utter only logical falsities, so K1 is false. On the other hand, there were inconsistencies in what he told during the trial, so K2 is probably true. Now suppose that, during the trial, Nixon uttered K3:

(K3) Everything I say during this trial is tautological.

---

are using it at all.

As odd as K3 might seem, it looks meaningful. In fact, it seems false and thus meaningful, for something nonsensical can neither be true nor false to begin with. Adding K3 to the set of sentences Nixon uttered during the trial does not change our evaluation of K1 or K2, either: they are still false and true, respectively.

Suppose now that our theory of validity is typed: it is expressed in a language L but does not apply to it. So, validity for L must be theorised in some distinct L^+. If someone wishes to check the logical validities of L^+, they need to go even 'higher', in some L^{++}, and so on. It is then impossible to make sense of Nixon's statement, for part of what is said during the trial is K3 itself. So, to be meaningful, K3 must be at least one level higher than any of Nixon's statements during the trial; in particular higher than itself. What makes things even weirder is that Nixon might not have said K3 during the trial. In that case, K3 would be perfectly meaningful and also false. So, whether K3 is meaningful or not depends on when it is said, something which not only looks unpalatable, but goes contrary to our intuitive evaluation of K3: K3 looks false irrespective of when it is said.

We can give another example which is not obviously self-referential. Consider these sentences:

(K4) K5 is a logical falsity;

(K5) K4 is a logical truth.

I am convinced that I understand what K4 and K5 mean. I can reason with them: suppose K5; 'being logically true' is factive, so K4 is a logical truth, and so K4, through the same reasoning; but then K5 is a logical falsity, and since logical falsities are not the case, not K5. Contradiction. Therefore not K5, and thus K4 is not a logical truth. This seems to me a perfectly reasonable argument. Also, contrary to Kripke's analogous example, K4 and K5 are more 'innocuous' and do not seem paradoxical, because we cannot infer from the fact that K4 is not a logical truth that K4 is false.[9] Yet, again, how can a non-self-applicable theory of validity make sense of these sentences? For, according to it, K4 must be spelled out in a theory one order higher than the theory of K5, and K5, by the same reasoning, must be spelled out in a theory one order higher than the theory of K4.

[9]Some think that truth-like paradoxes arise for a notion of naïve validity (Beall and Murzi 2013, Murzi and Rossi 2017). In the example discussed we just used the quite sensible axiom: if $\ulcorner\phi\urcorner$ is valid, then ϕ. Murzi and others think that an introduction rule also holds of their notion of validity: if $\ulcorner\phi\urcorner$ is proved, infer that $\ulcorner\phi\urcorner$ is valid. If so, a paradox follows by Montague's theorem. The notion of validity at hand however is 'logical validity', which, as the authors themselves make clear, it is not the notion of naïve validity they have in mind.

To make sense of sentences like $K5$ and $K4$ and similar, it is necessary and sufficient that a theory of validity applies to statements about validity: that it self-applies. In fact, we have a very well-known example where this happens: model theory. The reasoning above can be spelled out in model theory via diagonalisation. Also, ZFC proves the principle I have employed: if $\ulcorner\phi\urcorner$ is valid then ϕ. For even though no model is the intended one for the language of ZFC, reflection principles ensure that a formula holds in ZFC only if it is satisfiable in a set-based model of the von Neumann hierarchy, which by contraposition implies that if $\ulcorner\phi\urcorner$ is valid (i.e. true in all set-based models) then ϕ.[10]

4.3.2.1 Meaning and content

Here is a way to explain away the appearance of how K3-K5 might look meaningful, from the perspective of a typed theory of validity.[11] When we deal with context-sensitivity, we can reason with an expression even though the expression itself does not have a content at the present context. For example, 'This is that' and 'This is human' imply 'That is human', even though I have made no attempt to assign a referent to 'This' or 'That', so they have no content at the present context. In Kaplanian terms, we can distinguish between the character and the content of an expression: the character is the rule we follow when we fix the semantic content of an expression given a context; the content is the extension of an expression given the context (1989). 'This' and 'This is that' are content-less at the present context, and therefore meaningless in some sense. Yet, in another sense they are meaningful because they possess a character.

We do not need to assume that the meta-logical vocabulary is context sensitive to apply the character-content distinction in the present discussion, because Kaplan's thought extends beyond context sensitive expressions. If Stalnaker is right, 'The king of France in 1990' can be used to refer to something only if there is a king of France in 1990. Since there is no such person, the (use of the) expression is 'content-less', and so is (the use of) 'The king of France in 1990 is bald' – the sentence does not express any proposition. Yet, both the expression and the sentence are meaningful in some other sense, because we understand how the first can be used to refer to something, and how 'The sentence could be used, in certain circumstances, to say something true or false' (Stalnaker 1978, 330).

[10]Reflection principles were first proven independently by Montague (1961) and Levy et al. (1960).
[11]I thank Tim Williamson for the suggestion.

Similarly, one could argue that K3-K5 are meaningful even though they are content-less at every order of the hierarchy, because they retain a character. For the higher orderist, a sentence like K3 presupposes that its content belongs to an order in the hierarchy to which it cannot belong, so there is no proposition that can be expressed by using K3. This does not preclude reasoning with these sentences, however; we can reason with sentences where 'This' occurs even though the expression is content-less, and likewise we can reason with K3-K5 even though they are content-less.

One could complain that the argument above with K4-K5 does not only shows a correct pattern of conditional reasoning but it is an example of *consequentia mirabilis* – ¬K5 is a theorem. So, assuming that we trust what the theory is saying, ¬K5 should be true, which it cannot be if it is content-less. Yet, this line of response is not very convincing. First of all, one cannot just assume that we trust what the theory is saying, because the theory is assuming that validity is untyped, which is exactly what it at stake, here. Secondly, by following this line one could incorrectly argue that, since 'This is this' is logically true, 'Something is this' must be true as well. Yet, 'This' is content-less, so nothing is this, in the present context. The correct conclusion is that the fact that the sentence is logically true only shows that, provided it does express a content, it expresses a true one. Yet, it does not express anything at the present context so it is not true. Similarly, that ¬K5 is a theorem only shows that, provided it does express something, it says something true, *modulo* the acceptance of the axioms of the theory. Yet, for the typed-theorist it does not express anything, so it is not true.

Here is a better observation: Kripke (1976) would insist that the liar sentence is neither true nor false in every context: it is 'ungrounded', as he puts it. Yet we have some semantic grasp on it, it is not just gibberish. We can reason with it and we can create meaningful complex sentences out of it. This is an advantage of Kripke's theory over Tarski's, where the liar sentence is not even well-formed, so it is simply gibberish. Similarly, the character-content analogy here misses the point. The friend of self-applicability can agree completely that K3-K5 are content-less but meaningful. They can be neither true nor false, pretty much like the liar sentence in Kripke's theory. The problem, however, is that typed-theorists are like Tarski: they cannot do the same, because their theory is typed. To say that K3 is meaningful but content-less they would need to formulate K3 first, and that is exactly what they cannot do. In the theory of higher orderists K3-K5 are not even well-formed sentences. For example, if satisfaction is a concept of order six, it will be denoted by a predicate of

level six. K3-K5 presuppose self-applicability, so we need to interpret predicates at least of level six. To do so higher orderists insist that we need interpretations to be concepts of order six. Yet a predicate of level six cannot apply to variables of level six, so the formalisation of any of K3-K5 is not even well-formed: they are content-less but also meaningless.

4.3.2.2 Truth vs validity

One could argue that the argument from truth is a double-edged sword. If someone is not convinced by Kripkean arguments for type-free truth theories, they will require truth to be typed. If so, they will not be swayed by this argument, either and, in fact, they might try to turn the argument from truth against the proponent of self-applicability. For if truth is typed, then there are good reasons to think that truth under an interpretation ought to be typed, as well.

Things are not so simple, however. The orthodoxy is to believe that truth for a language L is undefinable in L, hence the need to type. Yet, it is commonly assumed that no such problem arises for validity. Anyone would admit that we cannot define truth for set-theory in set-theory, yet anyone who follows the standard model-theoretic definition of validity would agree that we can define logical validity for set-theory in set-theory. So, reasons for typing truth do not directly carry over to validity.

Also, as I have mentioned above, sentences like K3-K5 do not seem paradoxical, contrary to their Kripkean counterparts about truth. The problems that higher orderists spot about a self-applicable notion of validity are not like the problems of definability that concern a self-applicable notion of truth in an important respect: the second kind of problems arise when we try to define truth in a way that is at least extensionally adequate; on the other hand, we can define an extensionally adequate notion of logical validity for the language of ZFC in ZFC. The alleged problems that higher orderists spot are rather at the level of conceptual adequacy, at the intersection between interpretations, truth and generality, where much more philosophical claims have been assumed, and where we have room for manoeuvre if we wished to face the objections.

4.3.3 Expressibility and cross order universality

The argument from truth is only one of the many faces of a more general expressibility problem that affects all typed theories of validity: the way we form and express general theories about validity in philosophical logic requires self-applicability and a

type-free notion of validity. Consider the discussion between a monist, a pluralist and a nihilist about logical consequence.[12] It is hard to pin down exactly where the core of their disagreement lies in a way that works for everybody; yet, it is common to describe it as follows: monists believe that there is a 'one true logic' that is 'correct', in some absolute sense of the term. When spelling out their position, monists will likely quantify over some relevant cases. A case, in turn, will likely comprise of a context and a language. Roughly, they are claiming that there is a set of rules and axioms that apply to all cases whatsoever: to any context and in any language. Of course, there is a shallow sense in which the rules for quantifiers do not apply to the language of propositional logic, since it is not a language of quantification. Their claim can be refined as follows: in any language capable of expressing the relevant logical expressions, such and such rules and axioms about those expressions hold in any context whatsoever. This is the claim I am interested in, which I believe is quite common among monists.

Nihilists are in a similar position: they, too, are claiming that logical laws ought to hold in all relevant contexts and languages. However, they also claim that no law has such a feature, and they conclude that there are no logical laws (Russell 2018). Pluralists are in a somewhat different position: they claim that no rule or axiom is valid in all cases, but rather what is valid changes with the case at hand. The pluralist's position is not simply the negation of the monist's position, since this is compatible with the nihilist's position, as well. They are somewhat in the middle: they are committed to there being validities, but they deny that they are absolute.

When monists claim that excluded middle is unrestrictedly valid, part of what it means to understand their claim is to believe that their statement really covers any context and any language capable of expressing disjunction and negation: it is universal for this class of languages. Their claim needs a notion of validity that is universally applicable to any language with such properties, regardless of its order in the hierarchy. It requires what we might call 'cross order universality'. Assuming that the meta-theory itself is spelled out in a language with negation and disjunction, their statement is meant to apply to their language and context, as well. It would be odd and borderline incoherent for them to hold that classical logic is the one true logic, while at the same time refusing to admit that 'Excluded middle is valid or not' is valid; that is, valid in the language of the theory about what is valid and in the context of them pondering about what is valid. However, if the system cannot self-apply, then

[12]For an overview of the discussion, see Russell 2019.

the monists' claim cannot possibly cover their own language and context. Yet it does, so the monists' position is expressible only if their theory is self-applicable.

The same problem haunts the logical nihilists. They are, too, making a general claim, which is inexpressible unless the system self-applies: excluded middle is not valid in any context and language capable of expressing negation and disjunction, and in particular in the context and language in which they are speaking. The expressibility problem applies more indirectly to the pluralists: they will hold that validity is contextual in some way or another; maybe they will want to say that excluded middle is invalid in intuitionistic cases and valid in classical cases. From this claim the status of excluded middle in their specific case cannot be settled, so the lack of connection between what the theory should say and what it can say in the absence of self-applicability is less apparent. However, their claim is still meant to be cross order universal, so part of what they are saying is that, if their case is classical then 'Excluded middle is valid or not' is valid and, if their case is intuitionistic then 'Excluded middle is valid or not' is not valid. Again, this is not even statable unless self-applicability holds.

4.3.3.1 Self-applicability and normative constraints

The monist's, pluralist's and nihilist's positions bear normative consequences, which however do not follow unless validity is not typed. Particular care needs to be applied when we try to bridge from a theory of logical entailment to a theory of reasoning, as Harman (1884) has taught us. Yet, many think that the monist's or pluralist's position will create normative constraints on one's beliefs. I take Field 2009 as an example. If one is unhappy with Field's proposal, they can take their favourite 'bridge principle' instead – the argument will likely go through anyway, as it does not depend on the specifics of Field's theory. Field believes that validity constrains an agent's overall degree of belief. To model an agent's degree of belief, we can use probability theory; degrees of belief are represented by numbers in the $[0, 1]$ interval (Genin and Huber 2021). Where $cr(A)$ is the degree of belief in A, the uncertainty of A is defined as $1 - cr(A)$. Field roughly proposes the following rule, which applies well to the monist:

Belief rule If B follows from $A_1, .., A_n$ in the light of an agent's logic, then their degree of belief ought to be such that: $u(B)$ is less or equal to the sum of $u(A_1)$, .., $u(A_n)$.

If Maria is a monist who believes in *modus ponens*, then she ought to regulate her overall degree of belief so that it tracks the rule. If she is pondering on the shoes she should wear for the day, she should not be more certain of the fact that it rains and if it rains streets are wet than she is of the fact that streets are wet. For if she is a monist about *modus ponens*, she believes that the rule holds in any language and context, and in particular in English and in the context of her pondering about shoes.

If Maria is a pluralist, then the rule might be changed to arrange for shifts in cases: given the case at hand, her degree of belief ought to adjust to what is valid in that case. If a pondering-about-shoes case is classical, then what we said again applies; if it is not, then she should react accordingly. What happens, though, if self-applicability fails? Then whatever theory of validity Maria has in mind can never apply to her case, and the normative constraints that we expect the theory to have are cut off. Both the monist and the pluralist are unable to correctly infer the normative consequences their positions bear in their context, and to regulate their subjective degree of belief accordingly.

4.3.3.2 Schemata and universality

Prima facie the problem that haunts the typed theorist is similar to the one that haunts what we might call a 'generality-relativist'. Generality-relativists believe that quantifiers can never be absolutely general and their range is always limited. Many believe that generality-relativists are unable to express that everything whatsoever is self-identical, or that everything whatsoever is material, or similar logical and metaphysical claims. On pain of contradiction, they are also unable to quantify over all contexts and say that, for any context, there is something impossible to quantify over in that context. For then in their context it would be true that something is impossible to be quantified over, which is contradictory (Williamson 2003). Similarly, typed-theorists are unable to talk about what is valid in all languages and contexts, for that presupposes that they are able to define validity in their language and context, as well.

One attempted solution is to replace generalisations with suitable schemata.[13] We can try a similar approach here. However, care needs to be applied to distinguish between the two issues. The problem is not that the typed theorists cannot quantify over cases, but that they lack a universal notion of validity which applies to any case. All they have are different concepts of validity, which only apply to languages up to some fixed level.

[13]For a detailed discussion see Studd 2019, chpt. 5.

Here is how a schemata strategy could work. A minor obstacle is that different languages might have different symbols for disjunction or negation. We use $\sim^x$ and $+^x$ as meta-variables for the codes of negation and disjunction, respectively, of a given language x.[14] Say that L and c are meta-variables for languages and contexts, respectively. We restrict our attention to languages of finite order. Say that $\text{lev}(L, n)$ abbreviates the condition of being a language of level n, and 'VAL^n' abbreviates the notion of validity which applies to languages of level up to n. Then the following schema can be used to express the monist's position about excluded middle, given any n:

(S1) For any L and c, if $\text{lev}(L, n)$ then, for any (code of) formula ϕ of L, for any c, $\ulcorner \phi +^L \sim^L \phi \urcorner$ is VAL^n in L and c.[15]

The pluralist's position, on the other hand, is not directly expressible through a schema. To at least capture the consequences of the pluralist's position, however, the typed-theorist can point at the following schemata:

(S2) For any L and c, if c is classical and $\text{lev}(L, n)$ then, for any ϕ of L, $\ulcorner \phi +^L \sim^L \phi \urcorner$ is VAL^n in L and c;

(S3) For any L and c, if c is intuitionistic and $\text{lev}(L, n)$ then, for any ϕ of L, $\ulcorner \phi +^L \sim^L \phi \urcorner$ is not VAL^n in L and c.

There are well-known issues with this approach, for a schema cannot be embedded in a sentence like an explicit generalisation, so it cannot be embedded in a conditional or negated. In the monist's view, for example, what is valid is constant across cases, so a monist will want to say that, if in one case excluded middle is valid, then it is valid in all cases. A schema cannot handle this inference accurately, for the consequent is all of the instances of the schema at once. The main issue with this approach, however, is that nothing has been done to recover self-applicability. The instance of the schema where we substitute L and c with the meta-language and the present context still remains unavailable, because it requires validity to be self-applicable, and it is not.

[14]Note that by identifying a symbol in the language with an pre-interpreted expression (disjunction, negation etc), we are making some assumptions about how the language ought to be interpreted. That is, we are not considering languages as purely syntactic objects devoid of semantics, but as partially pre-interpreted objects: the meaning of the logical expressions is already fixed, even though the logical expressions themselves might be represented through different symbols (hence the need to quantify over meta-variables for those different symbols).

[15]I read $\ulcorner \phi +^L \sim^L \phi \urcorner$ as: the result of concatenating (the code of) ϕ with the symbol for disjunction of L with the symbol for negation of L with (the code of) ϕ.

The meta-language where we express the schema needs to contain variables of any level, so it is not of any finite level. We could try to extend higher order semantics to ordinal level languages, so that we can extend the schema to all ordinal level languages and therefore (hopefully) to our meta-language. We obtain the following, given any ordinal κ:

(S4) For any L and c, if c is classical and $\text{lev}(L, \kappa)$ then, for any ϕ of L, $\ulcorner \phi +^L \sim^L \phi \urcorner$ is VAL^κ in L and c.

Linnebo and Rayo (2012) make the case that a similar semantic 'expansion' will be needed for ordinal level languages, *contra* self-applicability. S4 covers validity for the meta-language only if the meta-language L is of some ordinal level θ. If so, then interpretations for L are of order θ, so truth under an interpretation for L is a concept of order strictly higher than θ. By the same reasoning as above, using resources higher than θ, we can show that there are more interpretations of truth under an interpretation than interpretations (of order θ), so the theory is not general. Since validity must be general, we have failed to capture validity for the meta-language, contrary to our hypothesis. Hence, S4 did not cover the meta-language.[16]

It seems that, in any case we consider, the schemata strategy cannot possibly recapture self-applicability, and therefore it cannot recapture neither universality nor the normative constraints of upholding (or rejecting) the monist's position.

4.4 The untyped theorist's perspective

4.4.1 Against A1

The Russellian objection shows that self-applicability, semantic openness and A1 are mutually inconsistent. Typed-theories rejects self-applicability, model-theory restrict semantic openness to a universe. In this section I suggest that the problem might be with A1. I argue that A1 might work if the semantic theory is not self-applicable but it is unsuitable if the theory is self-applicable. As such, it cannot be used to argue against self-applicability.

[16]Studd (2021) considers the possibility of 'stopping' the ascent as some fixed ordinal κ. There is no further ordinal level language, so we cannot use $\kappa + 1$ resources to prove Cantor's theorem. The strategy would work, but I do not know how to argue for the idea that languages stop at some κ. I am told that after ω^{CK} – the first non-recursive ordinal – the ordering of the indices of variables will stop being recursive, so it is hard to even state the syntax of this language (Halbach 1997). It is not clear to me, though, whether this is a problem for the language of level ω^{CK} or for the limitations of our notion of 'language'.

I am writing in English. By 'English' I do not mean the syntax of the English language, but a certain interpretation of that syntax; English as a 'system of signs in actual use', as Davidson (1967) would put it. English is not a formal language, but examples stated in English can serve as a useful philosophical guidance, to check if our notion of truth under an interpretation fits our pre-formal intuitions. If we are able to show that A1 is misguided using only the assumption that 'To be true under an interpretation' in English is untyped, then there are strong reasons to reject A1 when we are working in formally defined, untyped theories of 'To be true under an interpretation'.

Suppose that, as before, we read $\mathrm{Tr}(x, \langle y_1, .., y_n \rangle, z)$ as an abbreviation for 'x is true of $y_1, .., y_n$ under the interpretation z'. Then, as we speak in English, when z is English, 'Tr' should collapse into simple truth. That is, English can talk about English; under English, 'To be true so and so interpreted' means 'To be true in English', which in English simply means 'To be true'. Suppose I wish to state the obvious: in English 'To be a tiger' means to be tiger. According to A1, this requires me to say that, for any x, 'To be a tiger' is true of x in English exactly when x is a tiger. This is a T-schema for 'To be a tiger'. Suppose now that I wish to say the same for the predicate 'To be true' itself. In English, 'To be true' means what it means: to be true. According to A1, this must mean that, for any formula x, thing y and interpretation z, x is true of y under z exactly when 'v is true of u under w' is true of x, y and z under English. Where E is English, this is what the biconditional looks like in our dummy notation:

$$\forall x, y, z(\mathrm{Tr}(x, \langle y \rangle, z) \leftrightarrow \mathrm{Tr}(\ulcorner \mathrm{Tr}(v, \langle u \rangle, w) \urcorner, \langle x, y, z \rangle, E)) \qquad \text{(Iter-1)}$$

A similar rule Iter-n holds of n-ary conditions, for any n.

Is the Iter schema feasible? If we substitute E for z, then it is effectively saying that truth in English can be iterated and discharged at will; that if ϕ is true in English then the sentence that says of ϕ that it is true is true, and vice-versa. Yet, this is a very strong assumption to make, and something which is not obvious at all in an untyped language. For example, in a formal theory of truth we can show that iteration (ϕ is true iff $\ulcorner \phi$ is true$\urcorner$ is true)[17], bivalence (ϕ is true or $\ulcorner \neg \phi \urcorner$ is true) and consistency (if ϕ is true then $\ulcorner \neg \phi \urcorner$ is not true) are jointly inconsistent: see lemma 14.3 in Halbach 2014, 146.

[17] Quotations here are simplified for readability. In the language of PA what we should write is something like: (the code of) ϕ is true iff t is true, where t is a closed term for (the code of) $\ulcorner \phi$ is true$\urcorner$.

In our dummy notation, these three axioms are jointly inconsistent:

$$\forall x, y(\mathrm{Tr}(x, \langle x, y\rangle, y) \leftrightarrow \mathrm{Tr}(\ulcorner \mathrm{Tr}(v, \langle v, u\rangle, u)\urcorner, \langle x, y\rangle, E)) \quad \text{(Iter2)}$$

$$\forall x, y(\mathrm{Tr}(\ulcorner \neg x\urcorner, \langle x, y\rangle, y) \rightarrow \neg \mathrm{Tr}(x, \langle x, y\rangle, y)) \quad \text{(Cons)}$$

$$\forall x, y(\neg \mathrm{Tr}(x, \langle x, y\rangle, y) \rightarrow \mathrm{Tr}(\ulcorner \neg x\urcorner, \langle x, y\rangle, y)) \quad \text{(Biv)}$$

Consider $\delta = \neg \mathrm{Tr}(v, \langle v, u\rangle, u)$. We reason as follows:

$$\begin{aligned}
&\neg \mathrm{Tr}(\delta, \langle \delta, E\rangle, E) && \\
\leftrightarrow &\neg \mathrm{Tr}(\ulcorner \neg \mathrm{Tr}(v, \langle v, u\rangle, u)\urcorner, \langle \delta, E\rangle, E) && \text{def } \delta \\
\leftrightarrow &\neg\neg \mathrm{Tr}(\ulcorner \mathrm{Tr}(v, \langle v, u\rangle, u)\urcorner, \langle \delta, E\rangle, E) && \text{Biv, Cons} \\
\leftrightarrow &\mathrm{Tr}(\ulcorner \mathrm{Tr}(v, \langle v, u\rangle, u)\urcorner, \langle \delta, E\rangle, E) && \text{Biv, Cons} \\
\leftrightarrow &\mathrm{Tr}(\delta, \langle \delta, E\rangle, E) && \text{Iter2}
\end{aligned}$$

The first line is inconsistent with the last.

Maybe we should give up iteration, consistency or bivalence. Yet it is not clear at all that we should keep iteration and give up either consistency of bivalence. And yet, if we believe in A1, there is no other option, unless in English 'To be true' does not mean to be true, which is absurd. Yet, is it not strange to think that a predicate can mean what it means only if bivalence or consistency fail? Shouldn't it be obvious that 'To be true' means to be true, regardless of bivalence or consistency?

There are other issues. I can coherently imagine and talk (in English) of a language where predicates mean the opposite of what they mean in English: a Mirror-English – call it E^*. In Mirror-English, for any predicate P, P in Mirror-English means to not P. According to A1, however, there cannot be any such language. For suppose there was; then 'To be true' in Mirror-English would mean to not be true. We can then apply A1 as follows:

$$\forall x, y(\mathrm{Tr}(\ulcorner \mathrm{Tr}(v, \langle v, u\rangle, u)\urcorner, \langle x, y\rangle, E^*) \leftrightarrow \neg \mathrm{Tr}(x, \langle x, y\rangle, y))$$

We obtain a contradiction by instantiating x with $\ulcorner \mathrm{Tr}(v, \langle v, u\rangle, u)\urcorner$ and y with E^*.

