

2012-07-13

A study in the logic of institutions

Payette, Gillman

Payette, G. (2012). A study in the logic of institutions (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from <https://prism.ucalgary.ca>. doi:10.11575/PRISM/25040
<http://hdl.handle.net/11023/115>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

A Study in the Logic of Institutions

by

Gillman Payette

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

CALGARY, ALBERTA

June, 2012

© Gillman Payette 2012

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a Dissertation entitled “A Study in the Logic of Institutions” submitted by Gillman Payette in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY.

Supervisor, Dr. Richard Zach, Department of Philosophy

Dr. Nicole Wyatt, Department of Philosophy

Dr. John Baker, Department of Philosophy

Dr. Robert Kremer, Department of Computer Science

External Examiner, Dr. Allen P. Hazen, University of
Alberta

Date

Abstract

In my dissertation *A Study in the Logic of Institutions* I develop a logical system for reasoning about institutions and their consistency. Since my dissertation is a work in logic rather than one in socio-political philosophy, I don't defend a particular theory of institutions. Instead, I did as Yogi Berra suggested and simply took the fork in the road. A well-developed account of institutions is given by John Searle in (1995); and (2010). His account bases all social reality on language, and I use his account to provide a logic for institutional norms.

Briefly, social reality is constructed via language by making our intentions clear to one another. And we do this via speech acts. There is one particular type of speech act that is important to institutions: declarations. Declarations bring about new social objects and create social states of affairs. It is via declarations that social institutions are created. In so far as groups recognize an institution sustaining/making authority, that authority has the ability to generate new institutional rules via declarations.

According to Vanderveken (1990, 1991); see also Searle and Vanderveken (1985), speech acts have a logic. That is, performing one speech act can satisfy the conditions of having performed another speech act. A priest declaring a baby baptized will also make it so that the priest has asserted that the baby is baptized, for instance. More importantly, certain declarations will result in the declarations of some of the logical consequences of the initial declarations. I characterize the set of speech acts that stand in that relationship and develop a logical system around that characterization.

The formal framework incorporates action and permits representations of complex institution-dependent relations, e.g., rights and duties. I further develop this formalism to investigate the notion of normative consistency. I show how to represent at least a minimal conception of normative inconsistency within the formal framework, and characterize its properties. I conclude by comparing my work to that of others.

Acknowledgements

Nothing you write is ever as bad as you fear or as good as you hope.

Bertrand Russell.

I would first like to thank Pam. She has been patient with me while she was writing a book under a crazy deadline, and I lead her to believe I would be finished much sooner, and able to support her writing the way she supported mine. For that I am sorry. I would also like to thank my supervisor Richard Zach for his commitment to getting me funded through tireless reading and rereading drafts of my research proposals to SSHRCC (and others). That help made my life very comfortable as a PhD student, it also helped me to figure out what my dissertation project was. I would like to thank Richard for his extensive comments on my thesis and his help in my work generally.

I would also like to thank the Social Sciences and Humanities Research Council of Canada for both my Joseph-Armand Bombardier Canadian Graduate Scholarship and my Michael F. Smith Foreign Study Supplement which allowed me to spend time at the Institute for Language Logic and Computation in Amsterdam hosted by Johan van Benthem. I would also like to thank the Killam Foundation for the Honourary Killam Fellowship, it is an honour to be a Killam Scholar. I would also like to thank the John D. Petrie and Harry and Laura Jacques bursaries for funding this last year of writing. I must also thank three people for many discussions of my dissertation topic: Ann Levey, David Dick, and Allen Habib. I would very much like to thank my committee for their comments on earlier drafts of the first part of the thesis, without those comments it would be unintelligible: Thank you Nicole Wyatt and John Baker. I also want to thank Peter Schotch for being a supporter of mine. Without him I would not have gotten to this point. To all, again, Thank you.

I dedicate this work to my mother Cathy Bennett

Table of Contents

Abstract	ii
Acknowledgements	iii
Dedication	iv
Table of Contents	v
List of Tables	viii
List of Figures	ix
List of Symbols, Abbreviations, Nomenclature	x
1 Introduction	1
1.1 Some Remarks on Methodology	1
1.2 Fixing Some Notation and Terminology	2
1.3 Outline of the Essay	3
 I Philosophical Foundations of Institutions	 6
2 Foundational Considerations	7
2.1 Preservationist Consequence	7
2.2 Institutions and Consequence	9
2.2.1 Institutions	9
2.2.2 Norms	10
2.3 The Conception Matters	12
2.3.1 In Forceness Matters	12
2.3.2 Norms Matter	14
2.4 Historical Precedent	16
2.5 Our Choice	17
3 Searle on Social Institutions	19
3.1 Speech Acts	19
3.1.1 Speech Act Theory	20
3.1.2 The Specific Speech Acts	26
3.2 The Components of Searle's Conception of Institutions	29
3.3 Putting it Together	33
4 The Justification of Normative Consequence	37
4.1 Speech Act Entailment	38
4.2 Illocutionary Entailment and Declarations	40
4.3 Vanderveken on Strong Implication	42
4.4 Strong Implication and the Specific Illocutionary Acts	47
4.5 Summary	55
 II Formal Logic and Formal Dynamics of Institutions	 56
5 Constructing a Formal Language	57
5.1 A Formal Language of Social Reality	57
5.1.1 The Xstit Formalism	59
5.1.2 Institutional Facts and Roles	69

5.1.3	A Special Institutional Fact: The Violation Constant	73
5.2	Illocutionary Entailment	76
5.3	Completeness of \vdash_{SI}	82
6	Formalization of Normative Entailment	87
6.1	Introduction	87
6.2	Formal Languages for Institutions	88
6.3	Logic of \mathcal{L}_{\in}^I	91
6.3.1	Semantics for \mathcal{L}_{\in}^I	91
6.3.2	Proof Theory	95
6.3.3	Soundness and Completeness	99
6.4	How to Say Things Without Words: Expressing Legal Relations	103
6.5	Normative Entailment	107
6.6	Institutions and “The World”	113
6.7	Responses to Some Objections	120
7	Completeness of \vdash_{Ixp}	122
7.1	Completeness of \vdash_{xp}	122
7.1.1	Regular Models of \mathcal{L}	123
7.1.2	Neutral Models of \mathcal{L}	125
7.1.3	Completeness Relative to NU	131
7.1.4	Different Classes	143
7.1.5	The Same Old Models	153
7.2	Completeness of \vdash_{xp}^I	173
7.3	Completeness of the Rest: \vdash_{Ixp}	174
7.4	Proof of Proposition 6.6.1	179
8	Normative Consistency	182
8.1	Actual Inconsistencies	182
8.2	(In)Consistency of Codes	184
8.2.1	Von Wright	185
8.2.2	Hamblin	186
8.3	Formal Account of Normative Consistency	187
8.4	Proofs from Section 8.3	199
9	Reflections on the Logic	205
9.1	Formal Accounts of Norms	206
9.1.1	Other Work on Action	206
9.1.2	The Input/Output Camp	209
9.1.3	Grossi’s Formalization of Searle	212
9.1.4	Some Interpretation of Grossi’s Work	217
9.2	Philosophical Evaluation	221
9.2.1	Applicability to other Institutions and Systems of Norms	221
9.2.2	Norms: Interpretation and Defeasibility	222
9.3	Logical Evaluation	224
9.3.1	Future Directions	224
9.3.2	Aphilosophicality	225
9.4	Conclusion	226
	Bibliography	228
A	Languages, Models, and Logics	236

A.1	Languages	236
A.2	Models	236
A.3	Logics	238
B	Some Background	245
B.1	Kripke Models	245
B.2	Boolean Algebras	245
B.3	Fusion of Logics	246

List of Tables

6.1	Target Expressions	104
6.2	Institutional Relations	107

List of Figures

5.1	Universal, Regular \mathcal{L} -model	62
6.1	The Relation Between Institutions and Reality	113
6.2	Realization of Implementation	119
7.1	$\text{lub}(s)$ in a neutral model	130
7.2	Stretching Out h to h_N	145
7.3	$\text{lub}(h_{s+k})$ in \mathfrak{M}_K	163
7.4	Defining \approx_K in \mathfrak{M}_K	164
7.5	Counter Example to Indep-G	173

List of Symbols, Abbreviations, Nomenclature

- $A \cap B$ $\{x : x \in A \ \& \ x \in B\}$, page 2
- $A \cup B$ $\{x : x \in A \text{ or } x \in B\}$, page 2
- $A \subseteq B$ A is a subset of or equal to B , page 2
- Ax_i A Hilbert style axiomatisation of a logic L_i based on a language \mathcal{L}_i , page 246
- Cn Cn is the consequence operator of classical propositional logic, page 209
- D The domain from \mathcal{D} , page 77
- D_A The set of atoms from \mathcal{D} , page 77
- H The set of histories $\langle h, <_h \rangle$ from an xstit frame, page 60
- $P\varphi$ The previous state operator from \mathcal{L} , page 89
- R^+ If R is a relation this denotes the transitive closure, page 125
- S The set of static states from an xstit frame, page 60
- V The violation constant, page 74
- W_C The context of a context C in a context frame, page 214
- $[C] \langle C \rangle$ The universal context operator: for all worlds in the context $C \dots$, and the existential context operator: there is a world in the context such that \dots , page 214
- $AFT(s, h)$ $\{(s', h') \in \mathfrak{M} : s \in h' \ \& \ s \leq_{h'} s'\}$, page 188
- Ag, A** The set of agents for $\mathcal{L}_{\text{xstit}}$, and a set of agents/agent terms $\mathbf{A} \subseteq \mathbf{Ag}$, respectively, page 59
- At_B, At_I** The brute atomic sentences and institutional atomic sentences, respectively, page 70
- \mathbb{C}_{CL} Classical consequence operator, page 2
- DDL_F** A double discrete line function, page 127
- D_L_F** The class of discrete line functions, page 126
- Δ_V $\{\neg \Diamond X^n \Box X V \mid n \in \mathbb{N}\}$, page 197
- Δ_{if} $\{\Box X^n \delta \mid \delta \in \Delta \ \& \ n \in \mathbb{N}\}$, page 196
- $E(s, h, \mathbf{A})$ An effectivity function for \mathbf{A} evaluated at (s, h) , page 60
- \mathcal{L}_P The language of pure Boolean formulas, page 2
- \mathcal{L}_S The language for strong implication, page 76

\mathcal{L}_i	A modal language \mathcal{L}_i with a modal operator \Box_i , page 246
\mathcal{L}	An extension of $\mathcal{L}_{\text{xstit}}$, page 88
\mathcal{L}^B	The brute fragment of \mathcal{L} , page 89
\mathcal{L}^I	An revision of \mathcal{L} that uses role terms instead of agent terms, page 89
\mathfrak{F}	A frame for xstit, page 60
$\Gamma; \varphi$	This notation means the same as $\Gamma \cup \{ \varphi \}$, page 83
IDDLF	The class of IDDLFs, page 128
\iff	iff Metalanguage ‘if and only if’, page 2
\mathcal{I}	An <i>SI</i> -model or interpretation, page 77
$Ic(\varphi)$	The metalanguage institutional control predicate and function, page 108
$\mathfrak{I}(\Omega)$	An <i>implementation</i> of Ω , page 115
$\mathfrak{M}, (s, h) \triangleleft \Delta$	Δ is sustained in force after (s, h) in \mathfrak{M} , page 188
\mathfrak{M} / \approx	The quotient model of a Kamp \mathcal{L} -model \mathfrak{M} , page 161
\mathfrak{M}_K	A Kamp model generated from an NUZ \mathcal{L} -model \mathfrak{M} , page 161
NUZ	The class of NU models whose histories are all infinite IDDLFs, page 153
\Rightarrow	Metalanguage ‘only if’, page 2
Rol	The set of institutional role terms, page 70
E-Relative Normative Entailment	See definition 6.5.4, page 111
\approx	The modal alternativeness relation from a neutral \mathcal{L} -model, page 128
\approx^o	The equivalence relation for the canonical NU model, page 138
\mathbb{Z}	The set of integers $\{ \dots - 2, -1, 0, 1, 2, \dots \}$, page 61
\mathcal{D}	The po-set of propositional contents, page 77
\lesssim	The partial order from \mathcal{D} , page 77
\odot	The global infimum for \mathcal{D} , page 77
CON_N	Normative consistency predicate, page 199
$\text{CON}_N^{\mathfrak{I}}$	Normative consistency relative to the implementation \mathfrak{I} , page 198
\equiv	material biconditional, page 2
$\&$	Metalanguage ‘and’, page 2

- $glb(s)$ $\{ glb(s, h) : s \in h \ \& \ h \in H \}$, page 60
- $glb(s, h)$ The h -predecessor of s , page 60
- holds specifies which agents/groups hold which roles in an implementation of an institution, page 115
- $lub(s)$ $\{ lub(s, h) : s \in h \ \& \ h \in H \}$, page 60
- $lub(s, h)$ The h -successor of s , page 60
- $IC(\mathcal{L}^I)$ The set of sentences in \mathcal{L}^I that are under institutional control, page 109
- $\max^\vdash(\mathcal{L})$ The set of \vdash -maximally consistent subsets of \mathcal{L} , page 133
- $\Box\varphi, X\varphi, P\varphi$ φ is settled true, φ is true in the next state, φ is true in the previous state, respectively, page 59
- π a partition of **Rol**, page 115
- $\pi_{\mathbf{r}}$ The cell in the partition of **Rol**, π that contains \mathbf{r} , page 115
- \Rightarrow^C The proper-classificatory count-as conditional for the context **C**., page 215
- $\llbracket \cdot \rrbracket$ The semantic interpretation of \cdot , page 38
- $\llbracket \varphi \rrbracket_1, \llbracket \varphi \rrbracket_2$ The first and second coordinates of the semantic value of φ , page 78
- $\{x : \varphi(x)\}$ The set of objects satisfying φ , page 2
- \supset material conditional, page 2
- Ctx**, **C** The set of context variables for a context logic, and a context variable, page 214
- $\theta \subseteq \theta'$ The content of θ is contained in θ' , page 76
- \top a symbol for the logical constant that is always true., page 211
- \models_{Ixp} The semantic consequence relation for the logic of $\mathcal{L}_{\subseteq}^I$, page 95
- \models_{NU} The semantic consequence relation for neutral \mathcal{L} -models, page 130
- \vdash_{Ixp} The syntactic consequence relation for the logic of $\mathcal{L}_{\subseteq}^I$, page 95
- \vdash_N Norm Consequence (syntatic), page 112
- \vdash_N Norm Consequence (syntatic), page 244
- \vdash_N^{Ξ} Ξ -Relative Norm Consequence (syntatic), page 111
- \vdash_{CL} Classical consequence relation, page 2
- \vdash_{xp} The syntactic consequence relation for the language \mathcal{L} , page 99