One could deny that it made sense in the first place to talk as if 'To be true' was part of English, or to talk as if English was like one of our formal languages. The argument does not really rely on using English, though: English just served as a dummy language and a useful comparison. Also, we wish to talk about what it means to mean something; surely English is meaningful since we use it, and it looks like we interpret other languages in English all the time. So, if our assumptions about what it means to mean something fail in English, why should we keep them?

What the argument suggests is that self-applicability itself is a reason to deny A1, which makes the reliance of A1 in the argument against self-applicability suspicious. If we deny A1, we follow a type-free approach to truth: we give up the notion that truth is typed. To make space for a consistent type-free truth predicate, the usual Tarskian axioms for truth in Tarski 1933 must be changed to avoid paradoxes. Since to be true is a special way of being truth under an interpretation (the intended), it is likely that truth under an interpretation is also untyped. Therefore, it will need a similar treatment.[18]

If I am right, A1 is mistaken. It is not true that if P means that ϕ under $\mathcal{I}$, then P is true of any x under $\mathcal{I}$ exactly when x is ϕ. At most P means that ϕ under $\mathcal{I}$ if P is true of any x under $\mathcal{I}$ exactly when x is ϕ. The biconditional in A1 is sufficient, not necessary. Hence we cannot infer the RUS schema from semantic openness. At most, we can infer semantic openness from the RUS schema.

4.4.2 An alternative test

In a self-applicable theory, I suggest the following test to check if P means that ϕ. First of all, 'means-that' is an intensional notion, yet our semantics is usually extensional; we can never fully capture 'means-that'. For example, if in Pseudo-English 'To be human' meant to be a featherless biped, an extensional semantics would not recognise any semantic difference between English and Pseudo-English with respect to that predicate, since anything is human if and only if it is a featherless biped. However, we can provide an extensional notion that is sufficiently similar to means-that for the purpose of interpreting a language. Davidson (1967) suggests that truth can be such a tool. We can apply Davidson's insight as follows: imagine a situation where Maria wishes to explain John pseudo-English. In pseudo-English, 'To be human' means what 'Être mortel' means in French, which means to be mortal. One way to explain this to John is by talking in English about pseudo-English, using the English equivalent of 'To be human'; Maria can say (in English) that 'To be human' in pseudo-English means to be mortal. Our system does not have a means-that operator, all Maria can do is translate this semantic relation in an (extensional) equivalence using truth; she can say that, for any x, 'To be human' is true in pseudo-English of x exactly when x is mortal: we have obtained the biconditional in A1.

[18]Kripke's theory of truth was axiomatised by Feferman (1991). Under the Kripke-Feferman axiomatisation, we could retain A1 while still holding that 'To be true' means to be true in English. That is because Iteration holds in Kripke-Feferman, while 'Biv or Cons' fail. However, A1 is still wrong, because it is simply inconsistent with the existence of plausible languages like Mirror-English.

Another way for John to understand pseudo-English is the one we just used when we first introduced it and compared it with French: Maria can talk in English about the semantic relations between English and pseudo-English. They can say (in English) that 'To be human' in pseudo-English means what 'To be mortal' means in English. In this way Maria never *uses* 'To be mortal', but only talks about it and about its meaning in English, which happens to be the language in which both Maria and John are speaking. As before, we can translate this semantic relation into a biconditional using truth. Maria can say (in English) that, for any x, 'To be human' is true of x in pseudo-English exactly when 'To be mortal' is true of it, in English.

I argue that both descriptions of pseudo-English are sufficiently adequate, and both show that, in pseudo-English, 'To be human' means to be mortal, i.e. what 'To be mortal' means in English. While in this particular case the two tests are equivalent, since 'To be mortal' is true of x in English exactly when x is mortal, in problematic, liar-like cases a T-schema that allows to discharge truth is not available, and the two tests will come apart. From the perspective of this second test, A1 suppresses the step from the sentence being true in English to its content being the case, which is harmless if the formula itself is not liar-like, but might become paradoxical if it is.

Here is the new test.

(A2) An n-ary predicate P means that ϕ under $\mathcal{I}$ if, for any $x_1, .., x_n$, P is true of $x_1, .., x_n$ under $\mathcal{I}$ exactly when $\ulcorner\phi\urcorner$ is (simply) true of $x_1, .., x_n$.

The test assumes that the theory is able to specify an untyped theory of truth, otherwise a concept of simple truth of x would not be definable for formulae of the language of the theory.

A2 does not lead to Russellian worries. Consider English (E) under which 'To be true so and so interpreted' means to be true *simpliciter*. As it should be, when we plug in E and 'To be true' in A2 we obtain only an obvious truism: that 'To be true' means 'To be true' in English if, for any x, y and z, 'To be true' is true of x, y and z in English exactly when 'To be true' is true of x, y and z in English.

In the Russellian example, there is no issue in interpreting P as '$\ulcorner P(v)\urcorner$ is not true of u under u'. Using A2, all I obtain is that, under such an interpretation $\mathcal{I}$, $\ulcorner P(v)\urcorner$ is true of $\mathcal{I}$ under $\mathcal{I}$ exactly when '$\ulcorner P(v)\urcorner$ is not true of u under u' is simply true of $\mathcal{I}$. No contradiction obviously follows unless a T-schema is available, which is unlikely precisely because truth is type-free. Even so, it seems to me that we have successfully interpreted P like we wanted, despite a failure in A1, because A2 holds.

Mirror-English is now unproblematic. We cannot infer any instance of A1 from the fact that 'To be true' means to not be true, in Mirror-English. Rather, we could test that in Mirror-English 'To be true' means to not be true by checking A1. Finding that it leads to inconsistency, we can resort to A2.

Higher orderists might claim that we have been too fixated on the semantics of 'means that'. We can spell out the objection using other terms, if we wished to; we could have said that, if the theory is general, we should be able to pick any condition, and take that to be the applicability condition for P. Yet, unless we are given independent reasons to embrace such a principle, it is not obvious that generality requires it. Before we had independent reasons for accepting RUS – we are doing semantics, we are interpreting predicates; surely generality implies that we should be able to *interpret* P by any condition! This is what generality is about: the ability to capture *all the interpretations* of a predicate. Yet, if I am right we dispelled that worry: we are not failing to interpret, which is what matters to generality.

4.4.3 Re-evaluating the issue at hand

If I am correct, the Russellian objection to self-applicability fails. Yet the Cantorian objection still stands. Higher orderists will reason as follows: take the self-applicable theory you prefer, call it T1. If we allow ourselves to climb up one order higher than T1, we can build a theory of validity for the language of the self-applicable theory, from the standpoint of a 'higher' meta-theory – call this new theory T2. From the perspective of T2, we can 'see' that we were missing some interpretations in T1 due to Cantorian reasons. Also, we can now prove the RUS schema in T2, even for sentences with the (old) truth predicate of T1 – a Russellian objection is reinstated through semantic ascent. There is no need to justify RUS independently if we can simply prove it.

The move to T2 is compelling only if one has already taken a certain stance on semantic ascent and on absolute generality. A non-reductive view on higher level languages is required: concept-talk is not reducible to objects-talk; higher order quantification is not first order quantification in disguise. Only within this framework, higher orderists can claim that first order quantification ranges over absolutely everything, even though the range of second order quantification is 'bigger'. Let us define a 'first orderist' as someone who denies that higher order quantification is not reducible to first order quantification. A first orderist will simply deny that higher orderists can both build a second order theory with full comprehension and, at the same time, insist that their first level quantifiers are ranging over absolutely everything in the

meta-theory. In the eyes of the first orderist semantic ascent is not particularly problematic; all it shows is that, if one does not start with every object in the domain of quantification, the (first order) domain can be enlarged. Yet, the first orderist can also insist that one can start with every object in the domain of quantification, and in that case semantic ascent is incoherent. So, in any case, either semantic ascent is not stopped but it is unproblematic, or it is stopped so the Cantorian reasoning cannot go through.

Let us grant the higher orderist's view on higher level languages. Still, the Cantorian line of reasoning assumes that there is a viable notion of validity where validity is *not* self-applicable, to conclude that validity cannot be self-applicable because it is general. It is useful to make an analogy with the discussion of truth. Suppose I claim that truth is untyped, and therefore an unrestricted T-schema for sentences about truth is not viable, even though it might seem intuitive at first. Of course one can always build a Tarskian truth-predicate on top of whatever untyped theory of truth I decided to build, and prove that such a typed truth-predicate does satisfy an unrestricted T-schema for sentences that contain my untyped truth-predicate. Yet, bringing this up is not particularly useful: to call the typed predicate 'To be true' is already to assent that truth is typed, which is what I am denying. Similarly, suppose I wish to claim that 'To be true under an interpretation' is untyped. It might not be a sound theoretical move to argue against my position from the perspective of a typed theory of truth under an interpretation, for the point at hand is precisely whether 'To be true under an interpretation' is typed or not. I will simply deny that the notion discussed in the objection is worth calling 'To be true under an interpretation' unless we untype it.[19]

Yet, would this staunchness on my part be feasible? Does it make sense to say that there is no viable notion of truth that is typed? If our aim is to capture *the* concept of truth in natural languages, then maybe one could argue that truth in English is untyped. Yet, this is often not what we are doing; often, we use English as a heuristics, but we do not aim at an analysis of 'To be true' in English. Our meta-theory might be some formal language like Peano arithmetic, where sentences are complex arithmetic codes. Since we are working in artificial languages, we have more freedom in our definition of truth, and we might give up some intuitive features

[19]The Cantorian objection does not require truth to be typed. I could ascend to a stronger language and define an untyped theory of truth under an interpretation, and show that I gathered more interpretations of the language. The higher orderist needs the theory to be typed to stop the regress and make the positive claim that, contrary to the self-applicable theory we started with, theirs does not have the same issues, since it does not self-apply.

of truth in natural languages to obtain other features. This point is even stronger for validity: outside of logic classrooms, people do not have any clear notion of logical validity in mind; if anything, 'logical validity' looks like a technical concept, a purified version of our much less precise notion of 'good argument'. If our aim is not to give an analysis of 'logical validity' in English, then insisting that logical validity must be untyped or general to be conceptually adequate does not seem a sound theoretic move. Even if we could agree that there is a definite, unique notion of validity for first order logic because of the completeness theorem, the same cannot be said for higher order validity, where no completeness theorem is available. What is more reasonable is to say that, especially for higher order logic, there is no single, pre-theoretic notion of 'logical validity' that we are aiming to define. Rather, there is a viable notion of logical validity which is typed, and there is also a viable notion of logical validity which is untyped, and both have their own advantages and disadvantages.

Advocates of an untyped theory of truth often point at the advantages of their theory over the classic Tarskian typed hierarchy. With an untyped notion of truth at hand, we are able to reflect on the system, and make generalisations that are impossible to state within a typed theory of truth. The same was claimed here about validity: untyped validity is necessary to state generalisations across languages, generalisations that matter to our discussion of logical validity in philosophical logic. If there is no unified, pre-theoretic notion of validity that we are trying to analyse, maybe we should pick the notion that best suits our theoretical needs. If our aim is to cover *all* interpretations, then the higher orderist's, typed notion might be best suited. Yet, at least when it comes to philosophical logic, we need to make generalisations across languages and we need to be able to derive conclusions that cover our own meta-language. To suit these theoretical needs, a self-applicable theory like the usual set-based semantics might be a better candidate.

Chapter 5

Higher Order Semantics and Non-standard Interpretations

I discuss and develop a semantics for cumulative languages of arbitrary (finite) level from the perspective of a higher order meta-theory. I discuss the features of this semantics and I compare it to standard set-based semantics. I then discuss 'non-standard' higher order interpretations. A non-standard interpretation is one that does not follow the standard match between level of expressions and orders of the hierarchy. I focus on a particular subset of those, called 'raising interpretations'. I first discuss and defend their plausibility. I then show some interesting features of raising interpretations: they are models of higher order logic; there are strictly more raising interpretations than standard interpretations; it is not possible to gather all the raising interpretations at any order of the hierarchy. I conclude the article with a comparison between standard validity and raising validity.

5.1 Validity, generality and ascent

It is often said that an argument is logically valid exactly when, under all the acceptable interpretations of the language, if the premises are true so and so interpreted, so is the conclusion. In elementary logic textbooks this idea is usually spelled out in Zermelo-Fraenkel set theory, by defining an interpretation as a particular set.[1] However, some have argued that the standard set-theoretic framework does not really capture the philosophically cogent notion of validity we have in mind, as it is

[1] A theory of syntax is also needed: it can be primitively added, or it can be formulated in set-theory itself.

structurally unable to specify *all* the intuitively acceptable interpretations of a given language.[2] Some conditions that do not define a set look like the appropriate reinterpretations of a predicate, yet there is no set-based interpretation where the extension of a predicate is not a set. Under the intended reading, for example, we are presumably applying the predicate 'Set' to all sets, yet there is no set of all sets, so no interpretation renders this reading. Yet we are using this interpretation as we speak in set-theory, which bears witness to its acceptability. Here is the problem: if we are not quantifying over all interpretations when we define validity, how do we know that we are achieving the correct results? It might be that something is true in all set-based interpretations and yet false in some unreachable one.

Kreisel (1967) points out that in first order logic we have independent confirmation that set-based semantics provides the 'correct' results. Assume that we are, in fact, missing some intuitively acceptable interpretation in set-based semantics. Call 'intuitive validity' preservation of truth under all intuitively acceptable interpretations. We can squeeze intuitive validity between provability in first order logic and validity in set-based semantics using completeness. Surely provability in first order logic is sound for intuitive validity: if something is provable in first order logic it is true in all intuitively acceptable interpretations. And surely if something is true in all intuitively acceptable interpretations, it is true in all set-based interpretations, because at most the latter is a proper subset of the former. Since first order logic is complete, something true in all set-based interpretations is provable, and thus true in all the intuitively acceptable interpretations: the two 'validities' are equivalent.

We also know that set-theoretic reflection principles are provable in ZFC.[3] A reflection principle roughly states that a condition holds in the universe of sets only if it is satisfiable in a set-sized model of the Von Neumann universe, even if that condition fails to define a set. In this sense, reflection principles roughly say that a formula of ZFC is satisfiable in some model which is not set-sized only if it is satisfiable in a set-sized model (Shapiro 1987). If we restrict our attention to a single sentence ϕ of ZFC, a reflection principle implies:

$$\phi \rightarrow \exists \mathfrak{M}(\mathfrak{M} \vDash \ulcorner \phi \urcorner)$$

[2]Kreisel (1967) shaped the discussion on this issue. The higher order perspective I discuss is pioneered by Rayo and Uzquiano (1999), Rayo and Williamson (2003), Rayo (2006), among others.

[3]Reflection principles were proven independently by Montague (1961) and Levy et al. (1960). We talk of 'principles' because there are several such principles, with different strengths. We have to note, however, that reflection principles probably do not hold for ZFCU, if we do not add the assumption that the urelements form a set (McGee 1997, 67-8).

Where $\mathfrak{M}$ is a set-based model and $\vDash$ the usual notion of truth under a set-based model. There cannot be some falsity that turns out logically true or some truth that turns out logically false in set-based semantics for first level languages. For the contrapositive of the above is:

$$\forall\mathfrak{M}(\mathfrak{M} \vDash \ulcorner\phi\urcorner) \rightarrow \phi$$

Which means that if $\ulcorner\phi\urcorner$ is logically true in set-based semantics, then ϕ.

As we venture into validity for higher level languages, however, things get complicated. Second order logic is incomplete under standard set-based semantics spelled out in first order ZFC, so no squeezing argument is available. Reflection principles can be proven only for the first level fragment of the language of pure set-theory, as the meta-theory is first order and second order reflection principles are not even statable without using second level sentences in the meta-language.

One could hope that by strengthening the meta-theory – for example by using second order ZFC (2ZFC) – we could extend reflection principles to the second level fragment of the language. Yet, this is not the case, as reflection principles are not provable for second order ZFC on pain of inconsistency. Second order ZFC is finitely axiomatisable, so we can write all of its axioms as one conjunction, call it Δ. If a reflection principle was available, we could prove in 2ZFC:

$$\Delta \rightarrow \exists\mathfrak{M}(\mathfrak{M} \vDash \ulcorner\Delta\urcorner)$$

Δ is provable in 2ZFC because it is a conjunction of axioms, so we just proved in 2ZFC that 2ZFC has a model, something which violates Gödel's second incompleteness theorem.

With no reassurance of the results we get, given the absence of completeness results and reflection principles, and with evidence that we are missing acceptable interpretations, the set-based theory of logical validity for higher level languages now looks conceptually suspicious and maybe even extensionally inadequate.

5.1.1 Semantic ascent

Many who are skeptical of the idea that model-theory captures all the interpretations of a given language defend a version of 'Semantic ascent' (SA, for short).[4] Say that our aim is to interpret a language with variables up to some finite level. Call the level

[4]Among the skeptics who endorse semantic ascent are Rayo and Uzquiano 1999, Rayo and Williamson 2003, Rayo 2006, Linnebo and Rayo 2012 and Florio 2014.

of a predicate the highest level of the variables it can be applied to. Call the level of a language the highest level of its bound variables. The bare level of a language L is determined as follows: where n is the highest level of the variables of L (free or bound) and m the highest level of the non-logical predicates of L, the bare level of L is $\max(n, m+1)$.[5] The bare order of a theory is the level of its (intended) language. So, the language of ZFC is a first level language, but also a language of bare level 2, since '$\in$' is not logical. ZFC is a theory of bare order 2 (and a first order theory). Any predicate language is at least of bare level 2. Now, a conceptually adequate theory of validity for a language L should be 'general' for L: a theory of validity that is able to model truth under any acceptable interpretation. The higher orderist spells out generality as follows: say that we wish to interpret a language L^k of bare level k with no constants or predicate letters.[6] We define 'strict adequacy' as follows:[7]

Definition 7 (Strict adequacy)**.** *Where V^n is a variable of level n, a theory T of validity for L^k is strictly adequate just in case the following holds:*

- *For any x and V^1 there is an interpretation in T where the value of V^1 is x;*
- *For any second order concept X^2 and second level variable V^2 there is an interpretation in T where the value of V^2 is X^2;*
- *....*
- *For any k-th order concept X^k and k-th level variable V^k there is an interpretation in T where the value of V^k is X^k.*

We can extend strict adequacy to syntactically richer languages if we wished to, following this mapping: the possible values of a term of level n are concepts of order n; predicate letters of level n have the same semantic type of variables of level $n+1$. In the standard higher order framework we can show the following:

(SA1) A strictly adequate semantic theory for a language of bare level n cannot be given in a meta-theory of bare order n;[8]

[5] Rayo (2006, 237) uses a slightly different terminology. Mine is closer to Yi 2005.

[6] An 'interpretation' here also involves providing a value to variables. In model-theoretic terms, an interpretation would be a model paired with a variable assignment.

[7] I borrow the term from Rayo 2006.

[8] We need to specify that we are talking about bare level languages since we can provide a strictly adequate semantic theory for a language of level $n \geq 2$ in some meta-theory of order $n \geq 2$, as long as the object language does not contain non-logical predicates of level n or free variables of level $n+1$. We will need a primitive satisfaction predicate of level n, which is not explicitly defined at level $n+1$. Since we are not quantifying over level $n+1$, the meta-language still 'counts' as level n, however it is a language of *bare* level $n+1$ using my terminology.

(SA2) A strictly adequate semantic theory for a language of bare level n can be given in a meta-theory of bare order $n+1$.

In Linnebo and Rayo 2012, it is claimed that a stronger version of semantic ascent can be proven for languages of any ordinal level, and not just for languages of finite level. To prove this stronger version, one needs to ensure that, for any ordinal, there is a language of higher ordinal level. The authors claim this can be achieved through the following principle:

Union-principle For δ a limit ordinal, suppose there are languages of level β for any β less than δ. Then there is a language of level δ (2012, 276).

However, Studd (2021) raised some doubts on whether the union-principle is plausible, and whether it really entails semantic ascent in the transfinite in a way that is not question-begging. In the following, I focus only on languages of finite level, even though I suspect that what I say arguably extends to languages of infinite levels, provided that Linnebo and Rayo (2012) are right about semantic ascent in infinite level languages.

5.2 Standard higher order semantics

In the appendix (chapter 6 in this volume), I develop a system of interpretations in a higher order meta-theory which can be applied to arbitrary cumulative languages of finite level. The essential philosophical take of the higher orderist's proposal is that, rather than interpreting a higher level language in some other theory like set theory, we should assume a non-reductive position, and interpret higher level languages directly from within higher order theories. From this non-reductive standpoint, we can prove both SA1 and SA2, as well as other interesting features. In this section I hint at how the semantics works, and explain the main results and points of interest. In section 2.1 I explain the meta-language and meta-theory I am working with, as well as the philosophical framework I am adopting. In section 2.2 I point at some interesting results that can be proved within this framework.

5.2.1 Meta-language and meta-theory

In the following discussion my meta-language is a cumulative, higher level language built over the language of set-theory with urelements. The meta-language is built on levels: for any level n and arity m, there are variables of level n and arity m

(I often omit the arity for readability). The language is cumulative: the result of applying a variable of level n to a string of variables results in a well-formed formula, as long as the variables in the string are of strictly lower levels than n. The types of variables are not linearly ordered: there is more than one type for each level, in fact one for any arity at each level. No cross order generalisation is expressible in the language, since variables are always bound to a specific level. However, generalisations restricted to some finite n are expressible as conjunctions, and existential statements as disjunctions.

I choose as my 'starting base' second order Zermelo-Fraenkel set-theory with urelements (2ZFCU). I chose this base for two reasons: it makes the comparison with standard model-theory more obvious and direct; in addition, if we did not quantify over urelements, some intuitively acceptable interpretations of a first level variable would be missing: any that map it to something which is not a pure set.

The language of 2ZFCU contains the symbol '=' for identity, '$\in$' for membership and a monadic predicate 'Set' for sets. There are no constants. To this language I add variables of arbitrary finite level, and I update the definition of well-formed formulae accordingly. Following cumulativity, where X^n is a variable of level n, $\ulcorner X^n(Y_1, .., Y_n)\urcorner$ is well-formed if each of 'Y_1', ..., 'Y_n' are of levels strictly lower than n. I let quantifiers bind variables of any level: $\forall X^n \phi$ is always well-formed.

To the theory 2ZFCU I add (cumulative) comprehension principles at any order, which ensure that, for any formula ϕ that holds of a string of objects/concepts up to order n, there is a concept Y^{n+1} of ϕ-ing. More formally, where the maximum among the orders of X_1, .., X_m is n, and Y^{n+1} does not occur in ϕ, the following is an axiom:

$$\exists Y^{n+1} \forall X_1, .., X_m (\phi(X_1, .., X_m) \leftrightarrow Y^{n+1}(X_1, .., X_m)) \qquad \text{(CPn)}$$

ϕ might contain more free variables than $X_1, .., X_m$. Comprehension is impredicative: there are impredicatively defined concepts, as well as predicatively defined concepts, and quantifiers range over both kinds. The assumption that n be exactly the maximum between the free variables we abstract away from plus 1 allows us to avoid some counter-intuitive results; for example, that there be a different concept of being the empty set for any order, one for every different instantiation of comprehension at each order k, for $\phi = \ulcorner x = \varnothing \urcorner$. I do not set restrictions to which levels of quantification occur in ϕ, so we can define a second order concept using a fourth level ϕ, as long as ϕ contains first level free variables, and we abstract away only from those.

I add, for any level, the extension of the usual axioms of first order logic for the behaviour of quantifiers and identity (universal instantiation, Leibniz Law etc).

Identity is primitive and applies cross level: $X^n = Y^m$ is always well-formed, for any n and m. Where $n \neq m$, the system is silent on whether $X^n = Y^m$.[9] For many proofs we also need an extensionality principle, for any n:

$$X^n = Y^n \leftrightarrow \forall Z^{n+1}(Z^{n+1}(X^n) \leftrightarrow Z^{n+1}(Y^n)) \qquad \text{(EXTn)}$$

Another principle needed is a very strong form of the axiom of choice, for any n. It can be written in many ways. I adapt mine from Shapiro 1991, 3.2. Where F^n is a meta-variable for functions of order n (which I define in higher order logic) and $Y_1, .., Y_m$ and Z are variables of at most level $n-1$, the following is an axiom-schema:

$$\forall X^n \exists Z(X^n(Y_1, .., Y_m, Z) \rightarrow \exists F^n \forall Y_1, .., Y_m(X^n(Y_1, .., Y_m, F(Y_1, .., Y_m)))) \qquad \text{(CHn)}$$

The axiom of choice of ZFCU becomes redundant, if we add CH2.