- \vdash_{xp}^I The syntactic consequence relation for \mathcal{L}^I , page 101
- \vdash_{xp}^Ω is the consequence relation \vdash_{xp} restricted to the language of \mathcal{L}_Ω^B , page 102
- $[A \text{ xdstit}] \varphi$ The x-deliberative stit operator, page 68
- $[A \text{ xstit}] \varphi$ A sees-to-it-that φ is true in the next state, page 59
- $a \in A$ a is an element of A , page 2
- $at(\varphi)$ The set of atomic sentences in φ , page 89
- $d \vee d'$ The supremum of d and d' in \mathcal{D} , page 77
- v A valuation function, page 63
- F** A term that is interpreted as an illocutionary force, page 38
- Bad Situation **A** is in a bad situation at (s, h) when $E(s, h, \mathbf{A}) \subseteq \llbracket V \rrbracket$, page 189
- constitutive norm A constitutive norm defines a new action or object, page 32
- D-quandary Discriminatory Quandary is where one group is always in a bad situation, page 193
- declaration* The kind of declaration that Searle needs for his theory of institutions, page 47
- G-quandary Global quandary, page 193
- GR General Role: a cell in a partition of **Rol** such that all of the role terms in the cell instantiate a kind of role, e.g., citizen, page 72
- IDDLF Injective DDLF, page 128
- Kamp The property of an **NUZ** model where $\forall s, s' \in h \in H, s \approx s'$ only if $s = s'$, page 157
- master book The collection of all legal texts, page 9
- master system An interpretation of all legal texts, page 10
- MinCon Minimal consistency of the speaker, page 43
- Norm Content The content of a norm is what it says to do or bring about, page 12
- norm formulation A norm formulation is a sentence that is used to express a norm, page 10
- Normative Entailment See definition 4.4.1, page 47
- Normative Entailment See definition 6.5.4, page 112
- regulative norm A norm that regulates preexisting actions, page 32
- StComp Compatibility of strong implication with non-empty illocutionary points, page 43

T-quandary A total quandary is where there are no dynamic states after a certain point that aren't violation states., page 191

Universal Frame/Model A universal \mathcal{L} model is one where all of the histories are coincident at some point, page 61

Chapter 1

Introduction

The great thing about being a philosopher or a student of philosophy... [is that] it gives you a kind of licence to stick your nose into absolutely everything.

A.C. Greyling

The purpose of this essay is to raise and answer two questions: α) Is a logic of institutional norms possible?, and β) Given that a logic of institutional norms is possible, what does it look like? This introduction serves to frame our view of the methodology of philosophical logic, fix some terminology, and outline the progression of the essay.

1.1 Some Remarks on Methodology

The foundations of deontic and imperative logic, which are both often referred to as logics of norms, is one of the most contentious topics in philosophical logic. There is little agreement on how to approach the topic. This uncertainty suggests approaching the topic of a logic of norms in a manner rather different from much philosophical logic.

Generally, philosophical logics (modal logic, epistemic logic, temporal logic) are thought of in line with scientific theories, perhaps on a naive view of scientific theories. There is some set of data (the intuitively correct and incorrect inferences), and logicians attempt to find a theory that respects that set.¹ But we can raise the question: what is it that justifies those intuitive judgments in the first place? If logic is supposed to be indubitable, we need a better foundation.

Part of the reason for going after the inferences that we intuit to be right and wrong is to offer a theory that can be used in evaluating arguments that doesn't beg any questions. Such logical theories are meant to be aphilosophical, i.e., philosophically neutral. But even that seems to be problematic, specifically in deontic logic. Sayre-McCord (1986), for instance, argues that

¹See van Benthem (1983) for a more detailed account.

the axioms of many deontic logics take a stand on the possibility of conflicting obligations and dilemmas by saying that they are logically contradictory; thus, there can be no true deontic dilemmas. Such a stand represents a substantial ethical position.

The better way is to roll with the punches, reject the temptation to develop aphiosophical logics, and instead devise logics that formalize specific philosophical theories. This is similar to the project of Carnapian explication. Such a project has the advantage of not needing to match everyone’s philosophical views.

Of course it is important within these projects to recognize that not every detail will be taken care of by a philosophical theory, so sometimes intuitive ideas sneak back in, but when they do we notice that they are just naive philosophical theories, not *the way things really are*. Given these preliminary remarks, we will fix some notation.

1.2 Fixing Some Notation and Terminology

In the following we will use \vdash_{CL} and \mathbb{C}_{CL} to stand for the classical consequence relation and operator, respectively, over the language generated by the grammar

$$\varphi := \mathbf{p} \mid \varphi \wedge \varphi \mid \neg\varphi \mid \varphi \supset \varphi \mid \varphi \vee \varphi \mid \varphi \equiv \varphi$$

where $\mathbf{p} \in \mathbf{At}$ is an atomic sentence and \mathbf{At} is a set of atomic sentences. The symbol \supset is always interpreted as a material conditional. Similarly \equiv is the material biconditional. We call this language of pure Boolean formulas \mathcal{L}_P . We will often make use of some metalanguage notation as well: $\&$ for ‘and’, \Rightarrow for ‘only if’, \iff and iff for ‘if and only if’.

The standard mathematical symbols from set theory will be used: subset or equal to \subseteq , element of \in , union \cup , intersection \cap , and relative complement \setminus , as well as the set abstraction notation: $\{x \in A : \varphi(x)\}$. We will be assuming a underlying theory of ZFC set theory to go along with this investigation—yes, sometimes we need the axiom of choice. We will assume that the reader is familiar with the idea of a normal modal logic, and Kripke semantics for modal logics. Appendix B.1 reviews these topics.

Since we are talking about many different logics in this essay, we should note that each logic will give rise to a consequence relation, and each consequence relation \triangleright between sets of sentences and sentences from a language \mathcal{L} will have a corresponding consequence operator $\mathbb{C}_{\triangleright}(\Gamma) =_{df} \{ \varphi \in \mathcal{L} : \Gamma \triangleright \varphi \}$. \mathbb{C}_{CL} above is simply a specific case of this when $\triangleright = \vdash_{CL}$. We will define the notion of a subformula very generally for a formal language as follows:

Definition 1.2.1. If φ , θ and ψ are formulas in the language and $*$ is a binary connective, while $@$ is a unary connective, then the set of subformulas of θ , denoted $sub(\theta)$ is defined as follows:

1. $\theta = \mathbf{p}$, $sub(\mathbf{p}) = \{ \mathbf{p} \}$,
2. $\theta = @ \varphi$, $sub(@ \varphi) = sub(\varphi) \cup \{ @ \varphi \}$, and
3. $\theta = \psi * \varphi$, $sub(\psi * \varphi) = sub(\psi) \cup sub(\varphi) \cup \{ \psi * \varphi \}$.

Now that we have fixed this terminology we will outline the essay.

1.3 Outline of the Essay

The steps involved in answering questions α and β are very different. To answer α we need a philosophical discussion; to answer β we need to construct a logic/formal system. Since the answers are so different we have divided the essay into two parts.

Part I goes about answering ‘is a logic of institutional norms possible?’ in a rather round-about manner. First of all, by ‘possible’, we mean: is it epistemically possible for there to be a logic of institutional norms? The issue we confront with this question is a kind of scepticism regarding logics of norms. Scepticism about logics of norms can take two forms. First, one might be sceptical about the existence of norms at all, so a fortiori, the sceptic is sceptical about a logic of norms. Second, the sceptic may think that the logic of norms is trivial. By a trivial logic we mean a logic whose consequence operator is such that $\mathbb{C}(\Gamma) = \Gamma$ for any set of sentences Γ .

A motivation for thinking that the logic of institutional norms is trivial is to think of the real norms as simply those that are explicit in the institutional system. So in chapter 2 we argue that

we should look to investigate a particular theory about the logic of institutional norms. If we can then argue that that account of institutions gives rise to a non-trivial consequence relation, then we can argue that a logic of institutional norms is possible. That possibility hinges on the possibility and plausibility of the account of institutions on which it is based, but that is fine. We also focus on showing what is involved in any account of a logic of institutional norms. We end the chapter by pointing in the direction that we are headed, viz. Searle's theory of institutions.

Chapter 3 is primarily an exposition of Searle's account of institutions. Roughly, institutions are socially recognized norms, but they come about through various forms of symbolic representations. That means institutions come into existence in a manner with the same logical form as speech acts. Speech acts, however, do have a logic. That logic is discussed at length in Searle and Vanderveken (1985), and Vanderveken (1990, 1991). We follow Vanderveken (1990, 1991) for our account of speech act logic. We argue that the consequence relation for a logic of norms based on Searle's theory is a subrelation of classical logic, but it is non-trivial. The relation is what Vanderveken calls 'strong implication'. These philosophical discussions provide a philosophical theory that we can represent formally.

This brings the investigation to trying to get a handle on what the logic of institutional norms might look like, i.e., question β . In chapter 5 at the beginning of part II we develop all of the relevant components of a language for representing Searle's account of institutions. This involves an account of action in the form of xstit logic, Anderson's reduction of deontic logic to alethic modal logic with a violation constant, and how to formalize the relation of strong implication. We then use these pieces to generate a logic of norms.

We start chapter 6 by defining the various languages that we will use to define the consequence relation for a logic of institutional norms. Then we develop an extension of the xstit language and logic to represent the various pieces of reality: the institutional and the non-institutional or brute. We show how to interpret the various concepts of duty, right, privilege, etc. into our formal language. Finally, we reconstruct a notion of normative consequence and demonstrate how the institutional language and the brute language relate.

Chapter 7 is a completeness proof of the various logics introduced in chapter 6. We show that the logic is complete, and we draw some connections between this work and previous formal work in the logic of historical necessity.

The penultimate chapter is an application of our formal system. The notion of normative inconsistency hasn't been given a detailed study by moral philosophers and deontic logicians, for the most part.² As Donald Davidson once said concerning normative consistency: "It is astonishing that in contemporary moral philosophy this problem has received little attention and no satisfactory treatment" (Davidson, 1970, p. 105). Normative consistency is often reduced to consistency in standard deontic logic, but that is an uninteresting kind of consistency; it is plain logical consistency. That result issues from the limitations of the formalism of standard deontic logic. We show at least a minimal way of interpreting normative inconsistency within our formal framework, and prove some results about that interpretation.

Our final chapter deals with some questions of philosophical interpretation of our formal system, and comparison to other work in the area. We consider a number of formal works on the logic of institutions, but we focus on comparing our work with that of Stolpe (2008a), and Grossi (2007). These latter works relate most closely with ours. We argue that our work is superior in certain respects to that of Grossi (2007), and it doesn't succumb to problems raised by Stolpe (2008a).

²Hamblin (1972), Marcus (1980) and von Wright (1991) are some exceptions.

Part I

Philosophical Foundations of Institutions

Chapter 2

Foundational Considerations

When you come to a fork in the road, take it.

Yogi Berra

In this first part we set out to answer the first of our two questions: Is a logic of norms possible? Essentially, an answer to an “is there a logic of...” question is to provide a notion of consequence for the ‘...’ of interest. Roughly, our notion of consequence for institutions is that φ is a consequence of Γ whenever the conditions for all of the norms in Γ to be norms of an institution are met, then those conditions are also met for φ . This formulation needs to be expanded upon and explained, and that is our task in this first chapter.

2.1 Preservationist Consequence

We take what is called a preservationist view to consequence. The preservationist idea is that logic, generally construed, is concerned with understanding the way that conditions on sets of sentences transfer between sets of sentences. So any conditions whatsoever may be of logical interest, especially if we can find general principles to characterize that transfer. One property that has been a focus of logic is truth. What we teach in any first year logic course is the notion of validity: an argument from premises Γ to a conclusion φ is valid iff whenever all of the premises in Γ are *true*, then so is the conclusion φ . Of course a set of sentences isn’t true, so we define another notion called satisfaction, i.e., Γ is satisfied when all of the sentences in Γ are true. The preservationist program is to generalize this age-old notion of validity.

The preservationist generalization of validity is given as follows: $\{\varphi\}$ is a ϑ -consequence of Γ iff Γ has the property ϑ only if $\{\varphi\}$ has the property ϑ . What we say is that ϑ is *preserved* from Γ to φ . This generates a relation of ϑ -consequence, i.e., a set of pairs $\langle \Gamma, \varphi \rangle$ such that φ is a ϑ -consequence of Γ . We can refer to this relation as \models_{ϑ} .

The next steps we have to take are done for most work in (formal) philosophical logic:

A We must decide what the formulas of the formal language stand for.

B We must characterize the property ϑ in some mathematical idiom.

C We must be able to characterize (to some degree) the \models_{ϑ} relation using rules for manipulating the formulas of the formal language.

So we take the step A to provide an interpretation of the formal language, B to provide a mathematical (formal) semantics for the formal language, and C to provide a proof theory that is at least sound for the formal semantics. There are many other properties that logicians like a proof theory to have, but soundness seems like a minimal one.

An alternative preservationist project is to take an existing consequence relation \models , and a property ϑ , then see if there is a sub- or superrelation of \models that preserves ϑ . Ideally, in this alternative method one is looking for some relation $\models^+ \subseteq \models$ such that $\Gamma \models^+ \varphi$ iff whenever Γ has the property ϑ , then φ^1 has the property ϑ . regardless of the method used the steps A, B and C are undertaken in some manner.

So the natural extension of the preservationist project into a logic of institutions is to look for a property that can be preserved between sets of norms. The property that makes the most sense, because it is manifested in most if not all conceptions of institutions, is that of *inforceness*. We will refer generally to the notion of a norm being part of an institution (legal or otherwise) as that norm being *in force* for that institution. Generally we just say that a norm is in force leaving the relevant institution implicit. We will say more about this notion below, but for the moment we leave it in its abstract form since we are discussing logic.

So the notion of institutional consequence is given by a consequence relation between sets of norms. We then can define a notion of *norm consequence* as: For any set of sentences $\Gamma \cup \{ \varphi \}$ that represent norms, φ is a norm consequence of Γ ($\Gamma \models_N \varphi$) iff whenever all of the norms in

¹Or $\{ \varphi \}$.