Finally, we add the following axiom that states that the urelements form a set:

$$\exists x(\text{Set}(x) \wedge \forall y(\neg\text{Set}(y) \rightarrow y \in x)) \qquad \text{(URS)}$$

I will explain in a moment why we need this axiom. I call the result of adding to ZFCU, for any n, CPn, EXTn and CHn, URS, as well as all the axioms for the behaviour of quantifiers and identity 'Higher Order ZFCU', or HZFCU for short.

The axiom of choice in ZFC implies that any set can be well-ordered. CH2 in pure 2ZFC is equivalent to the axiom of global choice: we can well-order any property of sets, as well. In order to extend induction to arbitrary orders, we need to well-order them. Without URS, however, second-order choice does not imply second-order well-ordering (Rubin and Rubin 1963, II.2). Hence the need to assume URS. We do not stop there: CH3 in HZFCU implies that we can well-order the universe of concepts of order 2 which, by Cantorian reasoning, is strictly bigger than the universe (of objects): we well-order it using well-ordered concepts of strictly lower orders, which might outstrip objects. The choice schema allows us to extend mathematical induction to arbitrary orders: it allows us, given some order n, to apply induction on concepts of order n because we can well-order them.

2ZFC is an interesting theory. It is quasi-categorical: two models of the theory are either isomorphic or one is isomorphic to a proper initial segment of the

[9] I choose a primitive identity because it makes some definitions easier and straightforward. I do not expect this choice to be compulsory: we could define identity as indiscernibility, and obtain a same-order concept of identity. The definitions need to be handled carefully to avoid talk of cross order identity. We might have to introduce primitive variables for functions. Similar theorems can be proved in the new system.

other.[10] 2ZFCU is not quasi-categorical. However, McGee proved in 2ZFCU+ URS the following 'completeness principle': 'the pure sets are not isomorphic to a proper initial segment of the "pure sets" of any model of second order ZFCU' (McGee 1997, 54).[11] The completeness principle, paired with the quasi-categoricity of pure second order ZFC, shows that in this stronger theory (i.e. 2ZFCU plus the assumption that urelements form a set) we can prove the following 'categoricity theorem': 'Any two models of 2ZFCU+ "urelements form a set" with the same universe of discourse have isomorphic pure sets' (1997, 55).

5.2.1.1 The higher orderist's framework

In the scope of this paper we adopt a 'higher orderist mindset'. The higher orderist resists the attempt to read higher order theories through some first order theory. In their mind, higher order quantification cannot be understood as a form of first order quantification over some specific kind of objects, like sets or classes. More generally, the higher orderist believes that quantification of level n cannot be paraphrased adequately as quantification of any level strictly lower than n. On the contrary, higher order quantification should be understood directly, as *sui generis*. How exactly this is going to pan out depends on one's semantic interpretation of the range of higher order quantification. Two popular options are Fregean concepts and plural talk.[12]

In the following, in the informal explanation I use a Fregean terminology: I use the term 'concept' to qualify what we range over with higher order quantification. The term is a translation of the German *Begriff*, and it is widely used in discussions on Frege and higher order logic. The term is not meant to be an explanation of the range of higher order quantification, if not a circular one: Frege defines a concept as the value of a grammatical predicate. The term 'concept' is used for many other things: in philosophy of mind, for example, it is often used for mental representations, mental files and whatnot. This is not the reading I intend, of course. Concepts in my system are extensional. Thus, everything I say is arguably compatible with a plural reading of higher order logic, in the style of Boolos, even though there might be further philosophical difference between the two approaches. It is also compatible

[10] This was proven by Zermelo for pure 2ZFC. For a discussion, see Walsh and Button 2018, 8.3.

[11] By 'model' McGee essentially means what we would call here a standard higher order interpretation. However, McGee does not wish to make use of third order resources to define satisfaction under a model, so he has to be 'devious', as he puts it: his notion of consequence does not extend to arguments with infinite premises (McGee 1997, 50).

[12] Williamson 2003 discusses the first option, even though he does not embrace it. He thinks that Fregean concepts are unsuitable for higher order modal logic because of their extensionality. For a classic on the plural option, see Boolos 1984.

with more exotic readings of higher order logic, where we take the values of higher level variables to be properties or universals, as long as the latter are not reducible to objects.[13]

Under the intended reading of HZFCU, the higher orderist insists that quantification at order 1 ranges over any object whatsoever, without any restriction: it is 'absolutely general'.[14] Sets are objects, so any set is found among the objects, at order 1. Cantorian and Russellian arguments show that, in any given second order system with a sufficiently strong comprehension principle, there are strictly more values of the second level variables than values of the first level variables. This result is usually interpreted in two radically different ways. For Dummett (1991) and other skeptics of absolute generality, this result is a *reductio* of the assumption that we were, in fact, quantifying over every object whatsoever in the first place. Since there must be more objects that work as values of second level variables than objects that work as values of first level variables, when we shift to second order talk we must have extended our initial interpretation of the quantifiers. From there we can always shift again: the interpretation we start from is indefinitely extensible, and there is no end to the process. The higher orderist's take is different: since they do not give up absolute generality, that there are more values of second level variables than values of first level variables convinces them that the values of second level variables cannot be objects, but something else: what we call 'concepts'. There are strictly more concepts than objects, and strictly more concepts of order $n+1$ than concepts of order n, given any n.

For the higher orderist, it is misleading to say that concepts are the collection of things that satisfy them, unless we specify carefully what we mean by 'collections'. If a collection is an object, then the statement is false as there are more concepts than objects. Also, some object fails to define a collection: there is no universal collection, but there is a universal concept. This is so as long as we take collections to be (singular) objects: they might be sets, proper classes, etc.[15] Second order quantification does not range over the power-set of a set, because there is no set

[13]My meta-theory also has extensionality, which I use in some proofs. However, extensionality can be dropped and similar results can be proven with different definitions, provided we have some way of identifying concepts without relying on their extension: for example, if we have a primitive notion of identity.

[14]A classic on absolute generality is Williamson 2003. For a top-notch collection on absolute generality, see Rayo and Uzquiano 2006.

[15]One could have a higher order reading of proper-class theory, or a plural reading of proper-class talk: see for example Uzquiano 2003. Here, on the other hand, I am discussing a view where we take the language of proper class theory at face value, as a first order, singular language.

of everything, and first order quantification ranges over everything; what is more, elements of a power-set are things, and second order quantification does not range over things. In fact, we might apologise for the use of 'Something' to talk about concepts, for it is not clear that 'Something' can quantify in predicate position. There are some examples, like 'John became something he despises', but they are controversial. We can absolve ourselves here, because we are using English not as a purified formal language but rather as a ladder, as a way to gesture at what we mean.

The syntactic theory of higher level languages is spelled out through a syntactic hierarchy: there are levels of variables, which induce a level in predicates and a level in languages. On the other hand, the meta-theory defines a hierarchy of concepts, confined to specific orders. These two hierarchies are essentially different and should not be confused: one is about syntax, the other is about the world and how it is structured. As the reader might have noticed, I use the term 'order' for the order of concepts and notions related to this hierarchy (theories, quantification etc), and 'level' for the syntactic notion, which applies to variables, languages, predicates etc.

5.2.1.2 Why cumulative?

In a cumulative theory, concepts of order $n \geq 2$ can be meaningfully applied to concepts (or objects) of order strictly lower than n. Usually, higher order logic is spelled out in a non-cumulative language. A non-cumulative language is one where, given a variable or constant X of level m and a string of variables and constants $Y_1, .., Y_n$, the string $X(Y_1, .., Y_n)$ is well formed just in case each $Y_1, ..Y_n$ is of level $m - 1$. In this paper I use a cumulative theory. I don't place too much weight on this choice, but there are some reasons to prefer cumulativity over non-cumulativity.

A cumulative theory is very plausible and natural: satisfaction in a higher order semantics is conceived of as a relation between a first order, syntactic object (a formula) and two higher order concepts (an interpretation and an assignment), so it is natural to spell it out in a cumulative system. Informal examples of higher level sentences also corroborate cumulativity: in 'Socrates became famous', 'became' looks like a relation between Socrates (a first order object) and the second order concept of being famous; similarly, in 'It is true of Socrates that he is Greek', 'true of' looks like a relation between Socrates and the concept of being Greek. Interestingly, it seems that 'became' can also take different types of arguments in its second argument place, like in 'Socrates became Plato'.

Arguably, cumulativity is not needed to model such examples: they can be mimicked in a non-cumulative framework, albeit in a less natural way. Yet, maybe redun-

dancy is needed to ensure generality. The system at hand aims at being as general as possible, at any order. It might be that anything a theory can say about Socrates can be mimicked by talking about some set and its relations with other things. Socrates is then redundant, but first order quantifiers should range over him regardless, otherwise they would not be absolutely general. Similarly, (finite) cumulative concepts of order n might be redundant; nonetheless, we should have n-th order quantification range over them, if we wish to truly say that n-th order quantification is general.

From a proof-theoretic point of view, it does make a difference if we take a cumulative or a non-cumulative meta-theory. Degen and Johannsen 2000, 150 show how a cumulative theory of finite order $n > 3$ equipped with an extensionality principle does not allow cut-elimination, in contrast to its non-cumulative restriction which does allow cut-elimination. If we venture into theories of ordinal order, as far as I know it is an open question whether a cumulative theory (without extensionality principles) of ordinal level ω is conservative over its non-cumulative restriction.[16]

5.2.1.3 Plurals and higher order logic

Even though all the philosophers who defend a higher order theory of consequence tend to agree on semantic ascent and use a similar framework, they often use different meta-theories. The ideas behind these systems are already in Tarski (1933). Tarski already showed how to define satisfaction for an n-th level language in a meta-language with one level of predication higher. Tarski would use a notion of satisfaction under a variable assignment, not of satisfaction under an interpretation of the non-logical vocabulary and assignment. However, as Tarski (1936a) suggests, we obtain the same notion of validity if we swap uniformly each predicate letter with a variable of an appropriate level and each constant with a first level variable, and then evaluate the resulting argument.

The story behind the higher orderist's framework in recent years often does not mention Tarski, but draws more from Boolos (1985), who discusses a related notion of super-validity. However, Boolos' super-validity applies only to logical truths and does not generalise to arguments with infinite premises (since higher order logic is not compact under Boolos' semantics). Shapiro (1991) tries out a model-theory spelled out in a proper-class theory, however this system is not higher orderist unless we embrace a higher order interpretation of proper-classes.

[16]To the best of my knowledge, the state of the art on the matter is in Degen and Johannsen 2000, where the authors admit that they cannot answer the question (see page 149).

Rayo and Uzquiano (1999) offer an explicit, higher order definition of validity that generalises to arguments. They show how to build a semantics for second order set theory either implicitly, by adding a primitively axiomatised, second level satisfaction predicate, or explicitly via third order quantification. What's peculiar is that they claim to be extending Boolos' account, which however makes use of plural logic, while they use higher order logic. For them, this is but a notational squabble since, following Boolos, they analyse (monadic) second order logic as plural logic. Rayo and Williamson (2003) use the same meta-theory to define validity for first order logic. In a similar spirit, Rayo (2006) extends higher order semantics to non-cumulative languages of arbitrary finite levels, this time by making explicit use of 'higher plural' logic. In higher plural logic, the orders are not orders of predication but orders of 'pluralisation': at the second order we find pluralities, at the third order pluralities of pluralities, and so on. A higher order semantics has been extended to second order modal logic in Williamson 2013, 237-9. Williamson however helps himself to a categorial semantics in the style of Montague (1970), rejects the reduction of higher order logic to plural logic, and spells out his system one order higher: he takes interpretations to be third order concepts and satisfaction to be fourth order (2003, 2013).

It does not make too much of a difference if, for our semantic theorising, we use higher order logic or some 'higher plural' logic equipped with a theory of n-tuples of pluralities of arbitrary orders, like the one in Rayo 2006. From a technical standpoint, the theories are at least as much alike as we need them to be: an n-tuple of pluralities of order n can be interpreted as a plurality of order n, and any n-ary predicate of level n can be interpreted as a plurality of order $n+1$ of such n-tuples. However, plural logic and higher order logic have interesting philosophical differences. In higher order logic, the order of quantification is an order of predication, while in higher plural logic it is an order of 'pluralisation'. The two orders look distinct, and care needs to be applied not to confuse the two. One can doubt some axioms of plural logic without doubting their counterparts in second order logic, and vice-versa. Pluralities look very extensional; it is not unreasonable to argue that concepts are not extensional, albeit they are in the present system. Williamson (2003) argues that the extensionality of plurals suggests that they are rigid: if x is one of the yy, then x is necessarily one of the yy. Yet, the same does not seem to apply to concepts: 'Socrates is necessarily snub-nosed' does not follow from 'Socrates is snub-nosed'. Thus, while we might be able to analyse higher order logic as higher plural logic in extensional languages, the analysis fails in intensional languages.

It is also not clear whether one can iterate orders of pluralisation at all; some think that the idea of a super-plurality is semantically questionable – Lewis (1991, 70) is an old example, Ben-Yami (2013) a more recent one. Others think that super-pluralities are legitimate, and even present in English (see for example Linnebo and Nicolas 2008). If super-plurals are present in English, then the notion of super-plurality is hardly meaningless, unless English is meaningless. On the other hand, not many have questioned iteration in the order of predication. Thus, the two logical systems might not have the same level of plausibility, especially if we assume that there is an order in the hierarchy for any natural number, or even for any ordinal.

5.2.2 A sketch of standard higher order semantics and its virtues

In the appendix, I develop a standard higher order theory of validity for arbitrary, non-cumulative languages of finite order. The system is able to interpret any subset of higher order logic capped to some specific n, and cannot be applied to the whole meta-language.

The range of the quantifiers varies across interpretations. If it did not, cardinality claims about the size of each order would come out logically true, like 'There are at least n objects', for any $n \geq 2$. An interpretation chooses a second order, non-empty concept X, which represents the range of values of first level variables and acts as the domain of the interpretation. Since the domain X is a concept, the objects of the universe of an interpretation are not *elements* of the domain, but rather the domain *applies* to them, and they *instantiate* the domain: we are trading philosophical accuracy, the price is grammatical ugliness.

The range of values of expressions of higher levels is inherited from the range of the expressions of the levels below:

- Values of first level expressions: all the x that $X(x)$;
- Values of second level expressions: second order concepts that apply at most to the values of first level expressions;
- ....
- Values of n-th level items: n-th order concepts that apply at most to the values of expressions of level $m < n$.

Given the lack of cross order generality, one cannot define in general what it means to be an assignment of any order, and what it means in general to be valid or true, given a language of some level or other. One can only, given an order n, define assignments, validity and truth at that order for languages of levels strictly lower than n. Logical validity for a language L of level k is defined at order $k+1$; in particular, assignments need to be of order k and satisfaction is now a concept at order $k+1$. If we added either quantification of level $k+1$ or a predicate of level k to L, then the entire machinery would need to be redefined one order higher: assignments (and interpretations, if we have predicates of level k) would be concepts of order $k+1$ and satisfaction a concept of order $k+2$.

The semantic system has interesting properties, which I prove in detail in the appendix but I sketch here. Given any n, call the cardinality of concepts that inhabit that order 'order-n-many'. We can define this more precisely in a theory of higher order functions, where X has cardinality order-n-many exactly when there is a bijection between the concept/things that X and the order n. In the system, for any n-th level variable there are exactly order-n-many assignments. In contrast, in standard set-based semantics, for any variable of level n, there are at most as many assignments as there are sets; that is, at most order-1-many of them. Also, since comprehension is unrestricted, a variable V^n of level n can be interpreted as any n-th order condition ϕ. By that I mean that there is an interpretation under which, for any X^m where $m < n$, the value of V^n applies to X^m exactly when $\phi(X^m)$. In set-based semantics, on the other hand, the above happens only if there are at most as many X^m that ϕ that they can all be collected in a set.

Using comprehension principles, we can prove by induction on complexity of formulae that there is a higher order interpretation I under which truth under I satisfies the axioms of Tarski (1933) for the language of 2ZFCU. The axioms include the compositional axioms for connectives and quantifiers, and a T-schema for atomic formulae. The relevant interpretation I is the one where the first order domain applies to everything, where, for all x and y, x is an element of y exactly when '$\in$' is satisfied by x and y under I and where, for all x, x satisfies 'Set' under I exactly when x is a set. We can show that a T-schema holds for satisfaction under I, for all sentences of second order ZFCU.

Even more generally, we can prove that we can develop a truth-theory for any subset of the language of HZFCU capped to level k. The relevant interpretation is again the same I as above. Call the 'projected domain' up to n of an interpretation

the range of the quantifiers at each order up to n generated by the choice of first order domain of the interpretation. More precisely, given a second order concept D:

- First order domain: D
- Second order domain: all the X^2 that apply to sequences of objects that D.
- ....
- n-th order domain: all the X^n that apply to sequences of concepts $Y_1, .., Y_m$ of order $j_1, .., j_m$ up to $n-1$ such that that Y_1 is in the j_1-th projected domain of D, .., Y_m is in the j_m-th projected domain of D.[17]

Since the first order domain of I is the universal concept, the projected domain up to n of I is the intended one for any order up to n. That is, in I, the following holds:

- Range of first order quantification: objects;
- Range of second order quantification (of arity m): second order concepts of arity m;
- ....
- Range of n-th order quantification (of arity m): n-th order concepts of arity m.

5.2.2.1 Semantic ascent

In standard higher order semantics the possible values of a syntactic element are determined via the 1-1 translation between levels and orders: the possible values of a first level expression are first order objects, of a second level, monadic expression second order, monadic concepts, and so on. It follows that SA2 holds. Pick any variable of level $m \leq n$ and arity k and any concept $X^{m,k}$: via the definition of assignment and comprehension of order m, one can prove that there is an assignment where the value of the variable is $X^{m,k}$.

To show SA1, in the appendix we use a familiar Cantorian argument. In set theory Cantor's argument shows that there are strictly more subsets of a set than elements. The theorem is proven by contradiction, by assuming that there is a surjective function from a set to its power-set. The structure of Cantor's theorem is very general and can be replicated elsewhere. In particular, in higher order logic we can show that,

[17] The technical definition-schema is in the appendix, and it requires a theory of finite higher order sequences.

given some order n in the hierarchy, there are strictly more concepts one order higher than concepts of order n (or objects, if $n = 1$). To be exact, what is actually provable is not: 'At any order n, there are more concepts of order $n+1$ than concepts of order n' because quantification across all orders is impossible. Rather we can prove the schema 'There are more concepts of order $n+1$ than concepts of order n', given any n. I call the schema 'Higher Cantor', and its instance at order n 'n-th order Cantor'.

SA1 holds. Take a language with quantification restricted to at most level k and with free variables V^{k+1} (or, equivalently, with predicates of level k): this is a language of bare level k. Following the instructions I sketched in the preceding subsection, we can build a higher order semantics by defining assignments of order $k+1$ and satisfaction at order $k+2$. By the lights of that semantics, there are at least as many values of V^{k+1} as there are concepts of order $k+1$. Since there are strictly more concepts of order $k+1$ than concepts of order k, there are strictly more possible semantic values of V^{k+1} than concepts of order k. Thus, it is impossible to define a strictly adequate semantics for L in a theory of order k.[18]

5.3 Non-standard and raising interpretations

In the next section I present some non-standard interpretations of higher order logic. I first sketch a semantic framework that justifies them. I discuss their features, and compare them to standard interpretations. I finally define a new notion of validity where these non-standard interpretations are included, and compare it to the standard notion of validity.

5.3.1 Semantic structure

A semantic theory is a system that assigns extensions to each expression of the interpreted language. The way we assign extensions follows a certain pattern. Each expression has a semantic role, which determines its type and its semantic relationship with other expressions: the value of a formula must be able to play the role of

[18] In (Florio and Linnebo 2021, 3.A) there is an extensive discussion of alternative ways of proving this Cantorian result in the scope of plural logic. They note that, by coding the language of functions between order $k+1$ and order k in the language of order $k+1$, k-th order Cantor can be proven using only predicative instances of comprehension. In this case impredicative comprehension if not used, might be presupposed: to show that the language of functions between k and $k+1$ orders can be coded using sequences, one needs impredicative comprehension – without impredicativity, the theorem does not really say what we take it to say. Uzquiano (2015, 9) makes a very similar remark, when discussing a similar result in the scope of proper-class theory. However, by proving the theorem using a language of primitive functions from order k to order $k+1$, one can prove that there is no surjective function using only predicative instances of comprehension and no coding.

a truth-value; the value of a (singular) nominal, on the other hand, must be able to play the role of an object, and must be able to be applied to the values of second level expressions to produce a truth-value; the value of a second level expression, in turn, must be able to play the role of a second order concept: we must be able to apply it only to the values of first level expressions to produce a truth-value, and so on. The mapping between syntax and meanings must respect the following structure:

- formula: truth-value-role
- constants, first level variables: object-role
- first level predicate letters, second level variables: second-order-concept-role
- ....
- predicates of level $n-1$, n-th level variables: n-th-order-concept-role.[19]

In general, when we say 'the X of an interpretation' we mean what plays the role of an X under the interpretation: the 'second order concepts of an interpretation' are whatever plays the second-order-concept-role under the interpretation: i.e. the values of second level expressions under the interpretation.

In set-based semantics for higher level languages, we map the syntactic hierarchy of expressions to a hierarchy of sets. In higher order semantics, on the other hand, we map the syntactic hierarchy of expressions to a hierarchy of concepts. In standard higher order semantics, there is only one acceptable way to link the two hierarchies: the one recursively generated by matching first level expressions with (possibly a sub-concept of) objects. Only objects play the role of objects, only second order concepts play the role of second order concepts, etc. To play the role of an X in standard higher order semantics *is* to be an X. Call this 'the standard match'. I call a semantic rule 'non-standard' if it does not follow the standard match, and an interpretation 'non-standard' if it is generated by a non-standard higher order semantic rule. In non-standard interpretations to play the role of an X and being an X may diverge.

Most non-standard rules are not interesting and fail simple 'acceptability' requirements. When we start thinking about logical validity, we never start from scratch: we often have in mind some arguments that we take to be logically valid. In the case of higher order logic, we have some axioms that we take to be logically true: the comprehension schemata, the rules for the behaviour of quantifiers and identity,

[19] I am ignoring the arity of expressions.

etc. Given an argument that looks intuitively valid, an interpretation that makes its premises true but the conclusion false fails what we might call the 'plausibility constraint'.[20] Many non-standard matches generate interpretations that are simply implausible, in this sense.

Yet, I find some non-standard interpretations philosophically interesting. I call them 'raising interpretations'. A raising interpretation works pretty much like a standard interpretation, if not for the choice of first order domain. In raising interpretations we keep the way expressions are recursively matched given a choice of first order domain, yet we relax the requirement that the first order domain be chosen among (monadic) second order concepts. Consider a rule that chooses second order concepts to be the possible values of first level variables, third order (non-cumulative) concepts to be the possible values of second level variables, and so on. The interpretations generated following this rule disregard objects: they 'clip them out' from the hierarchy, so to speak. I call such a rule a '+1 raising rule' and the generated, non-standard interpretations '+1 raising interpretations'. I call 'raising interpretation and rule' an interpretation and rule that uniformly raises the values of expressions by some n or other.[21]

Suppose we are interpreting the language of 2ZFC through +1 raising interpretations. The domain of an interpretation is now a non-cumulative third order concept, since the 'objects' of an interpretation are second order concepts. We need to extend the notion of projected domain to +1 raising interpretations. Given a domain X^3, the projected domain of X^3 is defined as follows:

- First order domain: X^3
- Second order domain: all the Y^3 that apply to sequences of X^2 that X^3.
- ....
- n-th order domain: all the X^{n+1} that apply to sequences of concepts $Y_1, .., Y_m$ of order $j_1, .., j_m$ between 2 and n such that that Y_1 is in the projected domain

[20] Sometimes, it is important to shift focus on interpretation-generating rules rather than interpretations, because it is the rules that might be too restrictive. For example, consider a rule that mandates that the first order domain be finite. Every interpretation that the rule generates is plausible *per se*; it is the rule that fails the plausibility constraint, as there are many formulae that are not logical falsities but are satisfiable only in infinite models.

[21] The details about raising interpretations are in the appendix. As I discuss in the appendix, we are not able to say in general what it means to be a $+n$ interpretation, for any n, but only what it is to be a $+n$ interpretation where $n \leq k$, given some specific k.

of X^3 of order $j_1 - 1$, .., Y_m is in the projected domain of X^3 of order $j_m - 1$.[22]

Given a domain X^3, a +1 raising interpretation I^{+1} based on X^3 maps '$\in$' and 'Set$'$ to a binary and monadic third order concept in the second order projected domain of X^3, respectively. A +1 assignment A^{+1} maps any v to some X^2 such that $X^3(X^2)$ and any V^2 to third order concepts in the second order projected domain of the interpretation: to third order concepts that apply only to (sequences of) second order concepts that X^3. Satisfaction under an interpretation and assignment can be defined at order 4 with clauses that are similar to standard higher order semantics, just shifted one order higher.

5.3.2 Raising interpretations are models of higher order logic

We can check that all raising interpretations for a language of level n are models of higher order logic up to order n (theorem 4 in the appendix). Consider +1 raising interpretations. They satisfy the axioms for the behaviour of quantifiers and connectives because of how satisfaction is defined. The cardinality relations between different orders are preserved under the interpretations: for any n up to m, there are more values of expressions at level n than values of expressions at level $n + 1$: any instance of Higher Cantor up to m is preserved. Comprehension principles are true under those interpretations, as well. We can show that, for any instance ϕ of the comprehension principle of order $m \leq n$, we can find a matching instance ψ of comprehension at order $m + 1$ such that satisfaction of ϕ in a +1 interpretation is a consequence of ψ. Raising interpretations satisfy the plausibility constraint sketched above.

5.3.3 In raising interpretations the role of an expression is exclusive

In standard higher order semantics, something cannot be both the value of a nominal and the value of a predicate; it cannot both play the role of an object and the role of a concept. In a famous and controversial response to Benno Kerry's articles (1886, 1887), Frege (1892) explains the rationale for this exclusivity. Kerry claimed that the concept Horse can both have the role of a concept and an object. He gives the example of 'Socrates is a horse' and 'The concept horse is a concept easily attained'.

[22]The technical definition-schema is in the appendix, and it requires a theory of finite higher order sequences to be finitely definable for all arities of some given level.

That a concept can play both roles is surely an interesting thought, and some features of English might persuade us of this fact. We can transform any adjective into a noun by suffixing it with '-ness' or other similar morphemes. Frege however thinks that the surface structure of the informal translation is deeply misleading. He claims that, when we nominalise an attribute, we are not talking about the concept anymore, and therefore 'The concept *Horse* is a concept' is false (emphasised by author). Even when we successfully predicate something of the concept Horse, the concept is still playing the role of a concept i.e. it is still the value of a predicate: he makes the example of 'There is a horse', where 'A horse' still works as a predicate, since 'There is' is a predicate of predicates, according his analysis. Frege's example is unfortunate, since 'A horse' is still a common noun, and it can be used in nominal position, for example in 'A horse is mortal'. A better example can be given with adjectives: in 'Everything is human', 'is human' is surely not a nominal, and in any sentence it can only work as a predicate: 'Human is mortal' is ill-formed. So, the concept Human – the value of 'is human' – cannot possibly play the role of both an object and a concept. That is, being the value of a predicate it cannot be the value of a nominal, under the same interpretation.

The language of higher order logic is 'exclusive' in this sense: a concept never belongs to different orders. Thus, that something cannot play two different semantic roles is mandated if we follow the standard match: a concept of order n is only a concept of order n, so it can only play the role of a concept of order n. If we relax the relation between playing the role of an X and being an X, exclusivity becomes a choice. For example, consider a rule that squeezes two orders in one level as follows: level 1 is mapped to order 1 and 2, level 2 is mapped to order 2 and 3, level n is mapped to order n and $n+1$, and so on. A second order concept under these 'braid interpretations' is playing the role of an object but also the role of a second order concept, so its role is not exclusive.

When it comes to raising interpretations, however, the role is exclusive, pretty much as in standard higher order semantics. Something can play only one role under any raising interpretation. Thus, object-role and concept-role are never blurred and always kept distinct. Similar arguments work for any other type of expression. So, it looks like the match generated by the raising rule does preserve the semantic roles of expressions, and it is not so obviously open to Frege's criticism against Kerry.

5.3.4 Semantic roles and intended readings

Under raising interpretations, the structure of the semantic roles sketched above is preserved. The value of a formula is always a truth-value, so it can play a truth-value-role. The value of a first level expression (a second order concept) can only play the object-role; that is, it can be applied to the values of second level expressions (third order concepts), and it is what we quantify over when quantifiers bind first level variables, under that interpretation. Similarly, the value of a second level expression can only play the second order concept role: it can only be applied to the values of first level expressions (second order concepts), and it is what we quantify over when quantifiers bind second level variables.

As already stated, from the perspective of raising semantics, to play the role of an X is not the same as being an X. Sure, second order concepts are the values of grammatical first level predicates, under the intended reading of the language: the standard match is the intended one. However, further possible non-standard readings correspond to raising matches. From the perspective of raising semantics, that objects are the actual values of nominals does not entail that objects are the only possible values of nominals.

One could argue that a concept cannot play the role of an object in virtue of what it is. Yet, this question is hardly to be put in isolation. A role is the function assumed or part played by something. Roles are structural, they are not intrinsic properties and cannot be judged in isolation: we need a framework first. I cannot absolutely play a role: I only play a role in some framework of rules. I do not need to be a king to play the role of a king in some framework, I only need to function as a king in that framework. 'Can I play the role of King in a game of chess?' The answer to this question will depend on how we decided to model the different pieces and moves of the game. My ability to play that role is not an intrinsic property that can be decided on the basis of my nature; rather, it is something structurally determined by a framework of rules and conventions which needs to be put forward in advance, before we can even start asking the question coherently.

Similarly, to ask 'Can X play the role of an object?' in isolation is senseless. We need a framework first. The semantic role of an expression is determined by the contribution to the meaning of the sentences in which it occurs; as the famous Fregean quote goes, 'Never ask for the meaning of a word in isolation, but only in the context of a proposition (1884a).' So, to answer the question: 'Can X play the role of an object?', we first need a framework to interpret formulae. To have such a framework we first need to settle questions like: How do we model truth-conditions?

How does predication work at each order in the system? What are we quantifying over? How do we provide values to predicates? Etc. Given the structural nature of roles, as long as a semantic framework can make sense of the semantic roles of the sub-sentential expressions taken together, the nature of the specific values that are assigned to a specific type of expression does not matter, nor does it make too much sense to compare the nature of the values across different semantic frameworks. So, when put in isolation, the question: 'Can a concept play the role of an object' is senseless. Yet, after we are given a framework, the answer is apparent but not very informative: 'No' in standard higher order semantics; 'Yes' in raising semantics.

To distinguish between playing the role of an X and being an X is common to other semantic frameworks. Say that we are interested in models of ZFCU; the language of ZFCU is meant to talk about urelements: the quantifier is meant to range over both sets and urelements, under its intended reading. Yet, there are perfectly legitimate models of ZFCU where the domain does not contain urelements but only sets. This is legitimate since sets can *play the role of* urelements. If we added special variables to the language for urelements, the intended values of these variables would be urelements, yet not all possible values of these variables would be urelements: in models where sets play the role of urelements, the values are sets.

When judged from a higher order perspective, the usual set-based semantics is another example where to play the role of an X is not the same as being an X. In set-based semantics, the values of expressions of any level of the object language are always of order 1, i.e. they are all objects. Yet, albeit objects, they can play the role of truth-values, concepts etc. This is possible due to the machinery of the semantic framework, via the definition of model and truth under a model. An object (a set) can play the role of a concept: it is the semantic value of a second level expression. Membership functions like a tool of predication, which links the values of nominals and the values of predicates together to form an assertion. Since ZFCU is well-founded, in any model the value of an expression of some level can only be predicated of the values of expressions of strictly lower levels, since the domain of each order follows the power-set relation. In this way, in the model the hierarchy of sets can play the role of the hierarchy of concepts, even though it is a hierarchy of objects and not a hierarchy of concepts. Similarly, 1 and 0 can play the role of truth-values even though they are not truth-values, as long as we understand the mapping between a formula, a model and 1 as saying that the formula is true under the model.

5.3.5 Cardinality of orders and reduction

In set theory a set plays the role of a second order concept; yet, there is always a second order concept of which the set is playing the role. So, from the perspective of standard higher order semantics, we can still understand how the set-based theorist is interpreting second order quantification as a special case of what we are doing: second order quantifiers in set-based semantics still range over concepts, albeit indirectly through their (set-sized) extensions. On the other hand, raising semantics often finds no such 'translation' between frameworks; since there are strictly more concepts of order 3 than concepts of order 2, at the level of the framework some concept of order 3 is playing a role no concept of order 2 can play. In this sense, raising semantics is a proper extension of standard higher order semantics.

That raising semantics is 'irreducible' to standard higher order semantics in this way should not surprise us, since we rejected the tenet that to play the semantic role of an X is to be an X, and since our meta-theory is a higher order theory of arbitrary finite orders, where we are able to quantify over orders superior to the nth, for any n. The result fits the concept of role: when one plays the role of something, the role is often generic. I can play the role of *a* king, I do not need to play the role of some specific king. An infinite number of things can play the role of kings, even though there are and will be only finitely many actual kings. Similarly, concepts can play the role of objects, which is not to say that any concept needs to play the role of some specific object.

Not only the intended reading of X is not a 'cap' to the cardinality of what can play the role of an X in the system as a whole: in raising semantics there might be more things that play the X-role *inside the very same interpretation* than Xs. So, there are raising interpretations under which second order quantification is irreducible to second order quantification under its intended reading. On the other hand, in standard higher order semantics for any X-role we are always able to inject the universe of concepts that plays the X-role into the universe of actual Xs: the 'intended' universe of Xs serves as a size cap to any interpretation. The lack of cap to the size of interpretations in raising models has some philosophical plausibility. I grant that sometimes we wish to keep the actual cardinality of Xs fixed, because it is relevant: for example, infinitely many things can play the role of truth-values, yet we insist that only two things can play that role in each interpretation. Arguably the reason is that we wish to keep a classical semantics where bivalence holds, and we think that a sentence being either true or false is something relevant to logical consequence. We do not show the same respect for the cardinality of objects, however:

we usually build models with domains of any cardinality. All that we seem to care about is the relation between cardinalities of different orders, not the identity of the cardinalities themselves. For example, it is a logical truth of second order logic that there are strictly more concepts than objects, and that there are strictly more concepts of order $n+1$ than concepts of order n, given any n.

Often we simply violate the requirement that the intended universe of Xs should serve as size cap to any acceptable interpretation. Consider intensional logics. In topic logic we need something that plays the role of locations in space, yet it seems wrong to say that, therefore, an interpretation is acceptable only if, for any x that is playing the role of a location under the interpretation, there is an actual location out there of which x can play the role. The particular interpretation we give to the modal operator in the context at hand does not seem to have any impact on the choice of things that can play the role of points in the structure. S4 in temporal logic is still the same S4 in epistemic logic, what changes is only how we understand the modal operators: as talking about instants in time or epistemic alternatives, respectively. The two semantic frameworks arguably have two different role-places: instant-role and epistemic-alternatives-role. Yet, the definition of a model does not change: W is still 'any set'. Yet, if the requirement above is sound, to make sure that what we are doing makes sense we should check whether the elements of any set can be mapped to genuine Xs: locations, epistemic alternatives etc. No one does that.

Here is yet another example: say we are interested in multi-modal epistemic logic. We aim to provide a semantic theory for a multi-modal language with denumerably many belief operators, which we informally read as 's_1 believes that', 's_2 believes that', etc. Normally, a semantics comprises a set of worlds and agents. Yet, there are only finitely many epistemic agents out there, so how can we provide a semantics for this language without having something playing the agent-role without being a corresponding actual epistemic agent? Shifting the discussion to what is metaphysically possible instead of what is actual seems to me a step back rather than a step forward. The modal logician is not assuming that possibly there are infinitely many epistemic agents, or that there are infinite things that are possibly epistemic agents; the empty set is not possibly an epistemic agent, and yet it is among the agents of an interpretation, since that is just any set. It seems wrong to assume that the modal logician's semantics for epistemic logic is relying on some modal metaphysical fact about epistemic agents. It seems to me that all we need to assume is that anything can *function as* an epistemic agent in our semantic framework.

5.3.6 Logical expressions

Our meta-language contains only one quantifier which binds variables of arbitrary levels. As such, it cannot be given an extension explicitly. However, interpretations only interpret some restriction of the meta-language capped to some fixed order n, so the 'image' of the extension of $\forall$ restricted to order n can always be explicitly interpreted as a concept of order $n+2$. To help us understand how quantification of different orders work, however, suppose we add level-bound quantifiers $\exists^n$ and $\forall^n$ for any n which can bind variables of arity 1 of a fixed level n only.[23] These quantifiers are strong candidates for logicality, provided that we trust some popular invariance criteria. For example, the value of any universal and existential quantifier of any level would be invariant across domains, under arbitrary isomorphisms of the universe. Sher (1991) shows it for second level quantifiers, but we can apply the test to quantifiers of arbitrary levels and get the same result.

We expect logical expressions to mean the same operation across all interpretations. What 'same operation' means depend on the class over which we look for sameness. Most commonly, by 'same operation' we do not mean that the same extension is given to the expression in every interpretation. That does not happen as, in some sense, logical notions receive different extensions given different domains.[24] Rather, as Sher puts it: 'These terms are fixed not in the sense that they denote the same entity in each model [...]. Rather, while non-logical terms are defined within models, logical terms are defined by fixed functions over models' (1996, 675). The details heavily depend on the semantic machinery we are working with.[25] Sher (1991) showed how we can use cardinalities of sets to define the relevant functions for logical quantifiers. Her theory is general and powerful. We do not have an explicit theory of higher order cardinalities, however, as our definition of 'order-n-many' is implicit; developing a general theory of higher order cardinalities is beyond the scope of this paper. However, if we restrict our attention to existential and universal quantification, we do not need cardinalities to express the relevant operation across domains.

[23] I focus on quantification over monadic concepts to simplify definitions. What I say can be easily extended to quantification over concepts of arbitrary arity, with the use of a theory of finite higher order sequences.

[24] See Sagi 2018 for a thoughtful discussion on how we should not think of logicality as preservation of extension across models.

[25] Following McGee (1996), the value of connectives can be thought of as an operation on sets of assignments. Roughly, the value of a formula in a model is identified with the set of assignments on the domain under which the formula is true; connectives are functions on the values of formulae. I do not focus on connectives in the following, but only on quantifiers.

In the usual set-based semantics, given a set x, first order existential quantification denotes the following function $f : \wp(x) \to \{1, 0\}$:

$$f(y) = \begin{cases} 0 & \text{if } y = \varnothing \\ 1 & \text{otherwise} \end{cases}$$

On the other hand, first order universal quantification given a set x denotes the following function $g : \wp(x) \to \{1, 0\}$:

$$g(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{otherwise} \end{cases}$$

The operations across domains that correspond to $\exists^1$ and $\forall^1$ are the functions that take any set to its characteristic function f and g, respectively. These functions are not set-sized. We can generalise the above to all levels. The operation $\exists^n$, for example, takes any set x to the following function $f : \wp^{n-times}(x) \to \{1, 0\}$:

$$f(y) = \begin{cases} 0 & \text{if } y = \varnothing \\ 1 & \text{otherwise} \end{cases}$$

The operation $\forall^n$ takes any set x to the following function $g : \wp^{n-times}(x) \to \{1, 0\}$:

$$f(y) = \begin{cases} 1 & \text{if } y = \wp^{(n-1)-times}(x) \\ 0 & \text{otherwise} \end{cases}$$

In standard higher order semantics, it is straightforward to check that the operations for the quantifiers are structurally the same. However, instead of working on sets, the (higher order) function ranges over concepts. Instead of the powerset relation, we use the notion of projected domain (for monadic concepts). I write $\mathrm{Dm}^n(X^n, D)$ to say that the monadic X^n is in the projected domain of D of order n, where D is any (monadic) second order concept. I use standard λ notation: $\lambda X^n.\phi$ is the concept that applies to all Y^n that ϕ.

$\exists^n$ is the operation such that, given some domain X^2, $\exists^n(X^2)$ is the following higher order function F whose domain are all the monadic Y^{n+1} that $\mathrm{Dm}^{n+1}(Y^{n+1}, X^2)$ and whose range is $\{1, 0\}$:

$$F(Y^{n+1}) = \begin{cases} 0 & \text{if } Y^{n+1} = \lambda X^n.X^n \neq X^n \\ 1 & \text{otherwise} \end{cases}$$

Similarly, $\forall^n(X^2)$ is the following higher order function G whose domain and range are the same as F:

$$G(Y^{n+1}) = \begin{cases} 1 & \text{if } Y^{n+1} = \lambda X^n.\mathrm{Dm}^n(X^n, X^2) \\ 0 & \text{otherwise} \end{cases}$$

Now, consider an interpretation that raises the values exactly by k. Its domain is a concept of order $k+2$. The operation for quantifiers needs to be updated accordingly, but structurally remains the same, since it is only the range of possible choices of first order domain that is extended. $\exists^n$ is the operation such that, given some X^{k+2}, $\exists^n(X^{k+2})$ is the following higher order function F whose domain are all the monadic Y^{k+n+1} such that $\mathrm{Dm}^{n+1}(Y^{k+n+1}, X^{k+2})$ and whose range is $\{1, 0\}$:

$$F(Y^{k+n+1}) = \begin{cases} 0 & \text{if } Y^{k+n+1} = \lambda X^{k+n+1}.X^{k+n+1} \neq X^{k+n+1} \\ 1 & \text{otherwise} \end{cases}$$

Similarly, $\forall^n(X^{k+2})$ is the following higher order function G whose domain and range are the same of F:

$$G(Y^{k+n+1}) = \begin{cases} 1 & \text{if } Y^{k+n+1} = \lambda X^{k+n}.\mathrm{Dm}^n(X^{k+n}, X^{k+2}) \\ 0 & \text{otherwise} \end{cases}$$

It is straightforward to check that these are in fact the operations associated with these quantifiers, using the definition of (raising) satisfaction in the appendix. Consider +1 interpretations defined for the language L^k of level k. We can check that, for any m between 1 and k, under each +1 interpretation $\exists^m$ is assigned the operation corresponding to existential quantification at order m. The operations that logical expressions should match are defined given a class of possible choices of what counts as an object under the interpretation: a class of domains. Given one such choice, the operation of existential quantification over order 1 takes any value of first level predicates (or second level variables) to True if they are satisfied by some object-performing thing, to False otherwise. In the semantic framework at hand, second order objects are what is playing the role of objects, so $\exists^1$ receives its appropriate semantic role and value of quantification of order 1, under the interpretation. A similar argument works given $+n$ interpretations of any n.[26]

[26]Here is a dilemma that was brought up to me. If identity is not treated as a logical expression, an interpretation I picks a relation E as the value of identity. Often, this relation 'plays the role of identity' in the model because it implies indiscernibility in the relevant language. Aren't these interpretations discarded in normal practice, when we treat identity as logical? How can we discard them, if we do not insist that the value of = must be identity, and not just 'play the role of identity' inside the model? I see two possibilities here, and neither of them is problematic. Either E has the same extension of identity in some standard model or not. If the former, then what we are doing is just describing one of the usual models in a weird way, because interpretations are defined extensionally. If the latter, then I agree that the model should be discarded. For identity is logical, so it should mean the same operation across interpretations: the operation that maps each sequence of an object-performing thing and itself to True and everything else to False. If E maps the value of v to the value of u even though these values are not the same object-performing thing, E is not describing the operation of identity, regardless of whether the value of v satisfies the same predicates as the value of u. Thus, either the interpretation in question is not a new interpretation at all and we were already 'counting it', or it is a new interpretation but we do have reasons to discard it.

We can check that the usual logical expressions are invariant in raising semantics. Invariance is always defined given a choice of what counts as an object: given a choice of first order domain. In the more popular tests like bijection invariance, the transformation is defined across different choices of first order domain, given some isomorphism class (Sher 1991). Different models have different first order domains with different cardinalities, and thus different transformations that matter to invariance. The value of the first level existential quantifier is still invariant because, *given a certain isomorphism class defined on the class of what can be a first order domain*, it is always invariant under each relevant transformation defined over that class.

In the set based semantics, first order domains are sets, and the test for the invariance of the values of logical expressions needs to be applied accordingly. In standard higher order semantics, first order domains are second order concepts, and the test needs to be updated accordingly, to be applied across equinumerous second order concepts. In raising interpretations, the choice of first order domains is extended further: if we raise the values by k orders, a first order domain can be any concept of any order between 2 and $k+1$. So, the class over which we apply the test needs to be extended accordingly, to include the new raising domains. In any case, for any level n, any value of $\exists^n$ is invariant under the respective classes of transformations.

5.3.6.1 Types and extensions of logical expressions in raising semantics

Where I is some standard higher order interpretation whose domain is X^2, call $\exists^1(X^2)$ the extension of $\exists^1$ in I. There is an intended standard higher order interpretation whose domain is $U = \lambda x.x = x$ and where $\exists^2(U)$ is the intended extension of $\exists^2$. Consider the $+1$ interpretation I^{+1} whose domain is $\lambda X^2.X^2 = X^2$: the extension of $\exists^1$ under I^{+1} is the extension of the second level $\exists^2$, under its intended reading. This is not possible in standard, higher order semantics. However, that extensions can vary is normal; the extension of quantifiers can change from interpretation because their extension depends on what their range is, and the range can vary.[27] Similarly, in set-based semantics the extension of $\exists^1$ in some model whose domain is x is the extension of $\exists^2$ in a model where the domain is $\wp(x)$. Nothing surprising: the two

[27] I am happy to grant that in a constant domain semantics we could argue that the extension of quantifiers should remain the same in all interpretations, if we insisted that the domain should always be the intended. My aim is to show that, when it comes to logical expressions, raising interpretations are not that dissimilar to set-based interpretations. If the domain ought to be the intended, set-based semantics is doomed as well, so the two semantic frameworks are on a par when dealing with this issue.

extensions belong to different frameworks, with different choices about what the values of nominals are (i.e. who is playing the object-role).

In +1 raising semantics the value of $\exists^1$ does not even have the same *type* of the value of $\exists^1$ under its intended reading. That is, the characteristic operation of $\exists^1$ is not some kind of (actual) first order quantification, but rather some kind of (actual) second order quantification. On the other hand, in standard higher order semantics this never happens. From the perspective of standard semantics, one could argue that $\exists^1$ is supposed to mean 'Some-*object* in the first order domain', not 'Something that is playing the object-role in the first order domain'. Similarly, when we take 'Some-human' to be a logical expression, we demand that it means 'Some human in the domain', not 'Something the interpretation thinks it is human in the domain'. On the contrary, the latter is exactly how we would treat 'Some-human' if it was *not* logical.

Membership of a type of operation is not respected in set-based semantics, either, so raising semantics and set-based semantics are similar, in this respect. For what happens if we apply the same reasoning to $\exists^2$ in set-based semantics? Since it is logical, $\exists^2$ in every set-based model should always mean 'Some second order concept in the second order domain' and not 'Something that is playing the second order concept role in the second order domain'. Yet, in set-based semantics, only objects (sets) play the role of concepts, so there are no concepts in the second order domain of any set-based interpretation, and therefore $\exists^2$ cannot mean 'Some concept in the second order domain', if not vacuously. From the perspective of the higher orderist, strictly speaking the value of $\exists^2$ in any set-based model is a type of first order quantification, not a type of second order quantification.

From the perspective of standard higher order semantics we can 'translate' what the set-based theorist is doing in the standard higher order framework, by interpreting their operation of second order quantification as ours restricted to set-sized domains. On the other hand, trying to understand what the standard higher order theorist is doing from the perspective of the set-based theorist is hardly possible, unless we leave the higher orderist's perspective that insists that higher order quantification cannot be reduced to quantification over sets. Similarly, from the raising perspective we can translate what the standard higher order theorist is doing in the raising higher order framework, by interpreting their operation of second order quantification as ours restricted to domains of cardinality order-1-many. Trying to understand what the raising higher order theorist is doing from the perspective of the standard higher order theorist is hardly possible, unless we abandon the requirement that to play

the role of an X in the framework is to be an actual X, and thus second order quantification needs to be some (restriction of) actual second order quantification.

The difference between 'Some-human' and 'Something' is that, in the first case, inside the semantic framework we are disregarding what is decided at the level of the interpretation: what the interpretation thinks humans are; in the second case, inside the semantic framework we are trying to disregard what is decided at the level of the semantic framework as a whole: what is the pool of possible values of first level expressions *in some interpretation or other*. While the first move might keep us inside what is meaningful in the semantic system, the second tries to 'jump out of it' by not leaving it, so to speak – something which is destined to lead to confusion.

5.4 Features of raising semantics

5.4.1 Henkin models vs raising interpretations

In the literature, Henkin models are the most commonly discussed interpretations of higher level languages that are non-standard. In set-based Henkin models, the range of higher order quantification is restricted to some possibly proper subset of the powerset of the domain. For long philosophers acted as if a Boolos-style, plural semantics was immune to 'Henkinisation', yet this is not the case, as Florio and Linnebo (2016) recently pointed out. Of course, Henkin-like models could be developed within our own higher order semantics, as well, by restricting the range to some concept that applies at most to sub-concepts of the first order domain. If we allowed Henkin models, higher order logic would lose much of its interesting meta-theoretic properties. It would be complete, compact, and Löwenheim-Skolemisable.