Γ are in force, then φ is in force. This provides the general account of norm consequence, i.e., one that is independent of the conception of norms.

To make sense of this definition and formalize it precisely, i.e., characterize \models_N , a formal language and a mathematical characterization of in forceness are needed. But as we will discuss in section 2.2.1, both the conception of in forceness and the conception of the ontology of norms can be very different, and influence the consequence relation for norms. We will argue that the varying conceptions of norms and in forceness exert so much influence that it is pointless to unify the study of norms. It is better to pick a conception of norms and develop a consequence relation for that conception. So we can have a logic of norms; however, we will have different logics for different conceptions of norms.

2.2 Institutions and Consequence

2.2.1 Institutions

Institutions are ways of organizing human behaviour, but what that amounts to isn't exactly obvious. To focus the discussion, we restrict our attention to the law. The law is a social institution par excellence, and we take it as our paradigm example. What Anything we say about the law will generalize to all institutions, as we use the term in this discussion.

Since social institutions are ways of organizing behaviour, they are essentially sets of norms, and that is definitely the case for the law.² Already at this point things get messy since there are different conceptions concerning 1) what a norm is, and 2) whatever legal norms are, what makes them parts of The Law. Combinations of answers to these two questions provide a conception of legal institutions, but we can generalize the situation to arbitrary social institutions. Let's take a look at an example to get a better idea about what norms are.

A simple model of what the law is is given by the Master System interpretation of the law due to Alchourrón (1996). Many legal systems are codified, i.e., written down in legal texts. That is what is called the *master book*. But there are always various ways to interpret a master

²Dworkin (1978, Ch1–2) and his later work takes issue with this view.

book (because of vagueness, etc.), and a particular interpretation of all of the legal texts taken together is called a *master system*.

The distinction between master book and master system is related to the distinction between an indicative sentence and a proposition. A proposition is a semantic, abstract entity: a sentence is a linguistic and physical entity.³ Just as a particular sentence may express a proposition, we will say that the master book (when interpreted) *expresses* a master system. We will refer to the linguistic entities in a master book as *norm formulations*, and the things that are expressed by the norm-formulations in the master book as *norms*. This makes norms a kind of semantic entity. At the moment we haven't said anything about the composition of norms.

But there is one more distinction that we should note before moving on. When a master book is interpreted we get a master system, i.e., a collection of norms. We can say that each norm-formulation in the master book expresses one norm in the interpreted master system, and each norm in the master system is the expression via the interpretation of *one* norm-formulation in the master book. This way we can make sense of the idea of an explicit norm versus an implicit norm. For laws we have explicit laws, those that are interpretations of norm-formulations from the master book, and maybe we have implicit laws, norms that are also laws, but not explicitly on the books. We take the norm consequences of explicit norms to be one kind of implicit norm. Other kinds of implicit norms do not interest us for the moment, so we leave them aside. But the distinction between an explicit norm and an implicit one raises the question: are implicit norms norms at all? Put another way: are implicit norms in force? We will look at some examples below to make better sense of this question.

2.2.2 Norms

Not everyone agrees that legal systems treat norms as semantic entities, ontologically speaking. At base, this disagreement comes from a disagreement about the ontology of norms and

³More precisely, tokens of sentences are physical entities.

the primary distinction is that between the hyletic⁴ and expressive conceptions of norms, a distinction originating in Alchourrón and Bulygin (1981). “For the *hyletic conception* norms are proposition-like entities, i.e. meanings of certain expressions” (ibid. p. 96). Indeed, they are the semantic contents of norm formulations. “But [norm formulations], unlike descriptive sentences, have *prescriptive meaning*: that something ought, ought not, or may be the case (or done)” (ibid.).⁵ The original account of the hyletic conception holds that norms are the result of applying deontic operators, i.e., ‘is obligatory’, ‘is permitted, etc., to propositions, but we suggest allowing this view of the composition of norms to be one interpretation of the hyletic view. What is fundamental is that there is a semantic category distinct from that of propositions for interpreting norm formulations.

For the expressive conception, “norms are the result of the *prescriptive use* of language” (ibid.), and “are essentially *commands*” (ibid. p. 97). This makes norms expressions in a pragmatic mood, not the sense of a sentence. This view is held by many as Alchourrón and Bulygin (1981, p. 98) point out. But now we can look at what it is that makes one norm a norm *of* an institution and another not. We can get a clearer view to the distinction between explicit vs. implicit norms.

For the expressive conception norms are commands. Presumably, for a norm to be classified as a norm of the law means that it is the command of a legal authority. On the other hand, the hyletic view isn’t married to that way of norms becoming legal norms since norms have an independent existence from the commanding of some authority. There may be many ways that a norm can come to be a legal norm on the hyletic view. One such way might be something akin to a convention,⁶ if enough people all follow a norm implicitly then it is a legal norm.⁷ This view assumes that there are norms prior to there being an institutional norm. Or perhaps

⁴The word ‘hyletics’ is used in Ricoeur (1988, endnote 4, p. 281) in reference to uses by Husserl to mean “The study of matter or raw impressions of an intentional act; the abstraction from the form.” Goldenrowley (2009). Presumably Alchourrón and Bulygin (1981) mean to use it to say that norms are an abstraction from the form of the norm-formulation.

⁵A similar view is held in Castañeda (1975).

⁶This view is held by Postema (1982).

⁷See Lewis (1969).

God has handed down laws that we must follow, and it is only those norms that God has sent to us that are true legal norms. A norm can be legally in force, and when it is it will have the property of legal in forceness. The property of in forceness doesn't apply only in the hyletic case, however. For the expressive conception in forceness for a norm is just the act of making that command.

Now consider the example where it is sufficient for a norm to be in force when there is a convention in place. Then that can be a legal norm that is not in the master book, but still a legal norm. That means that it is an implicit norm. So on this account implicit norms exist, and are norms. One view of the expressive conception could be construed as saying that only the explicit commands are norms. On that conception of the expressive conception there are no implicit norms.

We will introduce one final bit of terminology. On either conception, hyletic or expressive, norms have content. The content of a norm is what it says to do, or what will satisfy or violate it. For the expressive conception the content of a norm is the proposition that is commanded to be brought about according to Alchourrón and Bulygin (1981). In the hyletic conception the content may also be the proposition that ought to or may be brought about, but that depends on the version of the hyletic conception.⁸

Now we will reiterate our claim that the conception of norms and in forceness will matter to a logic of norms. In the next section we will defend this claim. First we will argue how it is that in forceness matters and second we will argue that the conception of norms matters as well.

2.3 The Conception Matters

2.3.1 In Forceness Matters

The conception of what makes a norm in force, whatever a norm may be, can affect the logic, sometimes in radical ways. It is thought that logic should be philosophical. This means that a logic should make as few philosophically weighty commitments as possible, e.g., be committed

⁸Cf. Hare (1952), von Wright (1963), Braybrooke et al. (1995), Vranas (2008).

to a particular conception of possible worlds. This allows a philosophical logic, e.g., modal logic, to be widely applicable in philosophical arguments. To be widely applicable means not to validate any inferences/arguments that aren't validated by any conception of the target topic. In the current project, the way for a relation of norm consequence to be philosophical is for it to be compatible with any conception of norms. However, some conceptions of norms and conceptions of in forceness would restrict the norm consequences to just the explicit norms. That would leave no room for consequences distinct from the explicit norms, and so make for an austere and uninteresting logic.

To illustrate this point, we look specifically at conceptions of law. Various conceptions of law differ on what is called 'legal validity', which is what we would call *legal in forceness*. In the philosophy of law there are three leading schools of thought that all concern what it is that makes a law in force: Natural law theory, legal positivism, and legal realism. It is this third school that is of interest here. The legal realist doesn't think that the in forceness of laws is something that one may consider rationally; the law is simply the whim of judges. The legal realist allows us to make our point.

What seems to be important to the realist, however, is that there is no special, metaphysical basis for a law's being in force (such as a god or special source of morality). At best, a non-judge claiming that a norm is legally in force is a prediction about what a judge will use in making a particular ruling on a case, or what will be used as a substantial justification in a ruling.⁹ The legal realist cannot say to the judge 'the ruling that you made was wrong,' since the only thing that can properly be called 'the law' is what the judge rules. So regardless of the conception of norms, if the legal realist is right, then there are no logical connections between norms of law that are in force since it is only what is actually used by the judge that gives a norm its in forceness. Or to put it another way, the conditions for one norm being in force don't connect or fulfill the conditions for other norms to be in force since norm formulations for the other norms are not used by the judge. The best a logician could do is work with a psychologist to predict

⁹This was a view developed by Alf Ross (1958), see Peczenik (2009, p. 214).

the rulings a particular judge might make. This is an extreme interpretation of legal realism, so there is room for legal realist views that could permit a logic of norms that is less austere, but the austere view was held by at least Ross.

So we take it that we have established the two points we wished to make. First, the conception of in forceness can affect what the norm consequence relation is like. How the notion of in forceness affects the consequence relation may be very complex or very simple. But also, to do philosophically interesting work we should look to formalizing a particular conception. We should focus on one conception primarily because if we look to please every conception, the norm consequence relation would be uninteresting. Now we will argue that the conception of norms has an effect on the norm consequence relation as well. Before moving on we would like to make a brief aside about in forceness.

The phrase ‘in forceness’, as we are using it, is a placeholder term. It can be interpreted in many ways, hence why we can talk about different conceptions of in forceness. Later we will take a particular view on in forceness, but for the moment it is variable. It stands in for whatever the norm making property is within a conception of norms. That means it can be anything from a kind of common adherence to a particular regularity in action in certain context to the official decrees of a monarch.

2.3.2 Norms Matter

There is another way that norm consequence could be uninteresting: if it is logic as usual. If the logic of norms is just some standard kind of normal modal logic, then it might be seen as uninteresting. We know a lot about modal logics. But the conception of norms can influence what the logic looks like as well.

Consider a hyletic conception of norms where the notion of in forceness is that a legislator utters sentences like ‘A ought to see to φ ’, where φ is some proposition. Thus norms are special propositions that are made up by putting deontic operators on to propositions. It is a contentious point what the “right” deontic logic is, i.e., the logic of the terms ‘is obligatory’, ‘is permitted’,

and ‘is forbidden’, but deontic logic is often constructed by analogy with alethic modal logic. Recall that in alethic modal logic whenever a proposition ψ implies φ in classical propositional logic (i.e., ψ truth functionally implies φ), then if it is true that ψ is necessary, then it is true that φ is necessary. This is called the rule of inference RN. Thus by analogy, if ψ implies φ , then if it is true that ψ is obligatory, then it is true that φ is obligatory. This would be a case where the logic of norms is logic as usual. But the expressive conception offers a different view.

On the expressive conception norms are imperatives or commands. A command is a sentence, but it is not a sentence that *can* be true or false. So we have to ask, rhetorically, what can we say about a logic of commands/imperatives? This is the famous Jørgensen’s dilemma (Jørgensen, 1937), that either a logic of imperatives isn’t possible because logic only deals with truth, or logic in some way deals with things other than truth. There are many suggestions for a logic of imperatives cf. Jørgensen (1937), Hofstadter and McKinsey (1939), Chellas (1969), Searle and Vanderveken (1985), Hamblin (1987), Vranas (2008), to mention a few. Some of these try to derive a logic for imperatives from classical propositional logic by analogy to alethic modal logic, while others look for different foundations. The analogy with alethic modal logic interprets commands as sentences like: Bring about φ ! An imperative is interpreted by sentence operator applied to a proposition: $!\varphi$. But Ross’s paradox suggests that an analogy with alethic modal logic is problematic because of the inference rule RN. Ross’s paradox goes as follows.

Suppose that A is ordered to mail the letter B. So this has the form $!\varphi$, i.e., bring about that B is mailed. It is true of classical propositional logic that if B is mailed, then B is either mailed or B is burnt. This is just or introduction: φ to $\varphi \vee \psi$. If imperative logic was like alethic modal logic, i.e., RN held in all cases, then A would be ordered to mail B or burn B, i.e., $!\varphi \vee \psi$. But that result is unpalatable. This means that RN shouldn’t hold unrestrictedly for a logic of imperatives; the logic of imperatives isn’t logic as usual. So the conception of what a norm is, in this case a command, may affect the logic as well.¹⁰

The notions of in forceness and the ontology of norms may be very complicated. Suppose

¹⁰A similar problem arises on the hyletic conception when we try to use deontic logic in the case of hyletic norms, i.e., prescriptive propositions like A is obligated to bring about φ .

that the conception of in force-ness being used is that of a norm being a convention, and norms are regularities in action—whatever that is. Conventions work roughly as follows. A convention is a regularity in action, some action type, that gets repeated in similar recurring circumstances. The regularity in action provides a way of solving a problem involving coordinating human action. But for a convention there are equally adequate alternatives for coordinating the action. An example of this is driving on the right side of the road. That legal norm could have been otherwise, but in Canada that is the norm we follow. This is roughly a view held by Postema (1982, 1994).

We ask rhetorically, what would norm consequence be like for these conceptions of in force-ness and norms? That problem is incredibly complex and it seems that abstract discussion will not provide the answer. To give an answer would involve a detailed discussion of the problem. Essentially, to answer the question we would have to try and construct a logic. The point is that prior to investigating the details of the problem we lack a clear idea as to what the result would be. And trying to please every conception is uninteresting. So to do novel work it is best not to be philosophical.

2.4 Historical Precedent

That a logic isn't philosophical is not a totally novel view. Dummett (1991) holds a similar view concerning the difference between verificationist and realist accounts of truth. The realist conception of truth that propositions are true regardless of human knowledge of their truth leads to classical logic and verificationism, i.e., that truth *is* just the verification of a proposition's truth, leads to intuitionistic logic. Closer to home, Sayre-McCord (1986) argues that standard deontic logic, which will be discussed briefly in section 5.1.3, isn't philosophical because it doesn't permit there to be conflicting (moral) obligations. However, there are moral theories, i.e., conceptions of morality that allow there to be conflicting moral obligations.

The idea here isn't quite new. Indeed, Alf Ross (1944), offered the idea that there are really two logics of imperatives. One logic of imperatives deals with imperative satisfaction. To

satisfy an order is to fulfill the order, or do what is ordered. When A mails the letter he has been ordered to mail he satisfies his order. But any situation in which A satisfies his order, i.e., mails the letter, is also a situation where he satisfies the order to mail the letter or burn the letter. That is because ‘either B is mailed or B is burnt’ is true in any circumstance where ‘B is mailed’ is true. That means RN *does* hold when we are talking about satisfaction.