Many think that the categoricity of second order PA (2PA) makes it a better axiomatic foundation for arithmetic than its first order counterpart. In 2PA we are able to characterise the natural numbers up to isomorphism, and define many essential mathematical properties like well-foundedness and finitude. None of this is possible in first order PA. However, if we defined validity in 2PA through Henkin models, its expressive advantage would be in jeopardy, and many think this is reason enough to avoid it (Väänänen 2001). Similarly, second order ZFC is often seen as a better foundation for set-theory, for even though it is not a categorical theory, it is 'quasi-categorical', which many take to be enough.

Many claim that we have good reasons to believe that standard higher order semantics is the most natural and well-suited for interpreting higher level languages. Kreisel (1967) and Shapiro (1991, 5.3) suggest that we simply apply induction to

arbitrary functions, without checking whether they are first order definable, and the epistemic assurance in the induction schema is derived from the implicit trust we put in its higher order counterpart. Moreover, from the non-reductive perspective of the higher orderist one might claim that the question of whether the 'correct' interpretation is the Henkin or the standard cannot be put forward coherently. The meta-language we are using is a higher level language. Since we use it, it must be pre-understood, not understood through the semantics, standard or Henkin. Sure, when we develop a theory of logical validity, we do build interpretations of the meta-language, but this is not to say that it is through them that we understand it: rather, the tools we need to build interpretations presuppose our understanding of it.

The assumption that the meta-language is pre-interpreted rather than interpreted only through models is quite apparent in our higher order semantics, where an assignment of concepts of order at most n is a concept of order n; to interpret variables of level n as referring to concepts of order n, and thus to interpret quantifiers that bind variables of level n as quantifying over order n, we *use* quantification over order n. So, unless we pre-understood quantification over order n in a non-reductive way, our semantic machinery is thoroughly pointless. Similarly, questions like: 'Since both are models of the theory, are we speaking from a raising interpretation or a standard interpretation?' or: 'How can we be sure that by using $\forall^1$ we are quantifying over objects and not over concepts?' are bound to lead nowhere. It is like asking whether we really started counting from 1, or if rather we started counting from 3. To even talk of 3 and 1, we must have counted from 1 already.

I claim that the raising interpretations are more acceptable in the eye of the higher orderist and the non-reductionist. In the Henkin model we are artificially restricting the range of the higher level quantifier. In set-based talk, in some Henkin model a second level quantifier might not range over *all* the subsets of the domain; in non-reductive talk, in some Henkin model the second level quantifier might not range over *all* the sub-concepts of the first order domain. In fact, a higher orderist could argue that Henkin interpretations which restrict the second order domain are not acceptable; second order existential quantification is logical, and therefore should mean the same operation across all interpretations: the operation that maps each sub-concept of the first order domain to True if non-empty, to False otherwise. Yet, Henkin interpretations that restrict the second order domain fail to map some sub-concept of the first order domain to True.[28] In the raising interpretations this simply

[28]The Henkinist will complain, of course, and deny that this is the operation the quantifier should match. My point is not that the Henkinist is wrong: it is that even under this possibly unfair and

does not happen: we always quantify over the full projected order above. In every interpretation, the characteristic operation of the second order existential quantifier always maps to True all and only the non-empty sub-concepts of the first order domain. What might happen is rather that the full order above might not be the one we would be quantifying over, had we used the second level quantifier under its intended reading.

In raising semantics, higher order theories retain their expressive advantages, as well, and they are not multi-sorted first order theories in disguise. As I show in the appendix, higher order logic is still incomplete, non-compact, etc; second order PA is still categorical. In this respect, raising semantics is the same as standard semantics. However, raising semantics does carry its own meta-theoretic consequences, as the next sub-section points out.

5.4.2 Raising semantic ascent

In standard higher order semantics we can prove the following version of 'semantic ascent': a strictly adequate semantic theory for a language of bare level n cannot be given in a meta-theory of bare order n. A similar result can be stated from the perspective of raising semantics in regard to standard higher order semantics. It is easy to show that, given any n and m, where $m < n$, for any $+m$ interpretation there is an isomorphic $+n$ interpretation (lemma 5 in the appendix). Standard interpretations are +0 interpretations, so for any standard interpretation there is an isomorphic $+n$ interpretation, given any n. However, the opposite does not hold. In standard higher order semantics there are at most order-2-many possible values of second level variables; in a framework of raising interpretations, on the other hand, there are strictly more than order-2-many values of second level variables, whenever we are considering $+n$ interpretations where $n > 0$, so we are capturing more interpretations.

We are interpreting a language of level k. Consider the following definition-schema:

Definition 8 ($+n$ strict adequacy)**.** *Where V^n is a variable of level n, a theory T of validity for L^k is $+n$ strictly adequate just in case the following holds:*

- *Where $m \leq n+1$, for any X^m and V^1 there is an interpretation in T where the value of v is X^m;*
- *Where $m \leq n+2$, for any X^m and V^2 there is an interpretation in T where the value of V^2 is X^m;*

biased characterisation of what quantifiers should mean, raising interpretations pass the test while Henkin interpretations don't.

– ....

– *Where $m \leq n + k$, for any X^m and V^k there is an interpretation in T where the value of V^k is X^m.*

Given a $+n$ raising semantics, one can show that $+n$ interpretations satisfy $+n$ strict adequacy. On the other hand, standard interpretations do not satisfy $+n$ adequacy unless $n = 0$. The following version of semantic ascent can be shown:

(RSAn1) A $+n$ strictly adequate semantic theory for a language of bare level k cannot be given in a meta-theory of bare order $m \leq n + k$;

(RSAn2) A $+n$ strictly adequate semantic theory for a language of bare level k can be given in a meta-theory of bare order $n + k + 1$.

The positive result of RSAn2 is not particularly reassuring, because any semantic theory fails $+n$ strict adequacy for some n. If the interpretations of the semantic theory in question are at most $+m$ interpretations, the theory will fail $+n$ strict adequacy tests for any $n > m$. In raising semantics there is a phenomenon of indefinite raisability of higher order interpretations: irrespective of the order from which we choose to define validity, there are always more interpretations we are missing which are reachable from some strictly higher order. Consider even a first level language L. Suppose we develop a $+n$ raising semantics for L by defining interpretations as concepts of order $n + 2$ and satisfaction as a concept of order $n + 3$. However, specified as such, interpretations are unable to provide values of order $n+2$ to first level variables. We can show, though, that there are raising interpretations definable one order higher that do provide such values. Consequently, at no order we are able to say that we have gathered all the interpretations of L. One can make an analogy with indefinite extensibility: one starts from a set, builds up an interpretation for higher order logic, and realises that there is a bigger set available as a starting point (i.e. as a possible set that represents order 1). Here, one needs to start from an order, but one can always 'see' that are higher and higher orders available as possible starting points. Thus, given the axioms we provided for a logic of order n, there are raising higher order interpretations of cardinality order-m-many, for arbitrary $m > n$, which are models of the axioms provided: there is no cap to models of any subset of higher order logic.

Say that a semantic theory is raising-adequate if it is $+n$ strictly adequate for any n. Maybe no raising semantics is raising-adequate if we decide to define validity from

the standpoint of a specific order, but it might be if we consider all orders and all raising interpretations. If we are so keen to provide the most general theory of validity, would it not be possible to define validity for a language of level n as preservation of truth under any $+n$ raising interpretation, for any $n \geq 0$, rather than preservation of truth under any standard interpretation? After all, our meta-theory HZFCU itself is not capped to any finite order: we have all orders available, so why not use them all? The problem is that in any standard higher order system there is no way to quantify across all orders.[29] So, such a definition is impossible in the present system.

The problem of indefinite raisability is stubborn, and arguably it sticks even if we decide to extend orders to the transfinite. The old problem is solved: given a language of level n, we can quantify over all $+k$ raising interpretations, given any k. We do so by mimicking quantification at any order n at order ω. The new problem arises when we try to develop a theory of validity for languages of any ordinal level ω. As Linnebo and Rayo (2012) discuss, semantic ascent seems to hold even for infinite orders because a stronger version of Higher Cantor still holds. So, I suspect that infinite raising interpretations can be defined in the new system, with very similar philosophical consequences.[30]

5.4.3 Extensional equivalence?

Here is a dilemma: consider $+n$ validity, which is preservation of truth under all $+m$ raising interpretations, where $0 \leq m \leq n$. Standard higher order interpretations are among the raising ones (they are $+0$), so if something is false in some standard model it is false in some raising model: if something is not standardly valid, it is not raisingly valid. Yet, it is an open question whether something true in all standard models is also true in all raising models, and more in general whether raising validity and standard validity coincide for arbitrary arguments.

In the language of HZFCU it is an open question whether we get different results about logical validity using standard or raising semantics. We know that the answer is negative for Henkin semantics, because we have a completeness result. Here we do not: the logic is incomplete for raising validity. Higher order logic is not a categorical

[29] We also do not have one concept of satisfaction that works for any interpretation, of course, for it would need to apply to variable assignments of any order, and so it would not belong to any order. This is not an issue, though: we only need, for any interpretation, its respective concept of satisfaction, and then quantify over all interpretations and all their respective concepts of satisfaction. This is what we lack: quantification across all orders.

[30] Ascent is stopped if we insist that some fixed ordinal κ is the end point. Then Higher Cantor of course cannot be proved because it needs resources of order higher than κ. See Studd 2021 for a discussion.

theory, so the fact that we offer new models for it with different cardinalities does not help determine whether new validities come about when we consider 'bigger' raising models. Second order PA retains its categoricity under raising semantics, so we obtain only new isomorphic models in the new semantics. 2ZFCU is not categorical under the raising semantics, so again the situation does not change for raising models: we already had different models of it with different cardinalities, we are just adding on more of them. However, we do have some interesting differences when we consider the theory 2ZFCU+URS. As mentioned above, McGee (1997) showed how, under standard higher order semantics, any two models of 2ZFCU+URS have isomorphic pure sets, provided they share the same universe of discourse. However, McGee's proof fails for raising models, as I show in the appendix. That a proof fails does not mean that the result is not there: maybe it can be proven in some other way. I leave it as an open question, even though I suspect that the answer is negative.

While it is unclear whether raising and standard validity disagree on what is valid for the language of HZFCU, they might disagree in some extensions of it. Suppose we added *ad hoc* quantifiers that specify some cardinality by definition. A popular example by McGee 1992 extends the language with the quantifier $\exists^{AI}$, which informally means 'There are absolutely infinitely many'. The term is from Cantor: there are absolutely infinite ϕs when ϕ fails to define a set. So, the sets are absolutely infinite, as well as the ordinals (Cantor 1899). In higher order logic absolute infinity does not pin down one specific cardinality: the objects and the second order concepts are both absolutely infinite, even though there are strictly more second order concepts than objects. As discussed in the appendix, by using a theory of higher order functions we can pin down one cardinality by defining, given some n, the unique cardinality order-n-many. Something has cardinality order-n-many iff it is bijectable with order n. We can then add cardinality quantifiers $\exists^{O1}$, .., $\exists^{On}$ for any order. Unlike before, $\exists^{On}$ is not a quantifier of level n: it can bind variables of any level. We read $\exists^{On}v\phi$ as: 'There are at least order-n-many ϕs'. $\exists^{AI}$ is equivalent to $\exists^{O1}$. The following clause suggests itself in higher order semantics, standard or non-standard, where I is an interpretation:

Definition 9 ($\exists^{On}$)**.** *$\ulcorner\exists^{On}V^j\phi\urcorner$ is satisfied under an interpretation I iff there is a sub-concept X of the j-th order domain of I such that:*

1. *There is a surjection between X and order n;*

2. *For any Y such that $X(Y)$, Y satisfies $\ulcorner\phi\urcorner$ under I.*[31]

[31] Of course, the technical definition will require variable assignments.

The definition is a schema, given a j. Is this definition legitimate? It does not mention what it means to be order-n-many for the interpretation, nor does it check whether interpretations in some sense agree on what order-n-many means. By *fiat* it bypasses what any I 'thinks', by defining the cardinality directly. If our quest was to successfully determine what 'There are order-n-many things' means using the apparatus of models, the endeavour is particularly worthless. Fortunately, this is not our aim. We already know how to use these expressions: we can define them in our meta-language, which is pre-understood. By *fiat*, we are now able to categorically define order-n-many: any two models that make true $\ulcorner \exists^{On} V^j \phi \urcorner$ agree on the fact that the extension of ϕ is as big as order n. Standard and raising validity now diverge. For example, $\ulcorner \neg \exists^{O2} v\phi \urcorner$ is standardly logically valid but not raisingly valid. It is standardly valid because what plays the role of an object in a standard interpretation must be an object. So, we just cannot find an interpretation where there are more things that play the role of an object and satisfy ϕs than objects in the universe. On the other hand, in raising interpretations what plays the role of objects needs not be an object, so there is no cap to the cardinality of the first order domain in an interpretation: there can be more objects according to the interpretation than there are objects out there.

To assess the significance of the result we needs to understand what $\ulcorner \exists^{O2} v\phi \urcorner$ means, first. The definition makes $\exists^{O2}$ rigid: it always refers to the actual size of the second order domain. So, in any interpretation, what it says is not that there are strictly more objects than objects; rather, it says that there are strictly more objects (in the interpretation) than actual objects. The first claim is clearly logically false: in every interpretation, there are exactly as many object-performing things as there are object-performing things and no more. So, any interpretation makes true 'There is no surjection between ϕ and order 2', as long as we formalise the claim using only logical vocabulary. This is because we are asking to evaluate the claim 'internally', based on what happens inside the interpretation. On the other hand, that 'There are exactly as many objects as actual objects' is not logically true in raising semantics. It is false in an interpretation where there are more object-performing things than objects.

We can make an analogy with modal logic: while clearly 'There are as many dogs as there are dogs' is logically true, many would doubt that 'There are as many dogs as there are actual dogs' is logically true: it is not true if we are reading 'actually' rigidly (as it is intended here), if we check validity by looking at all worlds of every model, and if we are considering Kripke models where Barcan fails. The result is

independent of absolute generality: I can consistently claim that there are more things that can play the role of objects than objects, without entailing that somehow I am not considering all objects, because to play the role of an object is not to be an object, in raising semantics. To sum up, if $\ulcorner\neg\exists^{O2}v\phi\urcorner$ was logically true, it would mean that interpretations have somehow a cap on the size of their domain, that this cap is provided by the intended interpretation, and the rationale behind this cap is that, to play the role of an X, one must be an X, or at least each role-performing X must be playing the role of some specific X. When we lift the cap, $\ulcorner\neg\exists^{O2}v\phi\urcorner$ can be falsified.

Appendix A

Formal Notes on Higher Order Semantics and Non-standard Interpretations

A.1 Meta-language and meta-theory

A.1.1 Syntax

Our meta-language (L_{HZFCU}) is build from the language of ZFCU (L_{ZFCU}). The only predicates of the language are a unary predicate for sets 'Set' and two binary predicates '$\in$' and '='. There are no constants. The only connectives and quantifiers are $\forall$, $\wedge$, $\neg$. The other connectives and quantifiers can be defined via the usual meta-abbreviations.

To obtain L_{HZFCU}, to the syntax of L_{ZFCU} we add, for any $n > 1$ and m, denumerably many variables of level n and arity m, written $X^{n,m}$. The 'old' first level variables are replaced with variables of level 1 and arity 0. I denote variables by capital letters, with the possible exception of first level variables. We might omit superscripts where it is obvious from the context to which level or arity the variable belongs. I will often omit arity superscripts for readability.

An atomic formula of L_{HZFCU} is defined as follows: for any natural number m n, i and j, where $Y_1, .., Y_k$ is a string of n variables whose level is strictly smaller that m:

$$X^{1,0} \in Y^{1,0} \mid X^{n,i} = Y^{m,j} \mid \mathrm{Set}(X^{1,0}) \mid X^{m,n}(Y_1, .., Y_n)$$

Where ϕ and ϕ_{at} are meta-variables ranging over formulae and atomic formulae, respectively, the definition of well-formed formula is defined inductively for any n:

$$\phi_{at} \mid \neg\phi \mid \phi \wedge \psi \mid \forall X^{n,k}\phi$$

The identity predicate applies to any variable: it is cross-level. This makes some definitions simpler, for example the definition of functions whose domain and range extend across different orders.

For notational convenience, I express generalisations across a finite range of orders through the all-purpose meta-variables $\mathfrak{a}$, $\mathfrak{b}$, etc. which, in its first occurrence, is bound by an index which specifies the finite range of orders we are quantifying across. That is, we never write $\forall\mathfrak{a}\phi$, but rather $\forall\mathfrak{a}^{\leq n}\phi$, for some n. This trick lets me shorten disjunctions and conjunctions with a single formula. For example, $\forall\mathfrak{a}^{\leq n}\phi(\mathfrak{a})$ is short for the conjunction $\forall x\phi(x)\wedge...\wedge\forall X^n\phi(X^n)$. $\exists\mathfrak{a}^{\leq n}\phi(\mathfrak{a})$, on the other hand, abbreviates the disjunction $\exists x\phi(x)\vee...\vee\exists X^n\phi(X^n)$. We need to bound $\mathfrak{a}$ to some finite level, so that we do not exceed the limits of what is expressible in the language. Otherwise we would be able to write a schema as a single formula. For example, $\forall\mathfrak{a}\phi(\mathfrak{a})$ would express the following schema, where m is any natural number: $\forall X^m\phi(X^m)$.[1]

Finally, $\forall x_1,..,x_n$ is short for $\forall x_1..\forall x_n$, and $[X]_n$ is short for $X_1,..,X_n$. So, I abbreviate $\forall X_1,..,\forall X_n\phi$ as $\forall X_1,..,X_n\phi$, which I then abbreviate again as $\forall[X]_n\phi$. I assume that $[\mathfrak{a}^{<n}]$ can abbreviate a string of variables of different levels, up to $n-1$, and not necessarily a string of variables of the same level. So, $\exists[\mathfrak{a}^{\leq n}]_k\phi$ abbreviates a very long disjunction where every string of k variables of level up to n occurs. I use $\exists!X^{n,m}\phi$ as short for $\exists X^{n,m}\forall Y^{n,m}(\phi(Y^{n,m})\leftrightarrow X^{n,m}=Y^{n,m})$. Of course, I am using n,m,k etc. as meta-variables for natural numbers; natural numbers can be defined as usual in set-theory. I occasionally use $\forall(n\leq m)\phi$ as short for $\forall n(n\leq m\rightarrow\phi)$.

A.1.2 Meta-theory

I label the meta-theory 'higher order ZFCU', or HZFCU for short. The theory contains all axioms of second order ZFC with urelements (2ZFCU). The quantifiers in the usual axioms about sets are bound by the condition 'Set' when appropriate. For example, extensionality will look like this:

$$\forall x,y\big((\mathrm{Set}(x)\wedge\mathrm{Set}(y))\rightarrow(\forall z(z\in x\leftrightarrow z\in y)\rightarrow x=y)\big)$$

In second order ZFCU, the axiom schema of replacement is substituted with its second order universal closure. The axiom schema of separation becomes redundant, given the axioms of pure second order logic, so 2ZFCU is finitely axiomatisable.

[1]Some have made a case for developing a meta-theory of higher level languages in a untyped language to express cross-type generalisations, see for example Oliver and Smiley 2016, 15. I will stick with the orthodoxy, here, because it is the standard semantics I am discussing, and because it is not obvious that an unsorted language does not carry ideological weight. For example, it implies that we are able to quantify over all concepts of all orders, something which the higher-orderist usually denies.

The full theory we work on adds the axioms of higher order logic to 2ZFCU, and it is called HZFCU. To obtain HZFCU, we add the following comprehension schemata, for any n, where ϕ does not contain $X^{n,k}$:

$$\exists X^{n,k} \forall [\mathfrak{a}^{<n}]_k (\phi([\mathfrak{a}]_k) \leftrightarrow X^{n,k}([\mathfrak{a}]_k)) \tag{CPn}$$

Other principles needed are the extensions of the rules and axioms for the quantifiers of first order logic to variables of every level. Among those we have, given any n:

$$\forall X^n(\phi) \rightarrow \phi(Y^n/X^n)$$
$$\phi \rightarrow \forall X^n \phi, \text{ where } X^n \text{ is not free in } \phi$$
$$\forall X^n(\phi \rightarrow \psi) \rightarrow (\forall X^n \phi \rightarrow \forall X^n \psi)$$

Similar extensions are needed for the axioms of identity.

We also add extensionality, at any order n:

$$X^n = Y^n = \forall X^{n+1}(X^{n+1}(X^n) \leftrightarrow X^{n+1}(Y^n)) \tag{EXTn}$$

The system is silent on whether $X^n = Y^m$ when $n \neq m$. Extensionality is needed for some proofs.

To prove some theorems, we make use of a very strong choice schema, defined for any order. To write it down, we need a theory of functions. Higher order functions can be coded as special relations. Define the following condition, given any n and m:

$$\mathrm{Fun}(X^{n,m+1}, n, m) =_{df} \forall [\mathfrak{a}^{<n}]_m, \mathfrak{b}^{<n}, \mathfrak{c}^{<n}((X^n([\mathfrak{a}]_m, \mathfrak{b}) \wedge \wedge X^n([\mathfrak{a}]_m, \mathfrak{c})) \rightarrow \mathfrak{b} = \mathfrak{c})$$

The definition makes use of cross-order identity. We read $\mathrm{Fun}(X^{n,m+1}, n, m)$ as saying that $X^{n,m}$ is an m-ary function defined on concepts of order at most $n-1$. We use $F^{n,m}$ as a meta-variable for any $X^{n,m+1}$ that $\mathrm{Fun}(X^{n,m+1}, n, m)$.[2] We write $F^{n,m}(\mathfrak{a}^{<n})$ for the $\mathfrak{b}^{<n}$ such that $F^n([\mathfrak{a}]_m, \mathfrak{b})$. We add the following schema, for any n:

$$\forall X^{n,m+1}\big(\forall [\mathfrak{a}^{<n}]_m \exists \mathfrak{b}^{<n} X^n([\mathfrak{a}]_m, \mathfrak{b}) \rightarrow \rightarrow \exists F^{n,m} \forall [\mathfrak{a}^{<n}]_m X^{n,m+1}([\mathfrak{a}]_m, F^{n,m}([\mathfrak{a}]_m))\big) \tag{CHn}$$

With CH2, the axiom of choice of 2ZFCU becomes redundant. We also add the following axiom:

$$\exists x(\mathrm{Set}(x) \wedge \forall y(\neg \mathrm{Set}(y) \rightarrow y \in x)) \tag{URS}$$

[2]It is not possible to generalise in the n and m position for any n and m. At most, we can generalise up to a finite k and k', by taking the generalisation as a long conjunction.

Modulo URS, given any n, via CP$(n+1)$ we can well-order the concepts of order n. This fact allows us to extend mathematical induction to arbitrary orders.

Some definitions:

- nZFCU = The result of adding to ZFCU, for any $m \leq n$, CPn, EXTn and CHn and URS as well as all the other extensions of the axioms and rules for the behaviour of quantifiers and identity in first order logic;
- HZFCU = The union of all the nZFCU, for any n;
- Logic of order k = The result of clipping out all axioms about sets to HZFCU, and capping the theory to level k. The logic of order k contains all and only the axioms and rules of first order logic, along with CPn, EXTn, CHn and the other extensions of the rules and axioms for the quantifiers and identity, for any $n \leq k$;
- Higher order logic = The union of all the logics of order k, for any k.

Some final points. To simplify the definitions, I occasionally use (monadic) λ notation, governed by the following conversion principle, where ϕ does not contain Y^n free:

$$\forall Y^n(\lambda X^n.\phi(Y^n) \leftrightarrow \phi(Y^n/X^n))$$

A λ operator is not strictly needed in higher order logic, it is only a definitional expansion.[3] Instead of writing $\lambda X^n.\phi(Y^n)$ one can apply comprehension to ϕ and obtain the equivalent:

$$\exists X^{n+1}(\forall Z^n(\phi(Z^n/X^n) \leftrightarrow X^{n+1}(Z^n)) \wedge X^{n+1}(Y^n))$$

As already explained, I chose 2ZFCU as base for philosophical reasons: I wished to make the comparison with the set-based semantics as plain as possible, and I wished the theory to be as general as possible. However, the semantics I develop does not rely on set-theory: it can be defined using as base some weak syntactic theory and a theory of sequences. Peano arithmetic (PA) could be enough.[4]

[3]Comprehension schemata and a lambda conversion rules are not interchangeable; it is only polyadic comprehension and λ conversion that are so. The truth is that unary λ conversion is stronger than unary comprehension principles. Since λ constructions are iterative, in a lambda calculus an n-ary relation can be coded as a unary predicate: $R(x_1,..,x_n)$ is broke up as $\lambda x_1.\lambda x_2...\lambda x_n.R(x_1,..,x_n)$. Thus, a polyadic lambda conversion principle can be derived from a unary lambda conversion principle. On the other hand, a polyadic comprehension schema cannot be derived from a unary comprehension schema. See Dowek 1995, 2 for details.

[4]Arguably even with PA as base we can still make sense of absolute generality if we are careful enough, and bound the quantifiers in the axioms of PA to the condition of being a natural number. In this way, the quantifiers in the meta-theory can still be said to range over everything, and not only over the countable domain.