A logic of validity for imperatives is supposed to track what orders are given. So in ordering that φ be done, a logic of validity would provide the other orders that were given in virtue of the order to bring about φ . It is a logic of validity that Ross’s paradox applies to. When there is a valid order to mail the letter B, intuitively, there isn’t a valid order to mail or burn B.

We have generalized Ross’s idea of a logic of validity for imperatives to a logic of in force-ness for institutional norms. This way we can give a general schema to be filled in by various conceptions of in forceness and norms. Since now we have established that a logic of institutions can be affected by both conceptions of norms and conceptions of in forceness, we think it best to develop logics based on particular conceptions of norms and in forceness. We will be as general as possible where we can, and reflect on the extent of generality in the end. But we will not kid ourselves or our readers by passing this logic of institutions off as philosophical.

2.5 Our Choice

We have looked at what matters to a logic for institutions and found that there are two important factors. First is the conception of norms, and second is the conception of in forceness. We have laid out a principled methodology for our notion of institutional consequence, i.e., the preservation of in forceness, in the form of preservationism. We have also provided some of the context of discovery to help place this project within an intellectual heritage.

The primary point to take away from this discussion is that conception of institution matters to a logic of institutions. The conception of institutions determines the nature of the norms involved, and the nature of in forceness. But how the nature of those two components influences the resulting logic isn’t something that can be dealt with in broad strokes. We must get our hands

dirty. To that end in the next chapter we take up a conception of institutions and explain the philosophical background. The idea is to use Searle's account of social institutions to underpin a logic of institutions. We will now take the fork in the road.

Chapter 3

Searle on Social Institutions

Unlike shirts and shoes, institutions do not wear out with continued usage.

John R. Searle, (2010, p. 104)

As we said in the last chapter, we will choose a particular conception of institutions and develop a logic based on that conception. The account of the logic of institutions presented in this essay is founded on Searle's account of social reality. Briefly, Searle's account makes all social institutions systems of *Status Function Declarations*. Searle's account of institutional norms takes those norms to be institutional facts. And institutional facts have special properties and are brought into existence in a special manner.

Roughly, Searle's account makes use of three things: The speech act of declaration, collective recognition, and status functions. In what follows we will survey Searle's account of social reality by first looking at his account of speech acts, then at his account of norms, and finally at his account of social ontology. This will allow us to explain the notions of *declaration*, *collective recognition*, and *status function* used in Searle's account of institutions. In explaining Searle's account of institutions we will highlight his conception of norms and his account of inforceness. In chapter 4 we will discuss, in an informal manner, the logic of norms that issues from Searle's account.

3.1 Speech Acts

Following Searle, institutional powers are brought about by status function declarations. But norms in general, i.e., assignments of function in general, are created by declarations. In the following exposition we will follow Vanderveken (1990). We do this because Vanderveken is working from Searle's theory, but he presents a formal theory of speech acts and catalogues many types of declarations/performative types of speech acts that are relevant for our project.

3.1.1 Speech Act Theory

A speech act is any attempt to *do* something with words.¹ Speech acts must be evaluated relative to an interpretation (of the words involved) and a context of utterance. In fact, the interpretation of some of the words involved in any utterance will be fixed by the context of utterance, e.g., proper names. Each speech act so interpreted will then have three parts: the locutionary act, the illocutionary act, and the perlocutionary act. The locutionary act is the utterance itself, take for example the utterance of a sentence ‘bring me my coat’. The locutionary act is the saying of that sentence. The perlocutionary act, according to Austin, is what one does *by* saying something (Austin, 1962, p. 109). The perlocution is an effect of what the speaker is attempting to perform, or have recognized, or understood by the hearer(s) of an utterance; we will call all of these things *effects* of an utterance. But it is the intended effect of an utterance that could be attained through non-verbal means, which is the perlocution. So if someone yells ‘Duck!’, intending to cause the hearer to duck, the hearer may misinterpret the speech act and look for a duck. Nonetheless, the perlocutionary act in that case was to get the hearer to duck, although that goal was frustrated by the hearer’s “fowl interests”. But that brings us to our focus in this section, the illocutionary act. It is the illocutionary act that is the primary unit of meaning in the use of natural language. This is one of the foundational theses of Vanderveken (1990). The illocutionary act is what speakers intend to do with their utterances using language to communicate. An illocutionary act is made up of an illocutionary force *F* of the speech act, and a proposition *p* which is referred to as the content of the speech act. The type of illocutionary act can then be symbolized by $F(p)$.

Illocutionary forces come in five general flavours: Declaratives, Directives, Commissives, Performatives, and Expressives. The central examples of the five types are assertions, orders, promises, declarations, and emotive exclamations, respectively. The theory of Searle and Vanderveken (1985), claims that all other speech acts such as warning, conjecturing, demanding, begging, promulgating, christening, etc., can be represented by these five basic categories by varying the various components within the relevant category of illocutionary force.

¹The seminal theory of speech acts is found in Austin (1962).

The reason that there are only five general types is that there are only four general “directions of fit” between the world and our words. This sounds bizarre, but we will explain how this happens. According to Searle, language allows its users to represent their intentional states, such as belief and desire. A direction of fit describes the way that our intentional states relate to the world. In making an *assertion*, a speaker is trying to represent the world by their words. So the speaker wants their words to fit the world: word-to-world direction of fit. On the other hand, *promises* and *orders* have the opposite direction of fit. In making a promise or giving an order the speaker is trying to make the world come to fit their words, so the illocutionary forces have world-to-word direction of fit. Expressives have an empty direction of fit since they do not represent, they are simply used to express intentional states. But performatives, e.g., declarations, both represent the world, and at the same time make the world the way that the content of the speech act represents the world as being. Recall that the content of a speech act is always a proposition—at least speech acts with non-empty directions of fit. So performatives are said to have a *double* direction of fit. This provides four directions of fit: word-to-world, world-to-word, both, and none. To get the five basic categories we just have to note that there are two ways for the speaker to try and get the world to conform to its words,

1 promising (the speaker commits *him/her self* to making the world a certain way)

2 ordering (the speaker commits *others* to making the world a certain way)

So the extra world-to-word directions of fit provides the fifth category of speech act. We will focus on explaining the general account mostly in terms of the illocutionary force relevant to our project, viz. performatives.

Any illocutionary force, on the Searle-Vanderveken theory, is (or can be represented by) a function of six arguments: illocutionary point, mode of achievement, propositional content conditions, preparatory conditions, sincerity conditions and degree of strength of those sincerity conditions. It is important to note that the arguments are not always independent. For example, some sincerity conditions can determine preparatory conditions. However, these relationships

are not involved in the goal of our project so we won't discuss them. The illocutionary point is what act-type the speech act is, i.e., what the speaker is attempting to do with the words. The illocutionary point of 'I will return your book' is to promise to return the book. In the current case, i.e., for declarations, the speaker is attempting to bring "into existence a state of affairs by representing oneself as performing that action" (Vanderveken, 1990, p. 105). To put this in a better perspective, consider a sentence that represents a performative speech act like: (1) "I baptize you X". The speaker who utters (1) represents themselves as performing the act of baptizing. As Searle would have it, the speaker is representing that state of affairs as existing, namely the state of affairs in which that entity being baptized becomes a member of the Christian Church.² As mentioned above illocutionary acts have directions of fit, and that is what the illocutionary point represents: the direction of fit between the world and the words of the speaker.

The mode of achievement relates to the point of the illocutionary force. The mode of achievement of the illocutionary point restricts how that point is to be achieved. If A begs someone to do something, it is a directive point since the speaker, A, is attempting to get someone to do something; but that directive has to be accomplished from a humble, polite or even desperate position. It isn't quite how we would usually think of a directive, but it does have a world-to-word direction of fit in which the *speaker* tries to get the hearer to do what is being begged for. We will explore this more in an example below.

The act a (Catholic) priest performs by making an utterance such as (1) isn't something that stands on its own, however. There are many things that go into making such an utterance really an act of christening. The last three types of conditions concern those background requirements of illocutionary acts. Next we have the propositional content conditions. Propositional content conditions are those requirements on propositions that are the contents of certain illocutionary acts. In directives, e.g., commands, the propositional content must be a future proposition: no one can command someone to do something yesterday—although someone might say that to

²According to the Encyclopedia Britanica Online, christening is an admission ceremony of various sects of Christianity (Encyclopdia Britannica Online, Encyclopdia Britannica Online).

express the urgency of the desired result. The same is true about commissives.

For declarations the propositional content conditions have to do with the peculiar direction of fit. In Vanderveken's theory, declarations cannot have necessary propositions as their propositional content. One cannot declare a necessary proposition true, or a necessarily false proposition false, for that matter. The propositional contents of declarations must always be contingent propositions. There is another aspect of declarations that Vanderveken leaves out: the proposition that grass is green is contingent, but could someone *declare* such a proposition true? Presumably not since the proposition is independent of language on this view. This means that excluding the necessary propositions from the content of declarations isn't the whole story. The propositions that are contents of declarations must be amenable to being declared. As hard as A might try to declare that grass is blue, A will not succeed, the proposition expressed by 'grass is blue' just isn't amenable to declaration qua performative act. We will discuss this more in the sequel.

Preparatory conditions, as the name suggests, are the conditions that must be met to be in a position to succeed in performing a speech act. Promising, the commissive act par excellence, really only needs there to be a linguistic community that includes locutionary acts that count as commissives in certain contexts. But the existence of such a linguistic community isn't a small matter. The non-trivial cases of preparatory conditions, i.e., those beyond the existence of a linguistic community, are *very* important to institutional reality. Consider the priest case above. In order to properly baptize a child, the speaker must be a priest, which requires the existence of the church and all of its requirements—a highly non-trivial set of preparatory conditions.

Sincerity conditions are where the intentionality of the speaker is very crucial. They refer to the mental states that a speaker represents themselves as having in performing an illocutionary act. If someone says, 'I think there is life on Mars', they represent themselves as believing the content of the putative assertion, i.e., that there is life on Mars. Similarly with promises; if someone says 'I shall return your book' but doesn't have an intention to return the book, then this person has performed a locutionary act, and a commissive illocutionary act, but there is some-

thing wrong with that act. That speaker isn't sincere, i.e., doesn't have the mental/intentional states that they represent themselves as having.

Finally we come to the degree of strength of the sincerity conditions. The degree of strength applies to the mental/intentional states required by the sincerity conditions. Sometimes the mental states represented in the performance of one illocutionary act must be much stronger than those represented in another act. For instance, if someone conjectures that there is life on Mars then they haven't asserted something, but they have performed a speech act in the vicinity of asserting. Conjecturing requires representing less strength in the mental state than a bona fide assertion.

When a locutionary act is performed, and the putative illocutionary act is of the form $F(p)$, something might go wrong with some of the conditions previously outlined for the illocutionary force F . This relates to the notion of a speech act being *successful*. To quote Vanderveken at length, when:

(1) the speaker *achieves the illocutionary point* of F on the proposition $[p]$ ³ *with the mode of achievement* of F , and $[p]$ satisfies the *propositional content conditions* of F in that context;

(2) the speaker moreover *presupposes* the propositions... determined by the *preparatory conditions*... of F ; and

(3) the speaker also *expresses* with the *degree of strength* of F the mental states [necessary] with the psychological modes... determined by the *sincerity conditions* of F (1990, p. 129)

the speaker successfully performs an illocutionary act of the form $F(p)$. Vanderveken holds that one can successfully perform an illocutionary act, but the act may be defective in some way, e.g., he may presuppose a proposition needed by the preparatory conditions that is false, or may express that he has mental states that he doesn't. Suppose someone says 'I promise to return

³Here Vanderveken uses a ' P '. The '...'s remove notation that Vanderveken uses that we haven't introduced.

your book', but has no intention of returning the book. In such a case the same illocutionary act of promising to return said book is successful, but the necessary mental states are missing so the sincerity conditions are not met. So Vanderveken would say, we believe, that the illocutionary act was successful, but it was defective. On the other hand, suppose that the same locutionary act is performed, and the sincerity conditions are in place, but the book has been burnt up in a fire, and the speaker is unaware of that. That means there is a preparatory condition that fails to obtain for the act of promising to be non-defective. One might meet the sincerity conditions in the latter case, but there is no book to return. For Vanderveken a successful illocutionary act is one where the speaker says things properly, but he might not meet all of the preparatory or sincerity conditions, i.e., the world nor the speaker might not meet the preparatory conditions for the speech act to happen in a non-defective manner.

Vanderveken suggests that Austin's felicity conditions don't distinguish between the possibility of performing an illocutionary act successfully, but in a defective way.⁴ So what matters to successful performance is that the speaker represents itself and the world as meeting all of the relevant conditions. Nonetheless, the act can fail, in a certain sense, because the world hasn't agreed with the speaker's presuppositions or the speaker misrepresents him/herself. When a speaker meets the right conditions and the world agrees with the speaker's presuppositions, then that is what Vanderveken calls an illocutionary act being 'non-defectively performed' (Ibid. p. 130). In the performance of an illocutionary act a speaker represents the world and the speaker as satisfying all of the conditions necessary for non-defective performance.

The final topic of interest in relation to illocutionary acts is the notion of a speech act being *satisfied*. Satisfaction has to do with the direction of fit of the illocutionary point. When someone asserts *p* the direction of fit is from words to the world, i.e., words-to-world. When an assertion is satisfied the proposition asserted is *true*; thus, the satisfaction conditions for assertion reduce to the truth conditions for the proposition asserted. The case is different for illocutionary acts with a world-to-words direction of fit. If someone promises to bring about *p*,

⁴Vanderveken isn't correct on this point since Austin would say that a successful but defective illocutionary act was unhappy in some way. Cf. Austin (1962, pp. 12–38).

then the promise is satisfied when it is kept. Whether a promise is kept isn't simply spelled out by the truth conditions of propositional content *p* of the illocutionary act. Of course the truth of the relevant *p* plays a role in the satisfaction conditions of any speech act, but it may not be the whole story. Presumably, to fulfill a promise the promiser must *make* the proposition true, somehow. The satisfaction conditions for declarations are particularly interesting since they have both directions of fit. Again we quote Vanderveken:

[A] declaration is *satisfied* in a context if and only if the speaker performs the action represented by its propositional content by way of representing himself as performing that action in his utterance. On this account, a declaration could not be satisfied if it was unsuccessful [and conversely]. (Ibid. p. 133)

Thus with declarations, succeeding is coextensive with satisfaction, i.e., occurs in all of the same circumstances. This gives declarations a bootstrapping effect since they can make something out of almost nothing, and that effect is what is needed for creating institutions on Searle's theory. That concludes our brief tour through the components of illocutionary acts. Now we will look at some specific performative acts to bring out some details about declarations in relation to institutions.

3.1.2 The Specific Speech Acts

In the current project we are following Searle's theory about the construction of institutional reality, and it is through those special acts known as declarations that we construct that reality, as we will discuss below. Declarations are a specific form of performative speech act. In performing a declaration the speaker brings about the truth of the propositional content of the speech act.

But declarations aren't as simple as the Catholic priest example above might indicate, if anyone thought that was simple. The preparatory conditions for many kinds of declarations are extremely complex. But once there is the underlying social infrastructure, fewer preparatory conditions need to be added. Once there is a legal institution, judges have abilities to sentence

and provide rulings. Any speech act is looked at in relation to a context of utterance, and it is the context that handles many of the various conditions for successful speech acts, particularly the preparatory conditions.

In this section we discuss some acts which highlight important considerations in the use of illocutionary acts to make institutional facts. We will get a better understanding of their preparatory conditions, propositional content conditions, and sincerity conditions. In this we focus on the law, and only discuss the fragment consisting of one illocutionary act: promulgation. Following Vanderveken, to promulgate is

to declare publicly (mode of achievement) an enactment of some legal status (propositional content condition). (1990, p. 208)

We want to look at actions that change or create the composition of the institution as opposed to actions that are within the institution, like baptism. Of course speech acts that change an institution also have to be from within the institution itself in some sense, but we will leave that aside for now.

As we discussed in the previous section, there are six arguments in any illocutionary force: illocutionary point, mode of achievement of that point, propositional content conditions, preparatory conditions, sincerity conditions, and their degrees of strength. We will explain these arguments in what follows. The illocutionary point of a promulgation is to declare, i.e., it is a performative type illocutionary act. However, this type of declaration requires a specific mode of achievement involved, viz. *public* declaration. Thus to succeed in promulgating a new piece of law, the public must be able to become aware of the new law.

Concerning the propositional content conditions, Vanderveken claims that the content of the promulgation must have legal status and that status manifests as a restriction on the content of the promulgation. But there is another legal feature about promulgations: part of the legal-making feature of a promulgation is that it is promulgated *by institutional authorities* in their capacity as legislators. Thus the legal status of a promulgation is found in three places. The legal status is partly in the preparatory conditions, because such an authority must exist. It

is also partly in the mode of achievement since to perform the promulgation that authority is invoked.

The mode of achievement for promulgations is in the invoking the authority to make laws of the speaker to make the declarations, to create the laws, and in a public manner. Simple declarations don't have these additional requirement for their modes of achievement since someone can say 'I am asking if it has rained today' and so declare that one is asking a question: they represent themselves as asking a question.⁵ In making that declaration the speaker wouldn't need to invoke any special authority to make the declaration. However, in a promulgation the promulgator must indeed make use of the special authority in order to succeed in performing a promulgation at all. It is like when a priest says 'By the power invested in me...', the priest uses a special mode of achievement in pronouncing people married. A priest invokes the power given to them by a church to create something like a marriage; without that power a priest cannot create a marriage. The mode of achievement for promulgations is wrapped up with the special preparatory conditions since the promulgator must have such authority in order to invoke it: without that authority there can be no success.

The weird part of promulgations, and declarations generally, are the sincerity conditions. Mental states don't seem to play into the success of these declarations in general. One could imagine a dictator of a small country who forms no prior intention to enforce a particular law that they have promulgated. Then we might say that the sincerity condition of this isn't met since it is the dictator who is the chief administrator, and the ultimate enforcer of the laws in that country. But in democracies such as Canada it is not clear that its legislators need to have particular mental states concerning the laws they promulgate. Indeed, in democracies it seems that promulgation is a collective speech act. There is no one individual that performs the promulgation, it is a collective effort. We will discuss this more when it becomes relevant in section 4.2, but ultimately we want to say that there are no sincerity conditions for promulgations.

⁵This may seem odd since by someone saying "I am asking..." seems to be simply asking a question. But saying "I am asking..." asks a question *in addition* to making the declaration. Whereas saying "Did it rain today?" simply asks a question.

When we use ‘speech acts’ to refer to promulgations by governing authorities we use ‘speech’ loosely: it could refer to any number of symbolic acts to the same effect. In Canada, for instance, it is the signing of a bill that makes it law. But that doesn’t really affect the kind of action qua promulgation. We should note that promulgations may not be necessary to make laws. It is easy to imagine a ruthless ruler who makes new secret laws to entrap their public. Such laws may be genuine laws depending on which conception of law is right, but they would fail to be promulgated since they are not public.

So far we have dealt with the ‘declaration’ part of ‘status function declarations’. Now we will look at Searle’s conception of institutions and explain what a status function is, as well as the role of collective recognition in his conception of institutions.

3.2 The Components of Searle’s Conception of Institutions

Early in Searle (2010) he says

The claim that I will be expounding and defending in this book is that all of human institutional reality is created and maintained in existence by (representations that have the same logical form as) [Status Function] Declarations, including the cases that are not speech acts in the explicit form of Declarations. (Searle, 2010, p. 13)

We have seen what a declaration is, but to understand the notion of a status function we have to understand *collective recognition*.

Collective recognition is a form of collective intentionality, akin to collective belief and collective desire. Collective beliefs and desires are the kinds of beliefs and desires expressed in utterances like ‘we want to win the soccer tournament’ and ‘NASA believes the rocket launch will proceed on schedule’. Some collective intentions are reducible to each individual having the same intention, e.g., we believe that John is alive, i.e., each of us believes John is alive. These are called distributive collective intentions Meijers (2007). But some intentions don’t reduce in this way. Collective actions are sometimes like this, the playing of a piece of music

by an orchestra is done by each member playing their part. The collective intention is to play the piece of music, but each individual doesn't have an intention to play the same thing, since the role each instrument plays, i.e., the score each plays, is different. This non-distributive sense of collective intentions can be thought of as explicit cooperation.

The sense of collective recognition that Searle requires for the existence and maintenance of institutions is the distributive kind of collective intention: it doesn't require cooperation in general. As Searle says

[I]n an actual transaction when I buy something from somebody and put money in their hands, which they accept, we have full-blown cooperation. But in addition to this intentionality, we have prior to the transaction and continuing after the transaction an attitude towards the pieces of paper of the type I am placing in the hands of the seller, that we both recognize or accept the pieces of paper as money, and indeed, we accept the general institution of money as well as the institution of commerce. (Searle, 2010, pp. 56–7)

Further, for there to be cooperation within an institution there first has to be this a-cooperative collective recognition of the institution. The collective recognition of a status function, then, is the distributed recognition of the various status functions which make up the institution. Note that 'recognition' doesn't imply endorsement for Searle; it can be grudging acquiescence. But what is a status function?

Searle defines a *status function* as

a function that is performed by an object(s), person(s), or other sort of entity(ies) and which can only be performed in virtue of the fact that the community in which the function is performed assigns a certain status to the object, person, or entity in question, and the function is performed in virtue of the collective acceptance or recognition of the object, person, or entity as having that status. (Searle, 2010, p. 94)

It is essential to status functions that, as their name indicates, they are functions which depend on their status within a community. A status function could not serve the function it does without some sort of collective recognition. As one of Searle's key examples indicates, if some tribe of yesteryear has built a wall around their collection of huts, that wall serves as a boundary and its function as a way of keeping individuals out is achieved because of its physical properties. However, many years later, after the wall has crumbled, the outline of the wall may serve as a boundary, and people will not cross it unless authorized to do so. But that function of keeping people out isn't achieved because of the outline's physical structure. In that case, the function of the wall is achieved by a collectively recognized, symbolic status that the outline has. The outline has power because people recognize the power, and it wouldn't have it otherwise.

The symbolic power that the outline of the previous example has is part of what Searle calls a *deontology*. A deontology is an assignment of powers, rights, prohibitions, and obligations to various entities, especially people. In the boundary example above, the outline of the wall is assigned a special power because it imposes prohibitions on people: People are prohibited to cross the outline unless they have authorization. But deontologies function if and only if they are recognized to be binding via the community. As Searle puts it "*a deontology can exist only if it is represented as existing*" (Ibid. p. 95).

The importance of collective recognition for the existence of institutions is because without it an institution will not "lock into human rationality and will not provide reasons for action" (ibid., p. 102). Searle makes it an important part of institutions that an institution's subjects must see the deontology, e.g., the existence of an obligation, as providing a reason for action. This reason for action doesn't have to be a desire. Indeed, Searle goes against the Humean idea that all action must be underpinned by some desire. A full discussion of Searle's theory of action is unnecessary here, suffice it to say that Searle explains recognition in terms of recognizing deontological status as providing reasons for actions. We will come back to this point in the next section; for now we want to discuss the components of institutions.

Another type of status function that is particularly important is that of a *constitutive rule*. We will use ‘norms’ instead of ‘rules’ to be consistent with prior use in this essay, but Searle uses ‘rules’. Searle holds that there are two kinds of norms, regulative and constitutive. For the difference between regulative and constitutive norms we can follow Searle’s explanation:

As a start, we might say that regulative [norms] regulate antecedently or independently existing forms of behavior [...]. But constitutive [norms] do not merely regulate, they create or define new forms of behavior (Searle, 1969, p. 33).

We might say, and Searle does (ibid. p. 53), that the action or thing to which a regulative norm applies might have exactly the same description had the norm not existed. But that is not so in the case of constitutive norms. There are plenty of examples: speeding versus driving at over 80 kilometers an hour. For someone to speed there must be a constitutive norm that allows the classification of driving over 80 km/h as speeding. Playing chess is another action that would not be possible but for the norms that constitute the game. But a regulative norm prohibiting walking on someone’s lawn, promulgated by a sign posted that says ‘Don’t walk on the grass’, is something that regulates a preexisting action, i.e., walking on the grass. Walking on the grass is possible to do without there being any norms: walking and the grass exist prior to the norm. That the walking is classified as an offence of some kind occurs because there is the rule in force.⁶

These constitutive norms help define the basic components of institutions. If we go back to the general gloss of institutions, i.e., ways of organizing human behaviour, constitutive norms *are* a way of organizing human behaviour by classifying it according to the constitutive norms. For Searle, the general logical form of a constitutive norm is ‘*X counts as Y in C*’. The ‘count as’ formula in general is: action, object, or state of affairs *X*, counts as *Y*—again an action, object, or state of affairs—in context *C*. Constitutive norms allow people to connect physical reality to institutional reality. Institutional reality (*Y*s) are defined into existence out of physical or brute reality (*X*s) via constitutive norms.

⁶Actually, it takes a whole institution of private property in most cases to make such a rule in force.

We must pause a moment on the notion of context. In the count as formula the context functions in a few ways. Searle often phrases count-as statements in terms like: “doing X counts as a base hit in the context of a baseball game”. This use of context in the count as formula is a way of specifying the conditions of application for when X count-as Y . But we can change the logical form of this formulation to: under conditions C , X counts as Y . This makes the role of the context a little less mysterious. However, in a case where the context is a baseball game, there is an ultimate context that is needed: the existence of the institution of baseball. But that context is what is being defined by the collection of status functions, of which this one count-as formula is a part. Thus the use of context in this manner is redundant. We will treat count-as formulas as conditionals stating under what special circumstances X s count-as Y s. Now we return to our discussion of constitutive norms.

Grossi (2007) offers a formal system in which regulative norms are reduced to constitutive norms. We will develop this reduction in detail in section 9.1.3. Grossi’s idea, which is a formalization of the account of institutions in Searle (1995), makes for a simple formalism. According to Grossi, a regulative norm can be analysed as a constitutive norm that says certain actions or states of affairs are to be counted as violations. For instance, to give one of Grossi’s examples, operating vehicles in a public park counts-as a violation of the law. An important goal for a logical system is to show how to represent all of the required notions for providing the deontologies of status functions, i.e., rights, duties, powers, et cetera. We will do that in section 6.4.

In the next section we will see how all of these components fit together.

3.3 Putting it Together

Searle’s account of institutions sees institutions as collections of status functions. But these status functions are imposed on reality in a particular manner for Searle. “All institutional facts are created by the same logical operation: the creation of a reality by representing it as existing” Searle (2010, p. 93). Searle’s schema for this is:

We (or I) make it the case by Declaration that the Y status function exists. (Ibid.)

So the way that institutions come about is through, as the quote at the beginning of the previous section says, the declarations of status functions. Before moving on to mention how a logic of institutions would come out of this we have to take a look at Searle's account of the maintenance and creation of institutions.

In Searle (1995), he thought all institutional facts could be handled by constitutive rules. However, there was a counterexample to that hypothesis: limited liability corporations. Not all status functions are assigned to some preexisting thing. Searle's example of this is a limited liability corporation. Such corporations are just brought into existence through a status function declaration all the same, but although

[t]he corporation has to have a mailing address and a list of officers and stock holders and so on,...it does not have to be a physical object. This is a case where following the appropriate procedures counts as the creation of a corporation and where the corporation, once created, continues to exist, but there is no person or physical object which becomes the corporation. (Searle, 2005, p. 16)

So in this case there is no *X* that is counted as *Y*, hence the general formulation of institutional facts above. However, all institutional facts must connect in some way to brute reality. The connection is achieved by assigning the various roles in something like a corporation to people which gives them deontic powers, i.e., assigns a deontology to these people. Searle gives a more transparent formulation of institutional facts as follows:

We (or I) make it the case by Declaration that a Y status function exists in C and in so doing we (or I) create a relation R between Y and a certain person or persons, S, such that in virtue of SRY, S has the power to perform acts (of type) A. (Searle, 2010, pp. 101–2)

The relation R will often be different for different status functions. In the case of money S is the *possessor* of money, in the case of private property S is the *owner* of the property—to use

Searle's examples. In the case of a corporation, Bill will be a *share hold* in the the corporation Y, and so because of being a share holder, be permitted to vote to elect the president of the company. As a note on something that we will come back to much later (section 6.6), this formula indicates that in developing a formal system in order to interpret an institution we must have some way to assign agents roles in that institution.