A.1.3 Finite higher order sequences

We need a generalised notion of n-tuple that applies to concepts. A theory of finite sequences enables us to spell out a theory of higher order functions for a language of level k at the lowest possible order: viz. $k+1$. We need finite sequences to encode the extension of higher order concepts. Without a theory of sequences, we would need $k+2$ resources to define functions.

It is not possible to define in HZFCU what it means to be a sequence of concepts in general, yet it is possible to define what it means to be a sequence of order n, for an arbitrary n, where a sequence of order n contains elements of at most order n. First, a finite sequence of first order objects is easily definable in set-theory. To say that x is a finite first order sequence of exactly n objects we write $\mathrm{Sq}(x,1,n)$. If we wish to show explicitly the elements of a sequence of x_1, .. x_n we write the usual $\langle x_1,..,x_n\rangle$; to talk of the n-th element of a sequence x (if there is any) we write $x(n)$. We extend the definition to second order (finite) sequences as follows: a second order finite sequence is a ternary relation S between two numbers and (first order) n-tuples. The first number encodes a place in the sequence, the second encodes the order of the object in that place, the third place codes the extension of that object.[5]

Definition 10 (Second order sequence)**.** *Given any k, a second order sequence of length k is any second order relation S that satisfies the following conditions:*

$$\forall x,y,z(S(x,y,z)\rightarrow(x\leq k\wedge y\leq 2)) \tag{2S.1}$$

$$\forall n(n\leq k\rightarrow\exists!x\exists yS(n,x,y)) \tag{2S.2}$$

$$\forall x,y(S(x,1,y)\rightarrow\exists!zS(x,1,z)) \tag{2S.3}$$

$$\forall x,y\Big(S(x,2,y)\rightarrow\exists m\big(\mathrm{Sq}(y,1,m)\wedge\forall z(S(x,2,z)\rightarrow\mathrm{Sq}(z,1,m))\big)\Big) \tag{2S.4}$$

To say that S is a second order sequence of k we write $\mathrm{Sq}(S,2,k)$. We provide the following definitions:

$$\begin{aligned} x \text{ is the } n\text{-th element of } S &=_{df} S(n,1,x)\\ X^2 \text{ is the } n\text{-th element of } S &=_{df} \forall x_1,..,x_n(X(x_1,..,x_n)\\ &\qquad\leftrightarrow S(n,2,\langle x_1,..,x_n\rangle)) \end{aligned}$$

The n-th element of S is the empty concept exactly when $S(n,2,\varnothing)$, where $\varnothing$ codes the empty sequence. On the other hand, the empty set is the n-th element of S

[5]Having the second number that tells you directly the order of the object makes the definition slightly more complicated, but it is needed for book-keeping.

exactly when $S(n, 1, \varnothing)$: the second number is quite useful. The m in 2S.4 must be unique since, by definition, a first order sequence of exactly m elements cannot have but m elements. Since we recover concepts through their extension, a sequence of order 2 can itself be a second order concept, and we do not need to employ third order resources.

Now we provide an inductive 'definition-schema' for what it means to be a sequence of order n. In the definition, variables of order bigger than 1 always have arity 3, which is omitted for readability.

Definition 11 (Sequence of order n)**.** *Assume we have defined what it means to be a sequence up to order n. Given any k, a sequence of order $n+1$ of k elements is any relation S of order $n+1$ that satisfies the following:*

$$\forall x, y, \mathfrak{a}^{\leq n}(S(x, y, \mathfrak{a}) \rightarrow (x \leq k \wedge y \leq n+1)) \tag{nS.1}$$

$$\forall x(x \leq k \rightarrow \exists! y \exists \mathfrak{a}^{\leq n} S(x, y, \mathfrak{a})) \tag{nS.2}$$

$$2S.3 \tag{nS.3}$$

$$\textit{Where } 1 \leq r \leq n \textit{ and } t = r+1: \tag{nS.4}$$

$$\forall x, \mathfrak{a}^{i \leq r}\Big(S(x, t, \mathfrak{a}) \rightarrow \exists m\big(\mathrm{Sq}(\mathfrak{a}, i, m) \wedge \forall \mathfrak{b}^{i \leq r}(S(x, t, \mathfrak{b}) \rightarrow \mathrm{Sq}(\mathfrak{b}, i, m))\big)\Big)$$

$$\textit{Where } 1 \leq r \leq n \textit{ and } t = r+1: \tag{nS.5}$$

$$\forall x, \mathfrak{a}^{i \leq r}\big(S(x, t, \mathfrak{a}) \rightarrow \exists m \exists Y^r(\mathrm{Sq}(Y^r, r, m) \wedge S(x, t, Y^r))\big)$$

$$\textit{Where } r < n: \tag{nS.6}$$

$$\forall x, \mathfrak{a}^{r+1 \leq i \leq n} \neg S(x, r, \mathfrak{a})$$

We write $\mathrm{Sq}(S, n, k)$ to say that S is a sequence of order n and length k. In pseudo-English, nS.4 for $t = 4$ says that if the x-th element of the sequence is of order 4, then S maps x and 4 only to sequences up to order 3. nS.5 further specifies that there must be at least one third order sequence in the range (otherwise, the x-th element would not be of order 4, after all). nS.6 makes the sequence of $X_1, .., X_n$ unique, by excluding irrelevant values. The definitions above are extended as follows, where $m \leq n+1$:

$$X^{m,k} \text{ is the } i\text{-th element of } S =_{df} \forall [\mathfrak{a}^{<m}]_k (X([\mathfrak{a}]_k) \leftrightarrow S(i, m, \langle \mathfrak{a}_1, .., \mathfrak{a}_k \rangle))$$

We extend the common notational conventions to n-th order sequences: given stuff X_1, .., X_n of order up to k, we denote by $\langle X_1, .., X_n \rangle$ the concept of order k which is

the sequence of $X_1 ... X_n$. We write $S(n)$ as a short-hand for the n-th element of S, if there is any.

If in the definition we needed to quantify across all concepts of a fixed order *of some arity or other*, the definition would be infinite. Yet, all we need is to quantify over concepts of arity 3. For sequences are built up from sequences of lower orders. Sequences of any order (from 2 onward) are always ternary relations, so we know the arity of $\mathfrak{a}$ and $\mathfrak{b}$: it is either 0 if they are objects, or 3 if they are relations. This makes the definition specifiable by a single formula.

We can now show that higher order sequences are functional. That is, where S is a sequence of order m and length k, for any $n \leq k$ there is exactly an $\mathfrak{a}^{\leq m}$ that is *the* n-th element of the sequence. It is a tiresome exercise to show that, given any concepts $X_1, .., X_n$, we can build *the* sequence $\langle X_1, .., X_n \rangle$, using the axioms above, which is a concept whose order is the maximum order of its elements. This is a proof-schema, and it requires extensionality. Arguably, in an intensional higher order logic we would need to map a place in the sequence directly to the value of the sequence for that place: this means re-writing everything one level higher.

A.1.3.1 Uses

The language of sequences gives us a way to generalise over all concepts of a certain order 'all at once'. In the syntax, variables are tied to a specific arity. So, we cannot write $\forall X^n \phi$, but rather $\forall X^{n,m} \phi$, when we quantify over all concepts of arity m, for some fixed m. There are infinite arities, so at face value there seems to be no way of expressing the full generalisation in one sentence. However, with the definition of higher order sequence in place, we can get around the problem using a popular trick. Using the properties of sequences we can prove the following schema:

Lemma 17. *Given any m and n, for any $X^{n,m}$ there is a unique $X^{n,1}$ where the following holds:*

$$\forall [\mathfrak{a}^{<n}]_k (X^{n,m}(\mathfrak{a}_1, .., \mathfrak{a}_k) \leftrightarrow X^{n,1}(\langle \mathfrak{a}_1, .., \mathfrak{a}_k \rangle))$$

The lemma shows that we can talk about concepts of order n of arbitrary arity m by talking about the corresponding monadic concepts that apply to sequences of length m and order n. In the following, whenever I write $\forall X^k \phi$, I am implicitly using this lemma: I am quantifying over all monadic concepts that apply to the relevant sequences. Of course, this result does not imply that polyadic higher order logic is 'reducible' to monadic higher order logic, because to define sequences of length n and to prove lemma 17 we are *using* polyadic comprehension.

Finally, we provide the following definitions for functions. I focus on monadic functions: I omit the second superscript in the meta-variable for functions if the function is monadic. The definition is a schema: given two orders n and m, we define what it means to be a function from a monadic concept of order n to a monadic concept of order m. There are two definitions: a general definition that does not use any non-logical vocabulary, and a minimal definition that makes use of our theory of sequences. We already sketched the general definition to provide the choice schema:

Definition 12 (Functions, general definition)**.** *Where* $\max(n, m)$ *is* k*:*

- X^k *is a (monadic) function defined on stuff of order lower than* n*:*

$$\mathrm{Fun}(X^k, k, 1)$$

- $X^k(\mathfrak{a}^{<n})$ *: the* $\mathfrak{b}^{<m}$ *such that* $X^k(\mathfrak{a}, \mathfrak{b})$
- *Consider this condition* $\phi(X^k, X^n, Y^m)$*:*

$$\forall \mathfrak{a}^{<n}, \mathfrak{b}^{<m}(X^k(\mathfrak{a}) = \mathfrak{b} \rightarrow (X(\mathfrak{a}) \wedge Y(\mathfrak{b})))$$

  The domain of X^k *is* X^n *and its range is* Y^m *means:* X^n *and* Y^m *are the minimal concepts that satisfy* ϕ*.*

- X^k *is injective:* $\forall \mathfrak{a}^{<n}, \mathfrak{b}^{<m}(X^k(\mathfrak{a}) = X^k(\mathfrak{b}) \rightarrow \mathfrak{a} = \mathfrak{b})$
- X^k *is surjective:* $\forall \mathfrak{a}^{<n} \exists \mathfrak{b}^{<m} X^k(\mathfrak{b}) = \mathfrak{a}$
- X^k *is a bijection:* X^k *is surjective and injective.*

We can code functions through the minimal definition from a concept of order n to a concept of order m directly at order $m-1$, if $m > n$. That is, a minimal function needs to be of order strictly higher than the concepts in the domain, but it can be of the same order of the concepts in the range. This is done using sequences: instead of mapping X directly to Y, we map it to all the sequences of concepts that satisfy Y, which are at most one order lower than Y. The definition however makes use of non-logical terminology: we need some arithmetic (translated in the language of set-theory) to define first order sequences, numbers etc.

Definition 13 (Functions, minimal definition)**.** *Where* $m > n$ *and* F *is of order* m*:*

– F is a function between stuff of order $\leq n$ and stuff of order $\leq m$: F is a binary relation between some $\mathfrak{a}^{\leq n}$ and sequences of order up to $m-1$ such that, if some $\mathfrak{a}^{\leq n}$ is mapped to something, it is mapped only to sequences of the same length:

$$\forall \mathfrak{a}^{\leq n}\big(\exists \mathfrak{b}^{<m} F(\mathfrak{a},\mathfrak{b}) \rightarrow \exists! i \forall \mathfrak{c}^{t<m}(F(\mathfrak{a},\mathfrak{c}) \rightarrow \mathrm{Sq}(\mathfrak{c},t,i))\big)^{6}$$

– Where each $\mathfrak{b}^{<m}$ such that $F(\mathfrak{a},\mathfrak{b})$ is a sequence of length t, $F(\mathfrak{a}^{\leq n})$ is the unique $\mathfrak{b}^{i<m}$ such that:

$$\forall [\mathfrak{c}^{<i}]_t (F(\mathfrak{a}, \langle \mathfrak{c}_1, .., \mathfrak{c}_t \rangle) \leftrightarrow \mathfrak{b}([\mathfrak{c}]_t))$$

– Everything else works like before, simply by replacing $F(\mathfrak{a}^{\leq n})$ with the appropriate new definitions.

It is tiresome but straightforward to prove that, whenever the minimal definition can be applied, for any function in the sense of the general definition, there is a function in the sense of the minimal definition.

We (implicitly) define the cardinality of an order as follows:

Definition 14 (Cardinality)**.** *A concept X^m has cardinality 'order-n-many' when there is a bijection F such that:*

– *The domain of F is X^m;*

– *The range of F is $\lambda X^{n,1}.X^{n,1} = X^{n,1}$.*

By lemma 17 we know that there are exactly as many polyadic concepts of order n as monadic concepts of order n, so the definition works.

A.2 Standard, Higher order Semantics

It is not possible to state in HZFCU what it is *in general* to be a higher order assignment, a higher order interpretation or a higher order concept of truth in an interpretation, given a language of arbitrary level. Rather, we can only define through schemata what interpretation, assignment and truth mean for a language capped to a specific level k. It follows that we never interpret the whole language of HZFCU, but only a proper subset of it capped to some level. This is different from model-theory, where we can develop in a first order language a theory of validity for languages of arbitrary levels. For simplicity, I will consider only languages which are proper subsets

[6] This is needed to pin down exactly one value, which will need to have one specific arity.

of the language of HZFCU capped to some level, without any further syntax. These languages do not have constants or predicate symbols apart from '$\in$', 'Set' and '='. The theory can easily be extended in obvious ways to syntactically richer languages, if needed.

A.2.1 Coding

We assume that the language of HZFCU is coded in a natural way in set-theory.[7] In set-theory we do not have function symbols or constants, but codes can be written via suitable formulae. Given a (possibly improper) subset L of the language of HZFCU, 'Form(x, L)' stands for 'x codes a formula of L'. I use Greek letters as meta-variables for (codes of) formulae of the language at hand. 'Var(x, n, m)' stands for 'x codes a variable of level n and arity m'. I use $V^{n,m}$ as a meta-variable for things that satisfy Var(x, n, m). I use $V^{\leq n}$ as a meta-variable for codes of variables of levels up to n; that is, $\forall V^{<n}\phi$ stands for $\forall x, m, k((k < n \wedge \text{Var}(x, k, m)) \rightarrow \phi(x))$. I use semi-quotes; we read $\ulcorner\phi \wedge \psi\urcorner$ as: 'the result of concatenating the code of ϕ with the code of conjunction with the code of ψ'. So, $\ulcorner x\urcorner$ is the result of concatenating the code of x with nothing: i.e. it is the code of x. I often do not distinguish between the code of x and x.

A.2.2 Interpretations and assignments

We show how to interpret the restriction of HZFCU to level k, which is the subset of the language of HZFCU with variables up to level k. We call this language L^k. In the following we assume that k is at least 2, to simplify some definitions (in particular, the definition of satisfaction). It is straightforward to extend what will be said to $k = 1$, if we wished to. Interpretations and assignments always interpret n-ary concepts as monadic concepts that apply to n-ary sequences (otherwise satisfaction would not be finitely definable).

We build a higher order interpretation and assignment from a first order domain (which I will simply call 'domain'). The first order domain is any monadic second order concept D. Given a first order domain, we first need to define what it means to be in the projected domain at order n, for $n \leq k$. That is, what it means to be in the range of n-th level quantifiers under the interpretation. Something is in the first order projected domain of D exactly if it Ds; something is in the projected n-th order domain of D where $2 \leq n$ exactly if it applies only to sequences of a fixed length whose

[7] For example, using hereditarily finite sets; alternatively, by translating PA in ZFC, and then by using an arithmetical coding.

elements are in some projected domain of D for some m_1, .., m_j strictly smaller than n.

Definition 15 (Projected domains)**.** *Given a second order concept D, given any $n \leq k$, we define what it means for some (monadic) X^n to be in the n-th order projected domain of D – written* $\mathrm{Dm}^n(X^n, D)$ *– as follows:*

(DM1) $D(X^n)$ *if* $n = 1$

(DM2) if $n \neq 1$, $\mathrm{Dm}^n(X^n, D)$ *is short for the conjunction of the following conditions:*

$$\exists t \forall \mathfrak{a}^{m<n}(X^n(\mathfrak{a}) \rightarrow \mathrm{Sq}(\mathfrak{a}, m, t)) \tag{SM1}$$

$$\forall \mathfrak{a}^{m<n}, \mathfrak{b}^{i<m}, j((X^n(\mathfrak{a}) \wedge \mathfrak{a}(j) = \mathfrak{b}) \rightarrow \mathrm{Dm}^i(\mathfrak{b}, D)) \tag{SM2}$$

We write $\mathrm{Dm}^m_n(X^m, D)$ *to say that X^m is in the m-th order projected domain of D and applies only to sequences of length n. In 'Formalese':*

$$\mathrm{Dm}^m(X^m, D) \wedge \forall \mathfrak{b}^{i<m}(X^m(\mathfrak{b}) \rightarrow \mathrm{Sq}(\mathfrak{b}, i, n))$$

The definitions of assignment and interpretation given a domain are now straightforward:

Definition 16 (Higher order interpretation)**.** *An interpretation based on a domain D is any function F (in the sense of the minimal definition) such that:*

(I1) The domain of F is $\lambda x.x = \ulcorner\in\urcorner \vee x = \ulcorner\mathrm{Set}\urcorner$*;*

(I2) $F(\ulcorner\in\urcorner)$ *and* $F(\ulcorner\mathrm{Set}\urcorner)$ *are in the second order projected domain of D;*

(I3) $F(\ulcorner\in\urcorner)$ *applies to binary sequences and* $F(\ulcorner\mathrm{Set}\urcorner)$ *applies to monadic sequences.*

$\mathrm{Int}(X^2, D)$ stands for: X^2 is a higher order interpretation based on D. We use the meta-variable I for any X^2 that satisfies $\mathrm{Int}(X^2, Y^2)$ for some Y^2. Since there are no predicate letters of level $n > 2$ in the language, an interpretation is always second order regardless of level of the language interpreted.

Definition 17 (Variable assignment)**.** *Given a second order concept D, an assignment of variables up to level k based on D is a higher order function F (in the sense of the minimal definition) of order k such that:*

(A1) The domain of F is $\lambda x.\exists m, n(m \leq k \wedge \mathrm{Var}(x, m, n))$*;*

(A2) For any $V^{1,0}$, $D(F(V^{1,0}))$;

(A3) For any m: $2 \leq m \leq k$, *for any* n, *for any (code of) variable* $V^{m,n}$, $\mathrm{Dm}^m_n(F(V^{m,n}), D)$.

We write $\mathrm{Asg}(X^k, D)$ to say that X^k is an assignment of all variables up to level k based on D. We use the meta-variable A for any X^k that satisfies $\mathrm{Asg}(X^k, X^2)$ for some X^2. A is an assignment for a language L iff it is able to provide a value to all variables of L.

A.2.3 Standard satisfaction and standard validity

Definition 18. *Given any domain* D, *given an assignment* A *of order* k *based on* D, *the result of switching to* X^n *the value of* V^n *in* A *– written* $\pi(A, X^n/V_n)$ *– is the assignment defined as follows, where* U^n *is (the code of) a variable of level* n*:*

$$\pi(A, X^n/V_n)(U^n) =_{df} \begin{cases} A(U^n) & \text{if } U^n \neq V^n \\ X^n & \text{otherwise.} \end{cases}$$

Definition 19 (Satisfaction under an interpretation and assignment)**.** *Where* C *is a concept of order* $k+1$, *to be satisfaction for* L^k *– written* $\mathrm{SAT}(C)$ *– stands for the following: for any second order monadic concept* D, *for any higher order interpretation* I *and assignment* A *for* L^k *based on* D, *where* ϕ *and* ψ *are formulae of* L^k*:*

$$C(\ulcorner v \in u \urcorner, I, A) \leftrightarrow I(\ulcorner \in \urcorner)(\langle A(v), A(u) \rangle) \tag{Sat1}$$

$$C(\ulcorner \mathrm{Set}(v) \urcorner, I, A) \leftrightarrow I(\ulcorner \mathrm{Set} \urcorner)(\langle A(v) \rangle) \tag{Sat2}$$

$$C(\ulcorner V^{n \leq k} = U^{m \leq k} \urcorner, I, A) \leftrightarrow A(V^n) = A(U^m) \tag{Sat3}$$

$$C(\ulcorner V^{2 \leq m \leq k, n}([U^{<m}]_n) \urcorner, I, A) \leftrightarrow A(V^{m,n})(\langle A(U_1), .., A(U_n) \rangle)^8 \tag{Sat4}$$

$$C(\ulcorner \neg \phi \urcorner, I, A) \leftrightarrow \neg C(\ulcorner \phi \urcorner, I, A) \tag{Sat5}$$

$$C(\ulcorner \phi \wedge \psi \urcorner, I, A) \leftrightarrow (C(\ulcorner \phi \urcorner, I, A) \wedge C(\ulcorner \psi \urcorner, I, A)) \tag{Sat6}$$

Where $n \leq k$: (Sat7)

$$C(\ulcorner \forall V^{n,m} \phi \urcorner, I, A) \leftrightarrow \forall X^{n,1}\big(\mathrm{Dm}^n_m(X^{n,1}, D) \rightarrow C(\ulcorner \phi \urcorner, I, \pi(A, X^{n,1}/V^n))\big)$$

We use the meta-variable 'Sat' for any X^{k+1} that satisfies SAT. The existence of a satisfaction concept for L^k is ensured by comprehension schemata at order $k+1$. By induction on the complexity of ϕ, we can show that, for any D, for any I and A based on D, for any ϕ of L^k, ϕ, I and A holds of some Sat exactly when it holds of any Sat.

ϕ is true under I based on D iff ϕ is satisfied under I and all assignments based on D.

Definition 20 (Truth under an interpretation)**.** *Where I is based on D and ϕ is any formula of L^k:*

$$\mathrm{Tr}(\ulcorner\phi\urcorner, I) \leftrightarrow_{df} \forall \mathrm{Sat}, X^k(\mathrm{Asg}(X^k, D) \rightarrow \mathrm{Sat}(\ulcorner\phi\urcorner, I, X^k))$$

We define logical validity for languages of level k:

Definition 21 (Logical validity at order k)**.** *Where ϕ and Γ are sentences and set of sentences of L^k, respectively, that Γ logically implies ϕ – written $\Gamma \vDash^k \phi$ – is defined as follows at order $k+1$:*

$$\forall I(\forall\psi(\psi \in \Delta \rightarrow \mathrm{Tr}(\ulcorner\psi\urcorner, I)) \rightarrow \mathrm{Tr}(\ulcorner\phi\urcorner, I))$$

A.2.4 Virtues of the standard semantics

The following theorem-schema is provable, given any k:

Theorem 8 (Truth for L^k)**.** *There is a higher order interpretation I where* Sat *satisfies the Tarskian axioms for satisfaction for the restriction of the language of* HZFCU *to k.*

Proof. By 'Tarskian axioms' I mean the axioms in Tarski 1933: a T-schema for atomic formulae, and the compositional axioms for connectives and quantifiers. We consider the interpretation based on the universal concept: $\lambda x.x = x$. By comprehension, there is an interpretation I such that
$\forall x, y(I(\ulcorner\in\urcorner)(\langle x, y\rangle) \leftrightarrow x \in y)$ and $\forall x(I(\ulcorner\mathrm{Set}\urcorner)(\langle x\rangle) \leftrightarrow \mathrm{Set}(x))$. The theorem is proved by induction on the complexity of formulae. □

The system does not fall prey to the liar paradox, since SAT is not definable in L^k, but only in L^{k+1}. Since we do not have quantification across orders, the theorem remains a schema, and cannot be generalised explicitly for any n.

For any set-based interpretation of L^k there is an equivalent higher order interpretation. First, our base is an extension of ZFC, so we can define in HZFCU a set-based model $\mathfrak{M}$ as a pair $\langle d, v\rangle$ of a set and an interpretation function on d; truth under $\mathfrak{M}$ – written $\vDash$ – is defined in the usual way[9]:

Lemma 18. $\forall\mathfrak{M}\exists I\forall\phi(\mathfrak{M} \vDash \ulcorner\phi\urcorner \leftrightarrow \forall \mathrm{SatSat}(\ulcorner\phi\urcorner, I))$

[9]For a definition, see Chang and Keisler 1990.

Proof. The relevant I based on D is the following:

$$\forall x(x \in d \leftrightarrow D(x))$$
$$\forall x, y(\langle x, y\rangle \in v(\ulcorner\in\urcorner) \leftrightarrow I(\ulcorner\in\urcorner)(\langle x, y\rangle))$$
$$\forall x(x \in v(\ulcorner\mathrm{Set}\urcorner) \leftrightarrow I(\ulcorner\mathrm{Set}\urcorner)(\langle x\rangle))$$

The proof is by induction on the complexity of ϕ, using the definition of Sat and $\vDash$. □

Theorem 9. *Given a language L^k, the logic of order k in L^k is sound for validity at order k. It is not complete. In fact, no effective and sound deductive system for L^k is complete for validity at order k (where $k \geq 2$).*[10]

The theorem is proven below, where I prove soundness and incompleteness for raising semantics. Since standard semantics is a special case of raising semantics (by 0 orders), theorem 9 is a corollary.