Whereas theories of law like that of the expressive conception of norms, e.g., Alchourrón and Bulygin (1981), see legal norms as standing *commands*, Searle's theory sees institutional facts as like standing *declarations*. Institutions are systems of standing declarations because in order to persist, they must continue to be used again and again. This is achieved through collective recognition. As Searle says, "the whole apparatus [of an institution]—creation maintenance, and resulting power—works only because of collective acceptance or recognition" (ibid. p. 103). So the maintenance of an institution depends on facts which have the form:

We collectively recognize or accept (There exists Y in C, and because SRY (S has power (S does A))). (ibid.)

But Searle's theory faces a problem: Consider for the moment all the laws in force in Canada right now, or even just the bylaws of The City of Calgary. It seems to be an uncontroversial point that not every member or subject of these institutions recognizes all of the status functions imposed within those institutions. That is, not everyone is aware of all of the powers of all of those institutional agents in the institution. But Searle uses the distributive kind of collective recognition. So it seems that, on Searle's theory, everyone has to recognize all of the status functions for the institution to persist, and real institutions function without that kind of widespread recognition.

But Searle says that,

you do not need a *separate* attitude of recognition or acceptance for institutional facts within a preexisting institutional structure. If, for example, you accept the institution of baseball, then a given home run or base hit requires no separate acceptance. You are already committed to that acceptance by your acceptance of the

institution... The system, once accepted by participants, commits them to the acceptance of facts within the system because the system consists of sets of standing Declarations (ibid. pp. 102–3).

So recognition of an institution is at a very high level, and that general recognition of the institution is taken as a commitment to the whole collection of standing declarations which constitute the institution. But something has to impose those status functions initially. In the sequel we will refer to whatever imposes the status functions as an *institutional authority*.

Our use of ‘authority’ in this case is rather broad since it may be a whole community, or it may be a particular individual, e.g., a monarch. In the case where the authority is a monarch, what is recognized is a kind of second order status function that gives power to impose status functions on the behalf of the collective, e.g., legislators. But as we have said recognition of that power of the monarch constitutes the recognition of the status functions that monarch creates.

Searle’s conception of norms is that they can all be expressed as status functions, and with Grossi’s insight we can make regulative norms constitutive norms which are a kind of status function. But it makes the norms *propositions*, so these norms can be true and false. That means standard classical logic can apply to them. However, according to Searle, a norm is in force only when it has been made true by declaration. So although the norms themselves are simply propositions and so true or false, whether classical consequences of the norms are also in force depends on how the speech act of declaration (promulgation) interacts with classical consequence. We investigate that in the next chapter.

A final note. There are various theorists that suggest that institutional norms can come into force simply by a process like collective recognition, cf. Lewis (1969), Bicchieri (1997), Binmore (2010). We don’t want to deny the plausibility of those other views, but they are not Searle’s view. On Searle’s view collective recognition has the form of a declaration since what is important to institutions is that they are the result of representation, and representation is imposed by something like declarations. We are not concerned with these other views since they presume a different account of in forceness than Searle; therefore, we set it aside.

Chapter 4

The Justification of Normative Consequence

If you wish to make an apple pie from scratch, you must first invent the universe.

Carl Sagan, COSMOS(1980)

The first part of this essay deals with the philosophical underpinning for a logic of institutional norms. In chapter 2 we argued that a logic of institutional norms requires an account of what norms are and one of in forceness. Only with those two elements in place can anyone properly evaluate whether there is a logic of norms on that conception. Using Searle's interpretation of norms, we assume that all norms are constitutive norms, so norms are classifications used in the specification of status functions. These status functions are imposed on the world through speech acts, particularly status function declarations. This latter idea of a declaration of a status function provides a conception of in forceness. So a logic of norms in this case can be reduced to the logic of the speech act of declaration. This logic of speech acts is known as a logic of illocutionary acts, and even more specifically a logic for the performative acts of abrogation and promulgation.

In this chapter we are going to show that the consequence relation for norms isn't trivial. We will do that by first explaining Vanderveken and Searle's notion of illocutionary entailment, relying mainly on Vanderveken (1990, 1991). This notion of entailment is a general entailment relation between illocutionary acts. The deep question that we have to answer is: what is the relationship between classical consequence as a relation between propositions and illocutionary entailment? Specifically, for what set of pairs $\langle \Gamma, \varphi \rangle$ is it the case that when each proposition expressed by a sentence in Γ is declared/promulgated, is it also the case that the proposition expressed by φ is declared/promulgated? We want to find a subrelation of classical consequence that preserves declarations, i.e., preserves in forceness. In doing this we will justify a conception of normative consequence.

4.1 Speech Act Entailment

Let us represent contexts of utterance by i s and sentences representing illocutionary acts by A and B for the moment. A semantic interpretation \mathcal{I} assigns extensions and intensions to all of the terms involved in the A s and B s. Here we haven't given a full rendering of a formal language, but we use the notation to simplify the discussion. The notation $\llbracket \cdot \rrbracket$ then represents the semantic value of whatever is inside the brackets relative to the interpretation \mathcal{I} . Also $\llbracket A \rrbracket_i$ refers to the illocutionary act named by A in the interpretation \mathcal{I} in the context of utterance i . Since each illocutionary act has the form $F(p)$ consisting of the illocutionary force F and the propositional content p , we say that each sentence representing an illocutionary act has the logical form $\mathbf{F}(\varphi)$ where ' \mathbf{F} ' refers to an illocutionary force, and $\llbracket \varphi \rrbracket$ is a proposition.

The notion of strong illocutionary entailment (or commitment) from Vanderveken and Searle (1985) is put informally as, if A has performed the speech acts $\llbracket \mathbf{F}_1(\varphi_1) \rrbracket, \dots, \llbracket \mathbf{F}_n(\varphi_n) \rrbracket$ successfully in the context of utterance i , then A has also successfully performed the speech act $\llbracket \mathbf{F}'(\varphi') \rrbracket$ successfully in the context of utterance. Recall that a speaker successfully performs an illocutionary act when, in general, they achieve the illocutionary point of an illocutionary act while representing themselves as correctly presupposing the preparatory conditions and expressing that they meet the sincerity conditions to the correct degree of strength. So illocutionary entailment relates the success conditions of speech acts. Also notice that there is no restriction on what speech acts are represented. All of the speech acts could be from different categories.

Vanderveken (1991) recognizes many different notions of entailment between speech acts since success conditions can be related to satisfaction conditions, and vice versa. So there will be a kind of entailment relating satisfaction and success in both directions. Since for declarations, the success and satisfaction conditions are coextensive so we do not have to consider any other kind of entailment, illocutionary entailment will suffice.

We are interested primarily in Vanderveken's definition:

Definition 4.1.1. $\mathbf{F}_1(\varphi_1)$ illocutionarily entails $\mathbf{F}_2(\varphi_2)$ ($\mathbf{F}_1(\varphi_1) \models_I \mathbf{F}_2(\varphi_2)$) iff for any inter-

pretation \mathcal{I} , and any context of utterance i , if $\llbracket \mathbf{F}_1(\varphi_1) \rrbracket_i$ is *successfully performed* in i , then $\llbracket \mathbf{F}_2(\varphi_2) \rrbracket_i$ is successfully performed in i .

So illocutionary entailment holds *between sentences that express speech acts* when the success conditions of the illocutionary acts are related. Illocutionary entailment has to do with what it takes to successfully perform the various acts rather than merely the propositional contents of those acts. We want to know whether a logical entailment from a proposition $\llbracket \varphi \rrbracket$ to a proposition $\llbracket \psi \rrbracket$ guarantees an illocutionary entailment from $\mathbf{F}(\varphi)$ to $\mathbf{F}'(\psi)$.

Since we are focusing on declarations, we only need to focus on one illocutionary force which we will denote by $\llbracket \mathbf{D} \rrbracket$. That means definition 4.1.1 will suffice for our purposes. Now it may seem odd that the success and satisfaction conditions coincide for declarations: one cannot successfully declare something without satisfying that declaration as well. An example will make that fact a bit clearer. Suppose some authority says “I command you to do α ”. That is a declaration, and the content of that declaration is that the authority commands the subject(s) to do α .¹

For the declaration to be successful the authority must meet all of the preparatory conditions, etc., but if the authority does, then it has indeed commanded the subject(s) to do α . The success of the *declaration*, however, doesn’t guarantee that the *command* is satisfied; the subjects of the authority may not do what is commanded. But declarations bring about certain states of affairs. So when declarations are used to create status functions, the declarations are always satisfied when successful since they create institutional reality. Further, Vanderveken says, “*all successful declarations are eo ipso true, satisfied, sincere, and non-defective*” (1991, p. 73). The kind of declaration used in Searle’s conception of institutions acts exactly like this. There can be no defective but successful declarations, unlike insincere promises. The authority represents itself as creating institutional reality, i.e., it brings into existence a particular status function. How we and Searle analyse the bringing into existence of a status function is as the institutional authority making certain propositions true. By so representing itself as making a

¹The command associated with this declaration is ‘do α !’.

certain proposition true, the authority makes that proposition true. But now we can ask what logical relationships between propositions preserve that kind of making true?

4.2 Illocutionary Entailment and Declarations

Central to the creation of norms, on Searle's view, is the illocutionary act of declaration. If there is going to be a genuine logic of institutional norms, then it must describe some relationship between the norms themselves, so it must describe a relation between the success conditions for status function declarations since those declarations are the norms. Illocutionary entailment provides a way of understanding the relationship between those success conditions.

It would be convenient, however, if we could find some way to characterize the illocutionary entailments between declarations via another, better understood relation of consequence. The best understood relations of consequence that exist are those for *propositions* on one description or another. So we want to find a relation between propositions (or sets of propositions and propositions) characterizing illocutionary entailments between declarations. Let's call this relation \vdash_S . It would be spectacular if \vdash_S was a subrelation of classical consequence, and one that we could characterize in some mathematically precise manner. What we are looking for is a relation \vdash_S such that $\Gamma \vdash_S \varphi$ if and only if $\mathbf{D}[\Gamma] \models_I \mathbf{D}(\varphi)$. Ultimately, this relation will be (a restriction of) the consequence relation of strong implication from Vanderveken's work.

We will first note a necessary condition on \vdash_S that results because the contents of non-defective declarations can't be necessarily true or false. As mentioned in section 3.1.1 the propositional content conditions on declarations are rather interesting. First, no content of a declaration can be a proposition that is necessarily false or necessarily true. Thus the propositional content of any declaration must be something contingent. But that contingency must be of a special kind for promulgations. We often recognize the contingency of the proposition that grass is green, but it is counterintuitive to say that an institutional authority could *declare* that grass is green. An authority could, possibly, invent a new term 'green' and declare it to mean 'the colour of grass', but that isn't the same as declaring that grass is green. What is needed

is the recognition of a special part or segment of language that represents social reality, and is under the influence of those that create and maintain that institutional reality.

In order for declaration (promulgations) to be successful, an authority must have special access or control over certain bits of language. If we recognize a special vocabulary that can be included in the contents of declarations by the respective authorities, then that vocabulary can be the special bits of language that institutional authorities have control over.² This recognition of a special institutional language also makes sense of Vanderveken's requirement that the contents of promulgations must have a special *legal* content. We will say that this language is under *institutional control*.

This leads us to our first observation regarding the consequence relation \vdash_S . Suppose that some authority makes a declaration of the form $D(p)$. The p cannot be a necessary proposition. So if $D(p)$ illocutionarily entails $D(q)$, then q cannot be a necessary proposition either. So, if $\Gamma \vdash_S \varphi$, it is necessary that $\llbracket \varphi \rrbracket$ not be a necessary proposition, if $\mathbf{D}[\Gamma] \models_I \mathbf{D}(\varphi)$. This means that treating 'declared' as a modal operator results in a modal operator that doesn't distribute over classical consequence.

Before moving on with the rest of our justification we will look at Vanderveken's use of strong implication in relation to illocutionary entailment. Vanderveken's justification of strong implication is flawed, and we discuss that in section 4.3. However, we will argue in section 4.4 that strong implication does work for reasons different from Vanderveken's, and reasons that are particular to the case of declarations used in Searle's theory which is closer to promulgation than basic declaration.

²There will be more on this in section 5.1. In the construction of a formal language we introduce a special set of atomic sentences that serve as the institutional atoms, and a special set of agent-like terms called roles. We also introduce a relationship of institutional control that says when an authority has control over a certain bit of language.

4.3 Vanderveken on Strong Implication

Vanderveken's theory of illocutionary logic uses a consequence relation that he calls '*strong implication*' (1991, p. 38) which provides a consequence relation that declarations, and other illocutionary acts, will distribute over. Of course, in some cases illocutionary acts only distribute over a certain subrelation of strong implication because of the propositional content conditions on those illocutionary acts. For the most part, when Γ, φ meet the propositional content conditions, and \vdash_S is the relation of strong implication, then $\Gamma \vdash_S \varphi$, if and only if $\mathbf{F}[\Gamma] \models_I \mathbf{F}(\varphi)$.

Strong implication is a relation that encodes additional information about logical consequences. φ *strongly implies* ψ , when 1) $\varphi \vdash_{CL} \psi$, and 2) the propositional content of ψ is contained in φ . Vanderveken offers a formal account of propositional content and containment which we follow in section 5.2. The thought behind strong entailment is that requiring containment of propositional content in addition to logical entailment provides a relationship of equivalence between propositions such that they will be substitutable *salva felicitate*. That means: when φ strongly implies ψ , and vice versa, $\llbracket \mathbf{F}(\varphi) \rrbracket$ is successfully performed if and only if $\llbracket \mathbf{F}(\psi) \rrbracket$ is successfully performed. Recall that successful performance of a declaration implies the non-defective performance of a declaration.

The intuition behind strong implication and equivalence is that when the content of a logical consequence is contained in the premises, any agent performing illocutionary acts on the content of those premises also simultaneously apprehends the contents of propositions that are strongly implied by those premises. More specifically, if an agent believes p and p strongly implies q , then the agent apprehends q as well. This intuition is contentious, but we do not need it to make our point. We will outline the general debate and Vanderveken's position, then point out the issues with Vanderveken's position.

A fundamental problem in the philosophy of logic is the connection between logic and rationality.³ Is it irrational for A, a doxastic agent, not to believe that p when A has asserted that q and q classically entails p ? Definitely not: it would imply that A believed infinitely many

³Cf. Harman (2002), Field (2009).

propositions, e.g., $p \vee q$ for every other proposition p . Expecting a doxastic agent to believe all of the classical consequences of their beliefs is too high a standard for rationality.

Vanderveken recognizes this point, but he considers two principles to be fundamental to language use, and provide necessary conditions for a human to be a language user. These two principles are combined into what he calls “[t]he law of the rationality of the speaker” (1990, p. 141). “[*L*]anguage is the work of *reason*” he writes, reason “is *constitutive of linguistic competence*” (ibid.). This means that any speaker must have certain reasoning capabilities in order to have a command of language, i.e., in order to be capable of performing illocutionary acts. But since he says that reason is *constitutive* of linguistic competence, and these two principles are those for the rationality of the speaker, then these two principles taken together are necessary and sufficient for the level of rationality needed for linguistic competence.

The two principles that constitute the law of the rationality of the speaker are as follows:

- (MinCon) Each speaker is minimally consistent. The minimal condition is that speakers recognize the fact that illocutionary acts of the form $\mathbf{F}(\neg\varphi)$ and $\mathbf{F}(\varphi)$ cannot be successfully performed by the same speaker in the same context of utterance when the illocutionary point of $\llbracket \mathbf{F} \rrbracket$ is non-empty.
- (StComp) There is compatibility of strong implication with respect to illocutionary points with non-empty direction of fit. The idea is that when there is a strong implication between two propositions p and q , then a successful illocutionary act of the form $F(p)$ produces an illocutionary act of the form $F(q)$, provided that q meets the propositional content conditions required.

To justify these two principles, Vanderveken takes rather different tacks. For MinCon, Vanderveken reports that it has been confirmed by studies in cognitive psychology,⁴ thus, it has an empirical basis. StComp is given a different justification.

It is obvious that people’s beliefs are not closed under classical consequence, although it

⁴See Cherniak (1986).

is often assumed as an idealization in work on belief revision and doxastic logic.⁵ For Vanderveken, it is part of what it takes to be linguistically competent, however, to have one's illocutionary acts be closed under *strong* implication. Strong implication is supposed to capture what competent language users apprehend *a priori* as following from the propositional content of their illocutionary acts. It is worthwhile to consider this position in detail.

To start, let's look at what StComp commits us to in the case of assertion. Suppose that p strongly implies q . When A successfully asserts p , then A has achieved the illocutionary point of assertion on the proposition p : expressed the direction of fit between their words and the world holds, i.e., A has expressed that p is true. A has also expressed that they presuppose the necessary propositions needed for p to be true, and has expressed that they have the right mental states for asserting p , i.e., they express a belief in p —although they may not in fact believe p . Now StComp commits us to saying that A also has expressed that: q is true, they presupposed the necessary propositions needed for q to be true, and they believe q . To express these beliefs, however, Vanderveken is committed to saying that A apprehends q when apprehending p . So A must be aware of the strong implications of its assertions: they cannot express that they believe q without being aware of q .

Now that we have an example of StComp's implications, we contrast Vanderveken's condition (StComp) with another condition that is, perhaps, less controversial. If A asserts that p , then we would say that A is *committed* to the logical consequences of that assertion. The sense of 'committed' used here is not that A must believe the logical consequences of their assertion, but that when A is shown that something is a logical consequence of it, A should not deny its truth, given A's assertion. This conception of logical commitment is a *normative* requirement rather than a *descriptive* requirement—unlike Vanderveken's StComp condition. Vanderveken's StComp condition in this case would say that, if q is a strong implication of p , then A's successful assertion of p will result—eo ipso—in a successful assertion of q . Recall that one of the sincerity conditions of assertion on Vanderveken's theory is that A, the asserter,

⁵For example it is an assumption in Alchourrón et al. (1985) that belief sets are logical theories, i.e., closed under logical consequence. This is also the position taken by Hintikka (1962).

believe p . Thus in asserting p non-defectively, A also, in fact, believes q since A will have asserted q non-defectively as well. But there tends to be a general distrust of belief being closed under any collection of logical consequences.

Cherniak (1981b) credits the distrust of belief as being subject to closure under any non-trivial consequence relation⁶ at all to the assent theory of belief: A believes p iff A would assent to sentences expressing p . This theory of belief can clearly result in logical inconsistencies and logical gaps in the set of propositions that A believes. There would be gaps when there are propositions that follow logically from A's beliefs, but A wouldn't assent to sentences expressing those propositions because A has never considered those sentences. Similarly, A might assent to contradictory sentences that aren't conspicuously contradictory.

So Vanderveken's minimal rationality of the speaker must be interpreted as a description of what beliefs, intentions, and desires speakers have when non-defectively performing speech acts. Vanderveken isn't consistent on how the law of the rationality of the speaker is supposed to function, however. At one point he says

strong implication is *cognitively realized* in the minds of speakers. Whenever a speaker expresses a proposition in the performance of a speech act, he also expresses all propositions strongly implied by that proposition. . . a rational speaker cannot relate in thought a proposition to the world with the aim of achieving a success of fit from the direction of an illocutionary point without also relating to the world with the same direction of fit all weaker propositions which satisfy the propositional content conditions of that illocutionary point. Indeed, he and the hearer mutually know *a priori*, in virtue of their linguistic competence, that the truth of these weaker propositions is a necessary condition for the success of fit of the utterance. (Vanderveken, 1990, p. 143)

But at another he says of his project, which he calls 'General Semantics', that

⁶By trivial consequence relation we mean $\{ \langle \Gamma, \varphi \rangle : \varphi \in \Gamma \}$.

such an investigation of *cognitive aspects of meanings* is purely logical and relatively independent of empirical psychology. ...in thinking we do not produce propositions, rather we apprehend propositions with the mental capacities of the mind [and these propositions] are not private... General semantics deals only with the cognitive aspects of propositions which are related to linguistic competence, and neglects other empirical psychological aspects of the comprehension of meaning such as, for example, the contingent limitations due to memory. (Ibid. p. 84)

But linguistic competence is almost totally a question of the contingent abilities of humans, and quite a wide range of humans is linguistically competent. Thus, the questions of general semantics must be asked relative to the lowest common denominator of linguistic competence. But what the lowest common denominator is *depends* on empirical psychology, even if propositions are objective entities. Also, what level of linguistic ability constitutes linguistic competence isn't an a priori matter, it too should be established by investigating the variation of linguistic abilities.

But perhaps we can reinterpret Vanderveken's law of rationality of the speaker, particularly StComp. On Cherniak's view⁷ minimal rationality is a cluster concept; for A to be minimally deductively rational A just has to be able make *some* of a cluster of relevant inferences *some* of the time. Cherniak doesn't provide an account of what that cluster of inferences is, but he does suggest that the account is something that would apply to *human* rationality and language (see Cherniak, 1981a). So it might be suggested that Vanderveken's minimal rationality of the speaker should be interpreted as saying that the inferences validated as strong implications make up that cluster of relevant inferences for minimal rationality. But StComp can't act as a cluster concept as in Cherniak's theory, because a rational speaker on Vanderveken's view must apprehend *all* of the strong consequences. Otherwise there might be cases where some strong implication of an assertion was left unapprehended, and so unasserted, contrary to StComp. So Cherniak's version is not a way to interpret Vanderveken's position.

⁷See Cherniak (1981b).

But what are the alternative positions? One possibility is to change StComp so that it is a claim about *commitments*, in some weaker sense than actual success of implicit illocutionary acts. Altered in this way, A may not successfully perform illocutionary acts on the strong consequences of the contents of successful acts, but A is *committed* to those strong consequences of the contents of successful acts. This reinterpretation is problematic for two reasons. First, if we have changed the focus to commitment, then why not look at *all* of the propositions that are true given the contents of successful assertions? I.e., why not let $\vdash_S = \vdash_{CL}$, modulo propositional content conditions, rather than strong implication? Second, and more importantly, such a reinterpretation isn't going to help *us*. Recall that on our Searlean account, an institutional norm comes from a declaration, and that declaration must be *non-defectively performed* to be in force. So unless *commitment* is interpreted as *non-defective performance for declarations*, that move will not be of help. But to make that identification is to beg the question.

As Church points out,⁸ there is almost no end to how logically ignorant anyone might be, but still be a competent language user. At least that end isn't one that we can figure out from our armchairs. But we think that at least for promulgation-like illocutionary acts, and perhaps declarations more generally, we can argue that there is a non-trivial relation that \vdash_S can be interpreted by. Justifying the closure of special declarations under strong implication, or a subrelation thereof, via a different argument is the focus of the next section.

4.4 Strong Implication and the Specific Illocutionary Acts

What we want to show now is that in fact the particular speech acts discussed in section 3.1.2 are closed under strong implication. We will be concerned with the kind of declaration that Searle needs in his theory of institutions that we will call *declaration**, promulgation is an instance of *declaration**. What we need first is the following definition:

Definition 4.4.1 (Informal Normative Entailment). The relation \vdash_N represents norm entailment if and only if for all Γ, φ , that are under institutional control, $\Gamma \vdash_N \varphi$ iff $\mathbf{D}[\Gamma] \models_I \mathbf{D}(\varphi)$.

⁸Church (2009, p. 14).

So the relation \vdash_N will serve as an informal stand in for our notion of norm entailment. Note that \vdash_N is a relationship between propositions (or sentences that express propositions), not speech acts. We want to show that a non-defective declaration* of Γ implies a non-defective declaration* of φ , where φ is a strong implication of Γ and $\Gamma \cup \{ \varphi \}$ all satisfy the propositional content conditions. Recall that this is enough for the declaration* of φ to also be satisfied since the satisfaction conditions are coextensive with the non-defective performance conditions for performative-type speech acts, e.g., declarations. So what we have to show is that \vdash_N is the relation of strong implication. To show this we specifically need to show the following:

Observation 4.4.1. *(Informal) When Γ and φ are under institutional control for an authority, then*

1. *The illocutionary point of declaration* transfers between propositions that stand in the relation of strong implication,*
2. *the preparatory conditions for declaration* transfers between propositions that stand in the relation of strong implication,*
3. *The mode of achievement for declaration* transfers between propositions that stand in the relation of strong implication,*
4. *the sincerity conditions for declaration* transfers between propositions that stand in the relation of strong implication, and*
5. *the degrees of strength of those sincerity conditions for declaration* between propositions that stand in the relation of strong implication;*
6. *relative to all semantic interpretations.*

To establish this observation we first have to ask what the illocutionary point of declaration* is. Vanderveken doesn't give a detailed account of what the illocutionary point of promulgation is, simply that the point is to declare the content of the promulgation. By analogy, this means that the point of declaration* is for the speaker to represent itself as making the content of the

declaration* *the case*, and by so representing itself, it makes the content the case. For Searle and Vanderveken the illocutionary point is connected to the direction of fit, and the direction of fit for declarations in general is both of world-to-words and words-to-world. So the point of declaration* is to make the content of the illocutionary act true by fiat.

For a moment we will digress on this subject because there is a view that is worth mentioning. Other authors take a more liberal view of what occurs in the case of illocutionary acts like declaration*. Another way to think of the illocutionary act of a promulgation is to make the content of the promulgation correctly assertible *for the group* that the authority is an authority for. This view is held by Tuomela and Balzer (1999).

Tuomela and Balzer's work identifies the central property of social notions as being the products of collective acceptance and that for something to be collectively accepted it must be *true or correctly assertible-for-the-group*.⁹ We will use 'collective acceptance' to distinguish this notion from that of Searle's *collective recognition*. Tuomela and Balzer define the idea of collective acceptance as

COLLECTIVE ACCEPTANCE THESIS (CAT): A sentence *s* is collective-social in a primary sense in a group *G* if and only if (a) it is true for group *G* that the members of group *G* collectively accept *s*, and that (b) they collectively accept *s* if and only if *s* is correctly assertable (or true). (p. 181)

This makes the correct assertability of a proposition that is collectively accepted something that can be used by the group under any conditions in which the group, qua collective, persists. Put another way, as long as the institutional authority is recognized or the institution persists, the promulgated institutional facts can be used.

But Tuomela and Balzer say

CAT employs the notion of correct assertability (and truth as its special case) for the group. The forgroupness of an accepted, hence correctly [sic] assertable, sentence is the group's intentional attitude towards the accepted sentence, viz., precisely the group's taking that sentence to have been accepted for it by it. That a sentence

⁹This makes it a precursor to the work of Lorini et al. (2009).

is correctly assertable or true for G means, roughly, that the group members *qua* group members are entitled to treat it as correctly assertable or true in their various intellectual and practical activities in relevant group contexts, no matter whether the sentence is “objectively” true. (ibid., p. 182)

So if the illocutionary point of a promulgation is the collective correct assertability of institutional facts (i.e., the contents of the promulgation), then it is possible to promulgate *false* propositions. It also removes any propositional content conditions for promulgations. Indeed, there is an example of an attempt to make a *necessary* falsehood correctly assertable in the infamous Indiana Pi case.¹⁰ However, this is not within Searle’s view, so we will set it aside. For Searle, correct assertability is a side effect of declaration*.

In our current theory, when an authority has institutional control over some proposition, a promulgation of that proposition becomes correctly assertible *because* it is made *true*. But in cases where the authority doesn’t have institutional control, e.g., scientific and necessary propositions, then the authority can’t declare them since they will not satisfy the propositional content conditions.

Given that the illocutionary point of a promulgation is to make the contents true, we can see the following fact:

$$\mathbf{D}[\Gamma] \models_I \mathbf{D}(\varphi) \implies \Gamma \vdash_{CL} \varphi$$

i.e., all illocutionary entailments are classical consequences. That also means that $\vdash_N \subseteq \vdash_{CL}$. The reason for this containment is as follows. If the point of a declaration* of φ is to make the contents φ true, then whatever implicit declaration*s are made require that the same illocutionary point be achieved on their contents as well. But the only propositions that are true because of the declaration* of φ *in every semantic interpretation* are the classical logical consequences of φ . Thus, if $\Gamma \vdash_N \varphi$, then $\Gamma \vdash_{CL} \varphi$.

¹⁰In the Indiana Pi case a bill was put before the General Assembly (the state legislature) saying, in effect, that the circle could be squared via a particular method. However, it followed from that method that the value of Pi would be $\frac{16}{5}$. The bill was passed by the General Assembly, but was taken to the Indiana State Senate. There the bill was indefinitely postponed after national press were alerted to the bill. Ogilvy (1956, pp. 118–9)

Now we can demonstrate that \vdash_N is the relation of strong implication. What we just noticed is that the illocutionary point is achieved on all of the classical consequence, so *a fortiori* it is achieved on all of the strong consequences. What we have to show is that it is achieved only on the strong consequences.