We now prove 'There are more concepts at order $k+1$ than concepts/objects at order k'. We use the trick above to prove the theorem already at order $k+1$ instead of $k+2$.[11]

Theorem 10. *[K-Higher Cantor] In $(k+1)$ZFCU the following holds: there is no surjective function from order k to order $k+1$.*

Proof. A surjective function from order k to order $k+1$ is a function in the sense of the minimal definition for functions provided above, such that the following holds:

$$\forall X^k \exists Y^{k+1} F(X^k) = Y^{k+1}$$
$$\forall X^{k+1} \exists Y^k F(Y^k) = X^{k+1}$$

F is of order $k+1$. Proceed by contradiction. Suppose there was such an F. By (predicative) comprehension:

$$\exists X^{k+1} \forall Y^k(\neg F(Y^k)Y^k \leftrightarrow X^{k+1}(Y^k))$$

Since F is surjective, there is some Z^k such that $F(Z^k) = X^{k+1}$. When we instantiate Y^k in the comprehension instance with Z^k we obtain $X^{k+1}Z^k \leftrightarrow \neg F(Z^k)Z^k$. Yet, since $F(Z^k) = X^{k+1}$, we obtain a contradiction. □

[10] By 'effective' deductive system I mean one whose axioms and rules are recursive.

[11] I found a very similar result in Florio and Linnebo 2021, in the scope of a plural semantics, where sequences are defined differently.

The theorem shows that there must be strictly more X^{k+1} than X^k, since otherwise there would be such a function. It only needs predicative comprehension.

Via 10 we can show the following lemma:

Lemma 19. *There are strictly more higher order interpretations than set-based interpretations of any language of level k.*

Proof. The language must contain at least a predicate P. Since set-based models are sets and sets are objects, there are at most order-1-many interpretations of P. By definition of higher order interpretation, there are at least order-2-many interpretations of P. By 10, there are strictly more concepts of order 2 than objects, so there are strictly more higher order interpretations than set-based interpretations. □

A.3 Raising higher order semantics

We introduce a semantic system that allows one to raise the values of variables by some finite e order. In the following I define interpretations and assignments that raise the values up to some e order, instead of exactly by e orders.

We are again unable to define what it means in general to be an interpretation or assignment that raises the values by some n or other. All we can do is to provide a definition-schema for interpretations and assignments for L^k that raise the values up by some specific e.

A (first order) domain is now any monadic concept of order between 2 and $e+2$. We extend the definition of projected domain accordingly:

Definition 22 (Raising projected domains)**.** *Given a concept $\mathfrak{a}^{2\leq i\leq e+2}$, where $m \leq k$ and where $r = i-2$, we define what it means for $\mathfrak{b}^{r+m}$ to be in the m-th order projected domain of D – written $\mathrm{Dm}^m(\mathfrak{b},\mathfrak{a})$ – as follows:*

*(DM*1)* $\mathfrak{a}(X^{r+1})$ *if* $m = 1$.

*(DM*2) If* $m \neq 1$, $\mathrm{Dm}^m(X^{r+m},\mathfrak{a})$ *is short for the conjunction of the following conditions:*

$$\exists t \forall \mathfrak{b}^{s<r+m}(X^{r+m}(\mathfrak{b}) \rightarrow \mathrm{Sq}(\mathfrak{b},s,t)) \tag{a}$$

$$\forall \mathfrak{b}^{s<r+m}, \mathfrak{c}^{t<s}, j((X^{r+m}(\mathfrak{b}) \wedge \mathfrak{b}(j) = \mathfrak{c}) \rightarrow \mathrm{Dm}^{t-r}(\mathfrak{c},\mathfrak{a})) \tag{b}$$

We write $\mathrm{Dm}^m_n(X^{r+m},\mathfrak{a})$ to say that X^{m+r} is in the m-th order projected domain of D and applies only to sequences of length n.

$\mathrm{Dm}^m(X^n, \mathfrak{a})$ is undefined whenever $n \neq r + m$.

The definition of interpretation and assignment is changed accordingly in obvious ways, using the extended notion of projected domain. We write $\mathrm{Int}^*(X, X^i)$ to say that X is a raising interpretation based on X^i. When based on X^i, X is of order i and it is a $+(i-2)$ interpretation: it uniformly raises the values of variables of L^k by $i-2$. A $+0$ interpretation is therefore a concept of order 2: it is a standard interpretation. Given any n, we use the meta-variable I^{+n} for $+n$ interpretations. I use the following bit of notation: when I write $\forall(n \leq e)\phi(I^{+n})$ I mean that ϕ holds of any raising interpretation that raises the values up by $n \leq e$; i.e. I mean: $\forall \mathfrak{a}^{2 \leq i \leq e+2} \forall \mathfrak{b}^i(\mathrm{Int}^*(\mathfrak{b}, \mathfrak{a}) \rightarrow \phi(\mathfrak{b}))$. We need to bound n to e for the generalisation to be finitely expressible.

We write $\mathrm{Asg}^*(X, X^i)$ to say that X is a raising assignment based on X^i: a $+(i-2)$ assignment. We use the meta-variable A^{+n} for $+n$ assignments. The same notational conventions we stipulated for I^{+n} hold for A^{+n} when we bind n with a quantifier.

Finally, we need to extend the definition of satisfaction to cover the raising assignments and interpretations up by at most e. The new satisfaction for L^k is redefined explicitly at order $k + e + 1$:

Definition 23 (Satisfaction up to e)**.** *Where C is a concept of order $k + e + 1$, to be raising satisfaction for L^k up to e – written SAT*(C, e) – means the following: for any (monadic) $\mathfrak{a}^{2 \leq i \leq e+2}$, where $r = i - 2$, for any higher order interpretation I^{+r} and assignment A^{+r} for L^k based on $\mathfrak{a}$, where ϕ and ψ are formulae of L^k:*

$$C(\ulcorner v \in u \urcorner, I^{+r}, A^{+r}) \leftrightarrow I^{+r}(\ulcorner \in \urcorner)(\langle A^{+r}(v), A^{+r}(u) \rangle) \tag{Sat1*}$$

$$C(\ulcorner \mathrm{Set}(v) \urcorner, I^{+r}, A^{+r}) \leftrightarrow I^{+r}(\ulcorner \mathrm{Set} \urcorner)(\langle A^{+r}(v) \rangle) \tag{Sat2*}$$

$$C(\ulcorner V^{n \leq k} = U^{m \leq k} \urcorner, I^{+r}, A^{+r}) \leftrightarrow A^{+r}(V^n) = A^{+r}(U^m) \tag{Sat3*}$$

$$C(\ulcorner V^{2 \leq m \leq k, n}([U^{<m}]_n) \urcorner, I^{+r}, A^{+r}) \leftrightarrow A^{+r}(V^{m,n})(\langle A^{+r}(U_1), .., A^{+r}(U_n) \rangle) \tag{Sat4*}$$

$$C(\ulcorner \neg \phi \urcorner, I^{+r}, A^{+r}) \leftrightarrow \neg C(\ulcorner \phi \urcorner, I^{+r}, A^{+r}) \tag{Sat5*}$$

$$C(\ulcorner \phi \wedge \psi \urcorner, I^{+r}, A^{+r}) \leftrightarrow (C(\ulcorner \phi \urcorner, I^{+r}, A^{+r}) \wedge C(\ulcorner \psi \urcorner, I^{+r}, A^{+r})) \tag{Sat6*}$$

Where $n \leq k$: (Sat7*)

$$C(\ulcorner \forall V^{n,m} \phi \urcorner, I^{+r}, A^{+r}) \leftrightarrow \forall X^{n+r,1}(\mathrm{Dm}^n_m(X^{n+r,1}, \mathfrak{a}) \rightarrow \\ \rightarrow C(\ulcorner \phi \urcorner, I^{+r}, \pi(A^{+r}, X^{n+r,1}/V^n)))$$

We use Sat^{+e} as a meta-variable for any X^{k+e+1} that SAT* with e.

Definition 24 (Truth under raising interpretations)**.** *For any* $\mathfrak{a}^{2\leq i\leq e+2}$*, where* $t = i-2$*,* I^{+t} *is based on* $\mathfrak{a}$ *and* ϕ *is any formula of* L^k*:*

$$\mathrm{Tr}(\ulcorner\phi\urcorner, I^{+t}) \leftrightarrow_{df} \forall \mathrm{Sat}^{+e}, X^{k+t}(\mathrm{Asg}(X^{k+t}, \mathfrak{a}) \rightarrow \mathrm{Sat}(\ulcorner\phi\urcorner, I^{+t}, X^{k+t}))$$

Raising validity for L^k is defined as follows, at order $k+e+1$:

Definition 25 ($+e$ raising validity for L^k)**.** *Where* ϕ *and* Γ *are sentences and set of sentences of* L^k*, respectively, that* ϕ *follows from* Γ *in a raising* $+e$ *semantics – written* $\Gamma \vDash^k_{+e} \phi$ *– abbreviates the following:*

$$\forall(r \leq e), I^{+r}(\forall\psi(\psi \in \Delta \rightarrow \mathrm{Tr}(\ulcorner\psi\urcorner, I^{+r})) \rightarrow \mathrm{Tr}(\ulcorner\phi\urcorner, I^{+r}))$$

As mentioned, we are unable to define what it means to be true under all $+n$ interpretations, for any n. We can only define, given an n, what it means to be true under all raising interpretations up to $+n$.

A.3.1 Soundness

In the rest of the appendix, for readability I will often omit the superscripts to variables and meta-variables; instead, the information in the superscript is stated explicitly when I introduce the variable/meta-variable.

We show that raising models are actually models of the logic at hand.

Theorem 11 (Soundness)**.** *The logic of order* k *is sound for* $+e$ *raising semantics.*

Proof. The proof is by induction on the length of proofs of the logic of order k. First, since any Sat^{+e} commutes with the connectives, the formula that states that such axioms and rules are true and truth-preserving under any $I^{+(r\leq e}$ and $A^{+(r\leq e)}$ is equivalent to an instance of these axioms and rules, hence they are true and truth-preserving.

It is an obvious consequence of the definition of raising interpretations and assignments that every value of the free variables are inside their respective projected domains: universal instantiation and existential generalisations are sound, for any order up to k.

To deal with comprehension and choice schemata, we make use of the fact that, for any schema CPn and CHn, for any $r \leq e$, for any $+r$ I, satisfaction of CPn and CHn under I is a consequence of some CP$(n+r)$ and CH$(n+r)$. I show it for an instance of monadic CP1:

$$\exists V^2 \forall u(\phi \leftrightarrow V^2(u)) \qquad \text{(CP1)}$$

I know that ϕ does not contain V^2 free. Via definition of truth under a $+r$ I, I makes true CP1 exactly when CP1 is satisfied under I and all $+r$ assignments. Assume $i = r + 1$. By definition of satisfaction, this is so exactly when there is a Y^{i+1} in the second order domain of I such that, for any X^i in the first order domain of I, for any $+r$ A, the following holds:

$$\mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I, \pi(A, X^i/u)) \leftrightarrow \pi(A, Y^{i+1}/V^2)(V^2)\pi(A, X^i/u)(u)$$

By definition of π, the right-hand side of the biconditional is equivalent to $Y^{i+1}(X^i)$. So, where I and A are based on D, CP1 is satisfied under I and A when the following holds:

$$\exists Y^{i+1}\big(\mathrm{Dm}^2(Y^{i+1}, D) \wedge \forall X^i(\mathrm{Dm}^1(X^i, D) \rightarrow \\ \rightarrow \mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I, \pi(A, X^i/u)) \leftrightarrow Y^{i+1}(X^i))\big)$$

Where I is based on D, by definition of Sat^{+e}, ϕ is satisfied under I and $\pi(A, X^i/u)$ only if the domain of I applies to X^i. So, by logic:

$$(\mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I, \pi(A, X^i/u)) \wedge \mathrm{Dm}^1(X^i, D)) \leftrightarrow \mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I, \pi(A, X^i/u))$$

The following is an instance of CP$(i + 1)$:

$$\exists Y^{i+1}\forall X^i((\mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I, \pi(A, X^i/u)) \wedge \mathrm{Dm}^1(X^i, D)) \leftrightarrow Y^{i+1}(X^i))$$

Using the equivalence above, it is equivalent to:

$$\exists Y^{i+1}\forall X^i(\mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I, \pi(A, X^i/u)) \leftrightarrow Y^{i+1}(X^i))$$

By logic, if every X^i satisfies this biconditional then every X^i in the domain of I satisfies the biconditional.

$$\exists Y^{i+1}\forall X^i\big(\mathrm{Dm}^1(X^i, D) \rightarrow (\mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I, \pi(A, X^i/u)) \leftrightarrow Y^{i+1}(X^i))\big)$$

Finally, we check that Y^{i+1} is in the second order domain of I. It is because any X^i such that $Y^{i+1}(X^i)$ is in the first order domain of I and, by definition of projected domain, the second order domain of I must contain *all* concepts whose elements are concepts of the first order domain, so Y^{i+1} must belong to it. □

A.3.2 Indefinite raisability of the logic of order n.

The standard interpretations are among the raising ones, for $r = 0$. So, trivially, there are at least as many standard interpretations than raising interpretations. However, there are strictly more raising interpretations than standard interpretations whenever $r \neq 0$. In general, there are more $+n$ raising interpretations than $+m$ raising interpretations whenever $n > m$.

Lemma 20. *There are strictly more $+n$ interpretations than $+m$ interpretations, when $n > m$.*

Proof. By theorem 10, there are strictly more n-th order concepts than m-th order concepts. By definition of I^{+n} and I^{+m}, in a $+m$ semantics there are exactly order-$(m+2)$-many values of Set, while by definition of I^{+n}, in a $+n$ semantics there are exactly order-$(n+2)$-many values of Set. Since $n+2 > m+2$, there are strictly more I^{+n} than I^{+m}. □

The lemma shows that, for any given notion of validity defined at order k, there are strictly more interpretations reachable one order higher. Also, given the definition of $+n$ strict adequacy, any raising semantics that defines validity as preservation of truth under $+e$ interpretations is not $+n$ strictly adequate when $n > e$.

I am now going to prove that considering raising interpretations up to e or exactly by e does not change the extension of validity. First, some definitions. An isomorphism of models up to order $+e$ is defined as follows:

Definition 26 (Isomorphism for $+i \leq e$ models)**.** *For any two raising interpretations I^{+n} and N^{+m} for L^k with $n \leq e$ and $m \leq e$, we say that I^{+n} and N^{+m} are isomorphic when there is a function F such that:*

- *F is a bijection;*
- *The domain of F is the first order domain of I^{+n} and the range of F is the first order domain of N^{+m};*
- *For any X and Y in the domain of F:*

$$I^{+n}(\ulcorner\in\urcorner)(\langle X, Y\rangle) \leftrightarrow N^{+m}(\ulcorner\in\urcorner)(\langle F(X), F(Y)\rangle)$$
$$I^{+n}(\ulcorner\mathrm{Set}\urcorner)(\langle X\rangle) \leftrightarrow N^{+m}(\ulcorner\mathrm{Set}\urcorner)(\langle F(X)\rangle)$$

Two interpretations are isomorphic if there is an isomorphism between them. Two isomorphic raising interpretations I and N (up by $+n$ and $+m$, respectively) are equivalent. I sketch the proof.

Lemma 21. *Any two isomorphic raising interpretations I and N for L^k which raise the value by $n \leq e$ and $m \leq e$ respectively are equivalent. That is, the following holds.*

$$\forall \mathrm{Sat}^{+e}, \phi(\mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I) \leftrightarrow \mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, N))$$

We assume I and A stand for $+n$ interpretations and assignments and N and B for $+m$ interpretations and assignments, respectively, with $n \leq e$ and $m \leq e$. To prove the lemma, one first shows the following, for some Sat^{+e} (and therefore for all Sat^{+e}):

$$\forall A \exists B \forall \phi(\mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I, A) \leftrightarrow \mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, N, B)) \quad \text{(a)}$$

$$\forall B \exists A \forall \phi(\mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, I, A) \leftrightarrow \mathrm{Sat}^{+e}(\ulcorner\phi\urcorner, N, B)) \quad \text{(b)}$$

Proof. The proof is by induction. The relevant A and B are constructed following F. I show it for (a). For (b), one would 'follow' F^{-1} instead, which exists because F is a bijection. We build the function G by comprehension:

$$G(X^{n+1}) = F(X^{n+1}) \quad \text{(a)}$$

$$\text{Where } 2 \leq t \leq k: \quad \text{(b)}$$

$$G(\mathfrak{a}^{n+t,i}) = \text{the } \mathfrak{b}^{m+t} : \forall [\mathfrak{c}^{n<j<n+t}]_i (\mathfrak{a}([\mathfrak{c}]_i) \leftrightarrow \mathfrak{b}([G(\mathfrak{c})]_i))$$

(In the definition just given we are using sequences: we are quantifying over monadic concepts that apply to sequences of length i). Given any A, its correspondent B is the following:

$$B(V^{\leq k}) = G(A(V^{\leq k}))$$

We can show that the lemma holds by induction of the complexity of ϕ. The base case holds by definition of isomorphism for $\ulcorner\mathrm{Set}(v)\urcorner$ and $\ulcorner v \in u\urcorner$. Consider a formula $\ulcorner V^{j,i}([U]_i)\urcorner$, where $r < j$ and V is of strictly higher level than each of $U_1, .., U_1$:

$$\begin{aligned}
&\mathrm{Sat}^{+e}(\ulcorner V^{j,i}([U]_i)\urcorner, I, A) \\
\leftrightarrow &A(V^{j,i})(\langle A(U_1), .., A(U_i)\rangle) \\
\leftrightarrow &G(A(V^{j,i}))(\langle G(A(U_1)), .., G(A(U_i))\rangle) \\
\leftrightarrow &B(V^{j,i})(\langle B(U_1), .., B(U_i)\rangle) \\
\leftrightarrow &\mathrm{Sat}^{+e}(\ulcorner V^{j,i}([U]_i)\urcorner, N, B)
\end{aligned}$$

The induction step follows by definition of Sat^{+e} and G. □

We provide the following corollary:

Lemma 22. *[Push-up construction] For any $r \leq e$, for any $+r$ interpretation there is an isomorphic $+e$ interpretation.*

Proof. The proof is trivial for $r = e$. I focus on the case where $r < e$. We pick a specific X^{e+1}, for example $\lambda X^e.X^e = X^e$: call this concept C. Given a $+r$ model I based on D, an isomorphic $+e$ N is defined as follows. First, consider the function G whose domain is D such that $G(X^{r+1}) = \langle X^{r+1}, C^{e+1}\rangle$:

$$\text{Domain of } N\text{: } \lambda X^{e+1}.\exists X^{r+1} G(X^{r+1}) = X^{e+1} \tag{a}$$

$$\forall X^{r+1}, Y^{r+1} I(\ulcorner\in\urcorner)(\langle X^{r+1}, Y^{r+1}\rangle) \leftrightarrow N(\ulcorner\in\urcorner)(\langle G(X^{r+1}), G(Y^{r+1}\rangle)) \tag{b}$$

$$\forall X^{r+1} I(\ulcorner\mathrm{Set}\urcorner)(\langle X^{r+1}\rangle) \leftrightarrow N(\ulcorner\mathrm{Set}\urcorner)(\langle G(X^{r+1})\rangle) \tag{c}$$

Clearly G is an isomorphism between I and N. □

The following lemmata are straightforward consequences of lemma 22.

Lemma 23. *That $\Gamma \vDash_e^k \phi$ in L^k is equivalent to preservation of truth under all interpretations that raise the values exactly by e. That is, it is equivalent to the following:*

$$\forall I^{+e}\big(\forall\psi(\psi \in \Delta \rightarrow \mathrm{Tr}(\ulcorner\psi\urcorner, I^{+e},)) \rightarrow \mathrm{Tr}(\ulcorner\phi\urcorner, I^{+e})\big)$$

Lemma 24. *For any standard model I there is an equivalent raising model I^{+e}, given any e.*

The cardinality of an interpretation is the cardinality of its first order domain. We know that, for any k, the logic of order k is not categorical. However, it is assumed in the standard semantics that there is a limit to how 'big' models can be: they can be at most as big as the intended model. In general, for any language of level k, any two standard model of k are such that the first order domain is at most as big as the universe; the second order domain is at most big as order 2, etc. If we consider raising interpretations, this cap to the size of interpretations is lifted.

Theorem 12 (No-limit)**.** *Pick an n: there is a (raising) model for k-th order logic of cardinality order-n-many.*

Proof. Following the definitions of raising interpretations, assignments etc, we construct $+(n-1)$ interpretations. Consider any $+(n-1)$ model N whose domain is $\lambda X^n.X^n = X^n$. The cardinality of the domain of I is order-n-many. By theorem 11, N is a model of the logic of order k. □

A.3.3 Categoricity and incompleteness

We now show that validity defined over raising raising interpretations (i.e raising semantics) is incomplete, non-compact etc, pretty much like the standard semantics. First, we show failure of compactness.

Definition 27 (Compactness)**.** *Given a semantic theory T and an effective deductive system D spelled out in language L, we say that D is compact under T when, given a set of formulae Γ of L, Γ is satisfiable in T iff every finite subset of Γ is satisfiable in T.*

What compactness means depend on what 'satisfiable' means, and how we defined models, hence the need to specify the semantic system at hand. For example, some logical truths of standard semantics are not falsifiable in Henkin semantics, so some $\neg\phi$ that is satisfiable in Henkin semantics is not satisfiable in standard semantics.

For the next lemma, we need a definition of finitude for concepts. We can provide a definition-schema:

Definition 28 (Finitude)**.** *X^n is finite exactly when there is a surjective function from a proper subset of the natural numbers to X^n. X^n is infinite iff it is not finite.*

Theorem 13 (Non-compact)**.** *Given any effective D spelled out in a language L at least of level $n \leq 2$, D is not compact under $+e$ raising semantics.*

Proof. Consider the following formula of L^k of order 2, where F is defined using the general definition of functions, and not the minimal one:

$$\exists F(\forall x, y(F(x) = F(y) \rightarrow x = y) \wedge \exists x \forall y(F(y) \neq x)) \qquad \text{(DED)}$$

The formula says that there is a function on the universe that is injective but not surjective. This is the standard definition of Dedekind's infinity. The formula does not contain any non-logical vocabulary. We can prove that any raising interpretation satisfies DED exactly when the domain is infinite. In $+e$ raising interpretation I, the first order domain is some concept D of order e+2 and, by definition of satisfaction, the above formula is satisfied exactly when some function F of order $e+2$ in the projected second order domain of I satisfies the conjunction of a and b, where $t = e + 1$:

$$\forall J^t, H^t((D(J^t) \wedge D(H^t) \wedge F(J^t) = F(H^t)) \rightarrow J^t = H^t) \qquad \text{(a)}$$

$$\exists J^t(D(J^t) \wedge \forall H^t(D(H^t) \rightarrow F(H^t) \neq J^t)) \qquad \text{(b)}$$

Clearly the formula says that F has domain D and is an injective but not surjective function. With a proof very close to the usual one in set-theory, we can show that a concept X^{e+2} (such as the domain of I) is infinite exactly when there is an injective but not surjective function on the extension of $e+2$, which proves the theorem. The proof critically requires a choice schema at order $e+2$. Similarly, the set-theoretic proof requires choice, and it is not provable in ZF alone. The proof also relies on our definition of projected domain and satisfaction of a quantified formula: an interpretation satisfies $\exists X^2\phi$ exactly when some concept in the second order projected domain satisfies ϕ; the second-order projected domain contains *any* concept that applies to concepts/objects in the first order domain. Without this assumption the proof would not go through (for example in Henkin semanatics). □

Since compactness fails, no logic can be strongly complete for raising validity.

Lemma 25. *Take any effective deductive system D which is sound for a semantic theory T. If D is complete for T, then D is compact under T (Gödel 1930).*

Proof. One direction of compactness is obvious. For the other direction, we reason by contraposition: we show that, assuming completeness, if Γ is not satisfiable, some finite subset of Γ is not satisfiable. Suppose $\vDash_T$ models the validity relation in T. If Γ is not satisfiable, then $\Gamma \vDash_T \phi \wedge \neg\phi$, for any ϕ. D is complete for T by assumption, so $\Gamma \vdash_D \phi \wedge \neg\phi$, where $\vdash_D$ is provability in D. Proofs can use only finitely many premises, so there must be a finite subset Δ of Γ such that $\Delta \vdash_D \phi \wedge \neg\phi$. D is sound by assumption, so $\Delta \vDash_T \phi \wedge \neg\phi$, where Δ is not satisfiable and finite. □

The result is immediate:

Lemma 26. *There is no effective and sound deductive system D spelled out in a language of level $n \leq 2$ that is strongly complete for $+e$ raising semantics.*

Proof. Suppose otherwise. Then by lemma 25, D is compact under the raising semantics, contrary to theorem 13. □

The lemma only shows failure of strong completeness: a more interesting result is failure of weak completeness. There are many ways of proving this. I follow a popular one which consists in showing that $\mathsf{2PA}$ is categorical using the notion of validity at hand. Incompleteness follows through Gödelian reasoning (see Shapiro 1991 or Boolos et al. 2002).

We are interpreting the signature of $\mathsf{2PA}$ (L_{2PA}), which contains a constant 0 for zero, a successor function s and the identity symbol. Addition and multiplication can

be defined as meta-abbreviations in 2PA. The language is of level 2. We assume a standard axiomatisation of 2PA, which includes all the axioms of PA and where the induction schema has been replaced with its second order counterpart:

$$\forall X((X(0) \wedge \forall m(X(m) \rightarrow X(s(m)))) \rightarrow \forall n X(n)) \tag{2IND}$$

Pick any $e > 1$. We define a model M^{+e} of the language of 2PA that raises the values by e orders similarly to how we did it for L^k. We pick models that raise the values exactly by e rather than up to e, to make the definitions simpler, and because the result is equivalent via lemma 22. The value of 0 in M^{+e} (written $M^{+e}(\ulcorner 0 \urcorner)$) is a unique X^{e+1}. The successor function is mapped to some function over the domain of $e+1$ concepts. Raising assignments are defined as before.

The definition of raising Sat^{+e} is extended to L_{2PA} in the obvious way. 2PA is finitely axiomatisable, so we call AR the conjunction of its axioms. Clearly there is a model for AR, because there is a model $\mathfrak{M}$ of 2PA in the usual set-based semantics where $\vDash$ under $\mathfrak{M}$ satisfies the Tarskian axioms for truth. Therefore, there is an equivalent model in higher order standard semantics via lemma 18 and, therefore, an equivalent raising $+e$ model via lemma 22.

Theorem 14. *Let M and N be two raising $+e$ models of* 2PA*. Then M and N are isomorphic.*

Proof. I only sketch the proof, which is essentially a rendering in our higher order terms of the famous proof by Dedekind. For readability, I use X, Y etc for concepts of order $e+1$. I call D_M and D_N the domains of M and N, respectively. First, define H of order $e+2$ to be successor closed (SC-CL) if:

$$\forall X, Y(H(X,Y) \rightarrow (D_M(X) \wedge D_N(Y))) \tag{a}$$
$$H(M(\ulcorner 0 \urcorner), N(\ulcorner 0 \urcorner)) \tag{b}$$
$$\forall X, Y(H(X,Y) \rightarrow H(M(\ulcorner s \urcorner)(X), N(\ulcorner s \urcorner)(Y))) \tag{c}$$

We obtain the following concept C by instantiating impredicative comprehension with the following formula:

$$\forall Z^{e+2}(\text{SC-CL}(Z^{e+2}) \rightarrow Z^{e+2}(X,Y))$$

C works as the intersection of all SC-CL concepts. The proof consists in showing that C is an isomorphism between M and N. We do that by showing the following, in order:

1. C is SC-CL;
2. $\lambda X.\exists Y C(X,Y)$ is D_M;
3. $\lambda X.\exists Y C(Y,X)$ is D_N;
4. C is functional;
5. The range of C is D_N;
6. C is injective;
7. C is surjective.

I sketch some of the sub-proofs and highlight the points of interest.

(1) is really a consequence of the definition of SC-CL.

- Something is SC-CL, for example the binary $R(X,Y)$ such that the domain of M applies to X and the domain of Y applies to N is successor closed. So, $C(M(\ulcorner 0 \urcorner), N(\ulcorner 0 \urcorner))$, by definition of C: C satisfies (b);
- Suppose C did not satisfy (c). Then some X and Y are such that $C(X,Y)$ and yet
  $\neg C(M(\ulcorner s \urcorner)(X), N(\ulcorner s \urcorner)(Y))$. So there must be some H that is SC-CL and $\neg H(M(\ulcorner s \urcorner)(X), N(\ulcorner s \urcorner)(Y))$. Yet, since $C(X,Y)$ so $H(X,Y)$, so H is not SC-CL, contrary to our assumption;
- A reasoning similar to the above works for (a).

For (2): Call E the $\lambda X.\exists Y C(X,Y)$. We show that $\forall X E(X) \leftrightarrow D_m(X)$. By definition of C, if $E(X)$ then $D_M(X)$. For the other direction: since M is a model for AR, it makes true 2IND. Also, E is in the range of second order quantifiers in M, by definition of Sat^{+e}: this is what crucially might fail in Henkin semantics, where the proof does not work. So, by definition of Sat^{+e} and M, the following holds, when we instantiate 2IND for E, for any (+e) assignment A:

$$\big(E(M(\ulcorner 0 \urcorner)) \wedge \forall Y(E(Y) \to E(M(\ulcorner s \urcorner)(Y)))\big) \to \\ \to \mathrm{Sat}^{+e}(\ulcorner \forall v U(v) \urcorner, M, \pi(A, E/U))$$

$E(M(\ulcorner 0 \urcorner))$, since $C(M(\ulcorner 0 \urcorner), N(\ulcorner 0 \urcorner))$. Suppose $E(X)$; then $C(X,Y)$ for some Y; since C is SC-CL, $C(M(\ulcorner s \urcorner)(X), N(\ulcorner s \urcorner)(Y))$, so $E(M(\ulcorner s \urcorner)(X))$. X was arbitrary so we generalise for all X of order $e+1$. The antecedent holds, therefore we obain:

$\mathrm{Sat}^{+e}(\ulcorner \forall v U(v) \urcorner, M, \pi(A, E/V))$. By definition of Sat^{+e}, we obtain the following equivalences:

$$\begin{aligned}
&\mathrm{Sat}^{+e}(\ulcorner \forall v U(v) \urcorner, M, \pi(A, E/U)) \\
\leftrightarrow &\forall Y (D_M(Y) \rightarrow \mathrm{Sat}^{+e}(\ulcorner U(v) \urcorner, M, \pi(A, E/U, Y/v))) \\
\leftrightarrow &\forall Y \big(D_M(Y) \rightarrow \pi(A, E/U, Y/v)(U)(\pi(A, E/U, Y/v)(v))\big) \\
\leftrightarrow &\forall Y (D_M(Y) \rightarrow E(Y))
\end{aligned}$$

So, we obtain the other direction: if $D_M(X)$ then $E(X)$.

About (3). We show first that C is functional for $M(\ulcorner 0 \urcorner)$. Then, we prove that if it is functional up to X, it is functional for $M(\ulcorner s \urcorner)(X)$. The proof is a rendering of that in Shapiro 1991, 83, point 2, translated from set-theory into the higher order language. Using similar proofs, and making use of the fact that N satisfies the induction and successor axioms, we can prove the other points, too. Whenever we bring up the fact that the model satisfies 2IND, we make use of the fact that the second order projected domain contains all concepts that apply to the first order domain of the interpretation. Whenever we use induction, we are relying on CP(e+2), which allows us to extend induction to order $e + 2$. □

Theorem 15. *There is no effective and sound deductive system D spelled out in a language of level $n \leq 2$ that is weakly complete for $+e$ raising semantics.*

Proof. By theorem 14, any two models of 2PA are equivalent. AR is the conjunction of the axioms of 2PA. I^{+e} is a model of 2PA iff $\mathrm{Tr}(\ulcorner AR \urcorner, I^{+e})$. Thus, given any model M^{+e} of 2PA, $\mathrm{Tr}(\ulcorner \phi \urcorner, M^{+e})$ exactly when, for any I^{+e}, $\mathrm{Tr}(\ulcorner AR \urcorner, I^{+e}) \rightarrow \mathrm{Tr}(\ulcorner \phi \urcorner, I^{+e})$, exactly when $\vDash^n_{+e} \ulcorner AR \rightarrow \phi \urcorner$.

Suppose the theorem fails: for some sound and effective D, if $\vDash^n_{+e} \ulcorner \phi \urcorner$ then $\vdash_D \ulcorner \phi \urcorner$. A function that takes you from $\ulcorner \phi \urcorner$ to $\ulcorner \psi \rightarrow \phi \urcorner$ for some fixed ψ is clearly recursive, so the set of formulae ϕ such that $\vdash_D \ulcorner AR \rightarrow \phi \urcorner$ is effectively enumerable. That is, the set ϕ of arithmetical validities is effectively enumerable. Yet, as a consequence of Gödel's first incompleteness theorem, the set of valid formulae of arithmetic is not effectively enumerable. □

A.3.4 Other

Here is an interesting fact about raising interpretations.

Lemma 27. *The proof for the completeness principle in* 2ZFCU + URS *in McGee 1997, 63-4 fails when we consider raising interpretations.*

Proof. Recall that the completeness principle states the following: there is no isomorphism from the pure sets to a proper initial fragment of the pure sets of some model of 2ZFCU. Assume we are considering raising interpretations, as well, like +1 interpretations. We can define what it means to be a pure set in 2ZFCU, by defining what it means to be part of the hierarchy of sets built from the empty set. Call this condition PS. By $I^{+1}(PS)$ we mean the extension of the condition 'x is a pure set' in I^{+1}. Call $|PS|$ the concept of being a pure set. If the completeness principle fails for raising models, there is an isomorphism from the pure sets to a proper initial fragment of the pure sets of some I^{+1} of 2ZFCU. I^{+1} 'thinks' that $|PS|$ is the domain of some set-sized model of ZFC. That is, some κ in I satisfies 'v is an inaccessible cardinal' in I^{+1} and each pure set x and κ satisfies $\ulcorner v \in V_u \urcorner$ in I^{+1}. The proof by Mcgee relies on constructing three functions:

- I: an isomorphism from $|PS|$ to the elements of V_κ;
- H: a one-one map from the universe (i.e. from order 1) to $|PS|$;
- J: the inclusion map from the domain of I^{+1} to $\wp V_\kappa$.

Now, in a standard interpretation when we build the composition $I \circ H \circ J$, we notice that this is an injection from the elements of V_k to its powerset, which contradicts Cantor's theorem. However, this result critically relies on the fact that 'the universe' includes each member of the domain of the model. The reasoning fails in the case of raising interpretations, however, where the domain is strictly larger than 'the universe': for example if it has cardinality order-2-many. So, it might be that $H(J(X))$ is undefined, because X is not an object. □

Bibliography

Aristotle (1964). Prior analytics. In Ross, D., editor, *Analytica priora et posteriora*. Oxford Clarendon Press.

Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. In *Philosophy, language, and artificial intelligence*, pages 241–301. Springer.

Beall, J. and Murzi, J. (2013). Two flavors of Curry's paradox. *The Journal of Philosophy*, 110(3):143–165.

Beall, J. C., Beall, J., and Restall, G. (2006). *Logical pluralism*. Oxford University Press on Demand.

Beaney, M. e. (1997). *The Frege Reader*. Blackwell Oxford.

Ben-Yami, H. (2013). Higher-level plurals versus articulated reference, and an elaboration of salva veritate. *Dialectica*, 67(1):81–102.

Bencivenga, E. (1986). Free logics. In *Handbook of philosophical logic*, pages 373–426. Springer.

Berlinski, D. and Gallin, D. (1969). Quine's definition of logical truth. *Nous*, pages 111–128.

Bonnay, D. (2014). Logical constants, or how to use invariance in order to complete the explication of logical consequence. *Philosophy Compass*, 9(1):54–65.

Boolos, G. (1984). To be is to be a value of a variable (or to be some values of some variables). *The Journal of Philosophy*, 81(8):430–449.

Boolos, G. (1985). Nominalist platonism. *The Philosophical Review*, 94(3):327–344.

Boolos, G. S. (1975). On second-order logic. *The Journal of Philosophy*, 72(16):509–527.

Boolos, G. S., Burgess, J. P., and Jeffrey, R. C. (2002). *Computability and logic*. Cambridge university press.

Burgess, J. P. (1999). Which modal logic is the right one. *Notre Dame Journal of Formal Logic*, 40:81–93.

Cantor, G. (1899). Letter to Dedekind. In van Heijenoort, J., editor, *From Frege to Gödel*, pages 113–17. Cambridge Mass: Harvard. English translation by Stefan Bauer-Mengenberg.

Cappelen, H. and Lepore, E. (2008). *Insensitive semantics: A defense of semantic minimalism and speech act pluralism*. John Wiley & Sons.

Chagrov, A. and Zakharyaschev, M. (1997). *Modal Logic*. Clarendon Press.

Chang, C. C. and Keisler, H. J. (1990). *Model theory*. Elsevier.

Cresswell, M. J. and Hughes, G. E. (1996). *A New Introduction to Modal Logic*. Routledge.

Davidson, D. (1967). Truth and meaning. In *Philosophy, Language, and Artificial Intelligence*, pages 93–111. Springer.

Davies, M. K. (1978). Weak necessity and truth theories. *Journal of Philosophical Logic*, pages 415–439.

Degen, W. and Johannsen, J. (2000). Cumulative higher-order logic as a foundation for set theory. *Mathematical Logic Quarterly: Mathematical Logic Quarterly*, 46(2):147–170.

Dowek, G. (1995). Lambda-calculus, combinators and the comprehension scheme. Technical report, Institut national de recherche en informatique et en automatique (INRIA).

Dummett, M. (1991). *Frege's philosophy of mathematics*. Harvard University Press.

Eder, G. (2016). Boolos and the metamathematics of quine's definitions of logical truth and consequence. *History and Philosophy of Logic*, 37(2):170–193.

Etchemendy, J. (1988). Models, semantics and logical truth. *Linguistics and Philosophy*, 11:91–106.

Etchemendy, J. (1990). *The Concept of Logical Consequence*. CSLI Publications.

Etchemendy, J. (2008). Reflections on consequence. In Patterson, D., editor, *New essays on Tarski and philosophy*, pages 263–299. Oxford: Oxford University Press.

Feferman, S. (1991). Reflecting on incompleteness. *The Journal of Symbolic Logic*, 56(1):1–49.

Feferman, S. et al. (1999). Logic, logics, and logicism. *Notre Dame Journal of Formal Logic*, 40(1):31–54.

Field, H. (2009). I—What is the normative role of logic? In *Aristotelian Society Supplementary Volume*, volume 83.1, pages 251–268. Wiley Online Library.

Florio, S. (2014). Untyped pluralism. *Mind*, 123(490):317–337.

Florio, S. and Linnebo, Ø. (2016). On the innocence and determinacy of plural quantification. *Noûs*, 50(3):565–583.

Florio, S. and Linnebo, Ø. (2021). *The many and the one: A philosophical study of plural logic*. Oxford University Press.

Frege, G. (1884a). *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Beaney, Michael. Translated as: 'The Foundations of Arithmetic' in Beaney 1997, 84-129.

Frege, G. (1884b). *Die Grundlagen der Arithmetik. Eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Beaney, Michael. Translated as: 'The Foundations of Arythmetic' in Beaney (1997, 84-129).

Frege, G. (1892). Über begriff und gegenstand. *Vierteljahresschrift für wissenschaftliche Philosophie*, 16:192–205. Translated as: 'On Concept and Object' in Beaney 1997, 181-193.

Gauker, C. (1997). Universal instantiation: A study of the role of context in logic. *Erkenntnis*, 46(2):185–214.

Genin, K. and Huber, F. (2021). Formal Representations of Belief. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2021 edition.

Gödel, K. (1930). Die vollständigkeit der axiome des logischen funktionenkalküls. *Monatshefte für Mathematik*, 37(1):349–60. Translated as: 'The completeness of the axioms of the functional calculus of logic' in Van Heijenoort 2002, 582-91.

Gómez-Torrente, M. et al. (1996). Tarski on logical consequence. *Notre Dame Journal of Formal Logic*, 37(1):125–151.

Gupta, A. (1978). Modal logic and truth. *Journal of Philosophical Logic*, pages 441–472.

Hájek, P. and Pudlák, P. (2017). *Metamathematics of first-order arithmetic*, volume 3. Cambridge University Press.

Halbach, V. (1997). Tarskian and kripkean truth. *Journal of Philosophical Logic*, 26(1):69–80.

Halbach, V. (2014). *Axiomatic theories of truth*. Cambridge University Press.

Halbach, V. (2018). The substitutional analysis of logical consequence. *Noûs*.

Halbach, V. (2020). Formal notes on the substitutional analysis of logical consequence. *Notre Dame Journal of Formal logic*, 61(2):317–339.

Hanson, W. H. (1997). The concept of logical consequence. *The philosophical review*, 106(3):365–409.

Hanson, W. H. (2006). Actuality, necessity, and logical truth. *Philosophical Studies*, 130(3):437–459.

Harman, G. (1884). Logic and reasoning. *Synthese*, 60(1):107–27.

Harman, G. (1972). Logical form. *Foundations of Language*, pages 38–65.

Hilbert, D. and Bernays, P. (1939). *Grundlagen der Mathematik, Volume II*. Springer, Berlin.

Hinman, P. G., Kim, J., and Stich, S. P. (1968). Logical truth revisited. *The Journal of Philosophy*, 65(17):495–500.

Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In Almog, J., Perry, J., and Wettstein, H., editors, *Themes From Kaplan*, pages 481–563. Oxford University Press.

Kerry, B. (1886). Über anschauung und ihre psychische verarbeitung (ii). *Vierteljahresschrift für wissenschaftliche Philosophie*, 10:419–67.

Kerry, B. (1887). Über anschauung und ihre psychische verarbeitung (iv). *Vierteljahresschrift für wissenschaftliche Philosophie*, 11:249–307.

Kreisel, G. (1967). Informal rigour and completeness proofs. In *Studies in Logic and the Foundations of Mathematics*, volume 47, pages 138–186. Elsevier.

Kripke, S. (1976). Outline of a theory of truth. *The Journal of Philosophy*, 72(19):690–716.

Levy, A. et al. (1960). Axiom schemata of strong infinity in axiomatic set theory. *Pacific journal of mathematics*, 10(1):223–238.

Lewis, D., editor (1991). *Parts of classes*. Oxford: Blackwell.

Linnebo, Ø. (2006). Sets, properties, and unrestricted quantification. In *Rayo and Uzquiano (2006)*, pages 149–78. Oxford University Press.

Linnebo, Ø. and Nicolas, D. (2008). Superplurals in english. *Analysis*, 68(3):186–197.

Linnebo, Ø. and Rayo, A. (2012). Hierarchies ontological and ideological. *Mind*, 121(482):269–308.

MacFarlane, J. (2009). Nonindexical contextualism. *Synthese*, 166(2):231–250.

MacFarlane, J. (2017). Logical constants. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition.

Makinson, D. (1966). On some completeness theorems in modal logic. *Mathematical Logic Quarterly*, 12(1):379–384.

McGee, V. (1992). Two problems with tarski's theory of consequence. In *Proceedings of the Aristotelian Society*, volume 92, pages 273–292. JSTOR.

McGee, V. (1996). Logical operations. *Journal of Philosophical Logic*, 25(6):567–580.

McGee, V. (1997). How we learn mathematical language. *The Philosophical Review*, 106(1):35–68.

McKeon, M. (2004). On the substitutional characterization of first-order logical truth. *History and Philosophy of Logic*, 25(3):205–224.

Miyazaki, Y. (2007). Kripke incomplete logics containing ktb. *Studia Logica*, 85(3):303–317.

Montague, R. (1961). Fraenkel's addition to the axioms of zermelo. In *Essays on the foundations of mathematics*, page 91.114. Hebrew Univ., Jerusalem: Magnes Press.

Montague, R. (1970). Universal grammar. *Theoria*, 36(3):373–398.

Murzi, J. and Rossi, L. (2017). Naïve validity. *Synthese*, pages 1–23.

Oliver, A. and Smiley, T. (2016). *Plural logic: revised and enlarged*. Oxford University Press.

Peacocke, C. (1978). Necessity and truth theories. *Journal of Philosophical Logic*, 7(1):473–500.

Priest, G. (1979). The logic of paradox. *Journal of Philosophical logic*, 8(1):219–241.

Quine, W. V. (1986). *Philosophy of logic*. Harvard University Press.

Ray, G. (1996). Logical consequence: A defense of tarski. *Journal of Philosophical Logic*, 25(6):617–677.

Rayo, A. (2006). Beyond plurals. In *Absolute generality*, pages 220–254. Oxford University Press Oxford.

Rayo, A. and Uzquiano, G. (1999). Toward a theory of second-order consequence. *Notre Dame Journal of Formal Logic*, 40(3):315–325.

Rayo, A. and Uzquiano, G. (2006). *Absolute Generality*. Oxford University Press.

Rayo, A. and Williamson, T. (2003). A completeness theorem for unrestricted first-order languages. In Beall, J., editor, *Liars and heaps: New essays on paradox*, pages 331–356. University of Hawaii Press.

Rey, G. (2018). The analytic/synthetic distinction. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2018 edition.

Rubin, H. and Rubin, J. E. (1963). *Equivalents of the Axiom of Choice*. Elsevier.

Rumfitt, I. (2015). *The boundary stones of thought: An essay in the philosophy of logic*. Oxford University Press.

Russell, G. (2018). Logical nihilism: could there be no logic? *Philosophical Issues*, 28:308–324.

Russell, G. (2019). Logical Pluralism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition.

Sagi, G. (2018). Logicality and meaning. *The Review of Symbolic Logic*, 11(1):133–159.

Shapiro, S. (1987). Principles of reflection and second-order logic. *Journal of Philosophical Logic*, 16(3):309–333.

Shapiro, S. (1991). *Foundations without foundationalism: A case for second-order logic*, volume 17. Clarendon Press.

Shapiro, S. (1998). Logical consequence: Models and modality. In Schirn, M., editor, *The philosophy of mathematics today*, pages 131–156. Oxford: Oxford University Press.

Sher, G. (1991). *The Bounds of Logic: A Generalized Viewpoint*. MIT Press.

Sher, G. (2001). The formal-structural view of logical consequence. *The Philosophical Review*, 110(2):241–261.

Sher, G. Y. (1996). Did Tarski commit 'Tarski's fallacy'? *The Journal of Symbolic Logic*, 61(2):653–686.

Stalnaker, R. C. (1978). Assertion. *Syntax and Semantics*, 9:315–332.

Stanley, J. and Gendler Szabó, Z. (2000). On quantifier domain restriction. *Mind & Language*, 15(2-3):219–261.

Studd, J. (2019). *Everything, more or less: A defence of generality relativism*. Oxford University Press.

Studd, J. (2021). Infinite types and the principle of union. In *Modes of Truth*, pages 266–290. Routledge.

Tarski, A. (1933). *Pojęcie prawdy w językach nauk dedukcyjnych*. Warsaw: Nakładem Towarzystwa Naukowego Warszawskiego. Translated in German in Tarski 1935.

Tarski, A. (1935). Der wahrheitsbegriff in den formalisierten sprachen. *Studia Philosophica*, 1:261–405. Translation of Tarski 1933 by L. Blaustein, with a postscript added; translated in English as 'The concept of truth in formalized languages' in Tarski 1983, 152-278.

Tarski, A. (1936a). O pojęciu wynikania logicznego. *Przegląd Filozoficzny*, 39:58–68. Translated as: 'On the concept of logical consequence' in Tarski 1983, 409-20.

Tarski, A. (1936b). Über den begriff der logischen folgerung. *Actes du Congrès International de Philosophie Scientifique, fasc. 7 ((Actualités Scientifiques et Industrielles)*, 394:1–11. German version of Tarski 1936a; translated as: 'On the concept of logical consequence' in Tarski 1983, 409-20.

Tarski, A. (1983). *Logic, semantics, metamathematics: papers from 1923 to 1938*. Oxford Clarendon.

Tarski, A. (1986). What are logical notions? *History and philosophy of logic*, 7(2):143–154.

Uzquiano, G. (2003). Plural quantification and classes. *Philosophia Mathematica*, 11(1):67–81.

Uzquiano, G. (2015). Recombination and paradox. *Philosopher's Imprint*, 15(19).

Väänänen, J. (2001). Second-order logic and foundations of mathematics. *Bulletin of Symbolic Logic*, pages 504–520.

Van Heijenoort, J. e. (2002). *From Frege to Gödel: a source book in mathematical logic, 1879-1931*. Harvard University Press.

Wallace, J. (1975). Non-standard theories of truth. In Davidson, D. and Harman, G., editors, *The Logic of Grammar*, pages 50–59. Dickenson.

Walsh, S. and Button, T. (2018). *Philosophy and Model Theory*. Oxford, UK: Oxford University Press.

Warmbrod, K. (1999). Logical constants. *Mind*, 108(431):503–538.

Williamson, T. (1999). Logic and existence. *Aristotelian Society Supplementary Volume*, 173 (1):181–203.

Williamson, T. (2003). Everything. *Philosophical Perspectives*, 17.1:415–465.

Williamson, T. (2013). *Modal logic as metaphysics*. Oxford University Press.

Williamson, T. (2017). Semantic paradoxes and abductive methodology. In Armour-Garb, B., editor, *Reflections on the Liar*, pages 325–346. Oxford University Press.

Woods, J. (2017). Characterizing invariance. *Ergo*, 3:778–807.

Yi, B.-U. (2005). The logic and meaning of plurals. part I. *Journal of Philosophical Logic*, 34(5-6):459–506.

Zalta, E. N. (1988). Logical and analytic truths that are not necessary. *The Journal of Philosophy*, 85(2):57–74.