As a thought experiment consider the legal system of ancient Rome. Now take any φ expressing a status function from that system. Now suppose ψ says ‘Making duplicate digital recordings of Lady Gaga is a violation of the laws of Caesar’. Then $\varphi \vee \psi$ is a logical consequence of φ , so indeed $\varphi \vee \psi$ would be true given that the laws of ancient Rome are in force. Although $\varphi \vee \psi$ is harmless enough, it would be odd to claim that ψ was really something that would be intelligible to the law makers of ancient Rome. But it isn’t just odd, that institutional fact isn’t part of the institution that was Rome.

The reason that $\varphi \vee \psi$ isn’t an institutional fact has to do with the part that Toumela and Balzer get right: the truth is *for the group*. Recall from section 3.2 that Searle uses a notion of collective *recognition* rather than acceptance—‘acceptance’ has connotations of agreement or endorsement that Searle would like to avoid—to explain the collective or social nature of social institutions. Thus collective recognition can be tacit and indifferent awareness. This gloss of collective recognition doesn’t mean that the institutional facts can be unintelligible to the individuals that make up the institution. Particularly, the institutional facts must be intelligible to the authority, or the individuals that constitute the authority. Institutions are mind dependent since they require individuals’ minds to persist, and institutional facts whose propositional content is foreign to those individuals who maintain and create the institution couldn’t be properly realized by the individuals in the group. An institutional fact that has content *no* patient or subject of the institution understands isn’t a fact that belongs to *that* institution.

A formal way of restricting the implicit institutional facts given the explicit institutional facts is via strong implication. Strong implication is the consequence relation that encodes 1) logical consequence and 2) containment of propositional content of the conclusion in the content of the premises. To guarantee that none of the propositional content of implicit declarations is foreign

to the explicit institutional facts, we can require that the propositional content of the implicit declaration* is contained in the propositional content of the explicit declaration*s. Thus, the set of pairs $\langle \Gamma, \varphi \rangle$ that the illocutionary point of declaration* is achieved on is at most the relation of strong implication. This establishes 1 from observation 4.4.1.

Conditions 2–5 remain to be shown. Now conditions 4 and 5 involve sincerity conditions, and as we said in section 3.1.2 sincerity conditions aren't intelligible for promulgations, or they won't play a role. The reason for that has to do with the kind of relationship that declaration* has to its use. Vanderveken says,

the *successful* performance of of an illocutionary act with the primitive illocutionary force of declaration is necessarily *non-defective*. Indeed, if the speaker makes the propositional content true in a successful declarative utterance, then he has the capacity to make it true. Thus, the preparatory conditions obtain in the context of his utterance... Indeed, he cannot mean to make the propositional content true in virtue of his utterance without *eo ipso* believing, desiring, and intending his utterance to bring out success of fit between language and the world. (1991, p. 73)

So declarations are not only satisfied when successful, they are also non-defective. But desiring or believing that the utterance bring about a success of fit between the content of the utterance and the world is to desire and believe that content, i.e., the proposition declared. Although Vanderveken's claim may hold for some declarative-type illocutionary acts, it can't hold for those needed in Searle's theory. All that Searle's theory requires in sincerity conditions is what we will call *recognition*, derived from his notion of 'collective recognition'.

Declarations allow people to create institutional reality, so they allow us to introduce institutional facts where they didn't exist before. That requires successful status function declarations when there are no preexisting status functions. Recall that Searle's theory requires collective recognition of status functions. However, that collective recognition can be anything from enthusiastic endorsement to grudging acquiescence.

Sincerity conditions are attitudes or mental states of the agent(s) performing an illocutionary

act that must be expressed for the agent(s) to succeed in performing that illocutionary act. But if all that is required to sustain or even impose a status function is grudging acquiescence to that status function, that can't be something usually called 'a desire' for that status function. Thus desire is not necessarily connected to the success conditions of declarations.

Belief in a status function, qua the content of a status function declaration, is also problematic. In cases where there is no preexisting institution belief in the status function is necessary since there is no way for the status function to take effect without the collective recognition of the status function. However, in the case of status function declarations that go beyond those introducing institutions for the first time, the ability to impose a status function via declaration is held despite the speaker's inclinations and attitudes. The crucial conditions for success are the preparatory conditions: a legislator may enact legislation they think is false, e.g., by denying certain status functions to certain groups of people they think should actually have that status. In short: declarations can succeed despite the speaker's attitudes.

Finally there are the speaker's intentions. Again this is a case where the power to declare overcomes any attitudes. Thus there are two options. First, the sincerity conditions are non-existent, or, second, the degree of strength of those sincerity conditions must be neutral. If the degree of strength is neutral, it is satisfied by no belief (desire, intention) and stronger attitudes toward the propositional content. Either way the mental states of the speaker do not interfere with the success of declarations of the kind Searle needs.

Since the mental states don't matter, whether they are transferred across any consequence relation is irrelevant to the success of a declaration of the kind Searle needs. Thus, 4 and 5 are established. So all that remains to be established is that the mode of achievement and the preparatory conditions transfer, i.e., conditions 2 and 3.

Let's make a rather general, and perhaps obvious, point: illocutionary consequence will be—like standard consequence—*instantaneous*. There is no time-lag between the truth of premises and the truth of consequences. So when some φ is promulgated, it is at that moment that all of its consequences become true, if they weren't true already. That means all

implicit promulgations are also promulgated in the very moment that φ is promulgated, and any implicit promulgations that accrue from the combinations of explicit promulgations.

The instantaneous nature of implicit promulgations affects the preparatory conditions and the transfer of the mode of achievement in the following way. The preparatory conditions for a promulgation are that there is an institutional authority that is collectively recognized as such that can make promulgations. But such preparatory conditions are not specific to particular promulgations. The existence of an authority is a general position that has control over all propositions that are under that authority's institutional control. So when the preparatory conditions for one promulgation are met, then they are immediately met for all implicit promulgations. So the simultaneity makes the preparatory conditions met for strong consequences of propositions that have been promulgated.

For the mode of achievement the argument is similar. In making a promulgation, an authority must invoke the collective recognition that it is an authority. This is the special mode of achievement for promulgations. In invoking its authority, it makes the promulgation a collective act, i.e., that the content of the act is to be recognized by all of those who recognize the authority. That means, then, that they, by recognizing the authority, recognize its power to make the content of the promulgation true. But this ability extends to all of the propositions that are under institutional control. Thus, by assumption about the contents of the propositions in observation 4.4.1, we can conclude that that ability extends to φ , and φ is assumed to be a strong consequence of Γ . Therefore, we have demonstrated observation 4.4.1, and that \vdash_N is the relation of strong implication.

One more aside concerning the propositional content conditions before we move on. Just because the contents of declarations can't be necessary propositions doesn't mean that they can't be tautologies. Since the contents of promulgations could have not existed, i.e., had they not been promulgated, they are not necessarily true, and that means that they are not necessary propositions. There may not have been anything such as a Prime Minister of Canada, for instance. So the truth and falsity of institutional propositions or facts is institution dependant.

That means that tautologies that are under institutional control will also be institution dependent. So all tautologies that are under institutional control can also be the contents of implicit, or even explicit, promulgations.

4.5 Summary

In this chapter we have discussed the general problem of what a notion of norm consequence looks like from an informal perspective. We based our notion of norm consequence on illocutionary consequence because Searle's conception of institutional facts makes institutional facts the results of certain illocutionary acts, viz., declarations or promulgations. We argued that Vanderveken's account of why strong implication characterized illocutionary entailment was insufficient, but gave an independent justification of strong entailment as a characterization of illocutionary entailment in the cases we are interested in. In the sequel, strong implication will be given a formalization and used to underpin a formalization of norm entailment, see chapter 6.

Part II

**Formal Logic and Formal Dynamics of
Institutions**

Chapter 5

Constructing a Formal Language

...every language has, as Mr Wittgenstein says, a structure concerning which in the language, nothing can be said, but that there may be another language dealing with the structure of the first language, and having itself a new structure, and that to this hierarchy of languages there may be no limit.

Bertrand Russell Introduction to the *Tractatus*

So far we have argued that there is a logic of institutions, but it will depend on what philosophical account of institutions is correct. We have chosen a particular account, viz. Searle's, and we have discussed what consequence relation on propositions will mirror the consequence relation of in forceness on this conception of institutions. Now we will add some formal meat to these philosophical bones.

In this chapter we deal with two things 1) what is needed for a reasonably expressive object language to represent institutional facts, and 2) represent the formal counterpart for strong implication. In the first sections we discuss institutional facts and their representation. That includes discussing action and obligation. We conclude that obligation and institutional duty are different things—at least on this conception—and we introduce Anderson's reduction to deal with institutional duty. We also discuss logics of action, but decide to use a version of stit logic since it is less problematic than other logics of action, philosophically speaking.

In the final section we formalize strong implication and give a complete Hilbert style calculus for it. In the next chapter we will put these elements together to formalize norm entailment, and show the expressive capabilities of our language.

5.1 A Formal Language of Social Reality

According to Searle, institutions come about through a collective recognition of the abilities of some (or all) to perform certain speech acts. Through these special speech acts people can

create new kinds of actions and objects: speeding, judges, universities, contracts. Apart from the presumed correctness of Searle's account of institutions, we assume that there are special, institutional uses of terms. For example, written legal codes contain a set of definitions that define the relevant roles, subjects, actions, and situations. If these definitions are not written down, they are generally understood or assumed. We can see examples of this practice in the legal definitions of words like 'organic'.

To see exactly what we need to represent, let's recall Searle's general structure of status function declarations:

We (or I) make it the case by Declaration that a Y status function exists in C and in so doing we (or I) create a relation R between Y and a certain person or persons, S, such that in virtue of SRY, S has the power to perform acts (of type) A. (Searle, 2010, pp. 101–2)

In this definition there is the *we (or I)*, *declaration*, *context C*, *relation R*, *persons S*, and *power to perform acts A*. So to properly represent institutional facts we have to account for these things. The first three items on the list we can leave aside.

In the formalism we don't represent the authorities nor do we represent the acts that bring new institutional facts into existence. So the *We* and the *declaration* don't enter into the formalism. The focus for the representation is on the content of institutional facts. We are not concerned with representing either the processes that give rise to institutions or how the individual institutional facts come into existence. We focused on those questions in the previous section, and there we argued that the individual institutional facts obeyed the consequence relation of strong implication. So there is no need to represent those items in the formalism.

In section 3.2 we argued that the context, at the highest level, is defined by the existence of the institution. That means we needn't include the context in the formalism. This leaves us with representing the relation(s) R, the persons S, and the powers to perform actions. We will start by discussing how to represent actions.

Our formalism requires some representation of action, but we would like to remain as philosophically neutral with respect to action as possible. The most philosophically neutral logic of action is given by the “seeing-to-it-that” framework of Belnap and Perloff (1992). In this framework, an agent or set of agents see to it that a proposition becomes true. Whether that is via a particular action, or through some choice is really irrelevant. What is important for reasoning about action is 1) that an agent stands at the causal nexus of some change in the world via their choices or actions, and 2) that the formulas of formal language represent which agents can bring about which changes.

There are many logics of action, but we are interested in using as simple a formalism as possible that still represents certain intuitive properties of actions and how they interact with the world. In the next section we will introduce the specific stit formalism, called xstit, that we will use. At the end of the next section we will discuss how this formalism meets our ends and how it deals with the persons S from above. In section 5.1.2 we will discuss institutional facts and the relations R .

5.1.1 The Xstit Formalism

The stit language that we use is called xstit in Broersen and Meyer (2011), the reader can compare this logic to standard stit logics in Belnap and Perloff (1992). In xstit logic a group of agents’ actions or choices determine possible future states in the state following the current state of the world. The language is constructed from a set of agents \mathbf{Ag} and atomic propositions from \mathbf{At} . Sets of agents are denoted by capital Roman letters $\mathbf{A}, \mathbf{B}, \mathbf{C}$. We can then define a language $\mathcal{L}_{\text{xstit}}$ as follows:

$$\varphi := \mathbf{p} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [\mathbf{A} \text{ xstit}] \varphi \mid X\varphi$$

Informally $[\mathbf{A} \text{ xstit}] \varphi$ means that \mathbf{A} sees to it that φ in the next state. $\Box\varphi$ means that φ is historically necessary. That just means, however, that φ is true relative to every history *at this moment*. This will become clearer when we introduce the semantics. $X\varphi$ means that in the next state relative to the history we are in, φ is true.

Like in Belnap and Perloff (1992), the semantics evaluates formulas relative to histories *and* moments. A history is a set of moments that is linearly ordered. This is modelled as follows:

Definition 5.1.1. An xstit frame is a triple $\mathfrak{F} = \langle S, H, E \rangle$ such that:

1. $S \neq \emptyset$ are called the *static states*.
2. $H \neq \emptyset$ is a set of ordered sets $\langle h, <_h \rangle$ such that for each $h \in H$
 - (a) $h \subseteq S$ and $\langle h, <_h \rangle$ is isomorphic to \mathbb{Z} with its usual order, and
 - (b) if $s \in h \cap h'$, then $\{s' : s' <_h s\} = \{s' : s' <_{h'} s\}$. Since each order is isomorphic with \mathbb{Z} , there is a unique successor and predecessor in h for each $s \in h$, we refer to these by $\text{lub}(s, h)$ and $\text{glb}(s, h)$, respectively. We can generalize these concepts in the following way: $\text{glb}(s) = \{\text{glb}(s, h) : s \in h \ \& \ h \in H\}$ and $\text{lub}(s) = \{\text{lub}(s, h) : s \in h \ \& \ h \in H\}$. These give the set of successors and predecessors of s , respectively.¹
3. $E : S \times H \times \mathcal{P}(\mathbf{Ag}) \rightarrow \mathcal{P}(S)$, the h -effectivity function, assigns a set of static states to each triple (s, h, \mathbf{A}) . It must obey the following conditions:
 - (a) If $s \notin h$, then $E(s, h, \mathbf{A}) = \emptyset$
 - (b) If $s' \in E(s, h, \mathbf{A})$, then $s' \in \text{lub}(s)$ (i.e., $E(s, h, \mathbf{A}) \subseteq \text{lub}(s)$)
 - (c) If $s \in h$, $\text{lub}(s, h) \in E(s, h, \mathbf{A})$
 - (d) If $s \in h$, $E(s, h, \emptyset) = \text{lub}(s)$
 - (e) If $s \in h$, then $E(s, h, \mathbf{Ag}) = \{\text{lub}(s, h)\}$
 - (f) If $\mathbf{A} \subsetneq \mathbf{B}$, then $E(s, h, \mathbf{B}) \subseteq E(s, h, \mathbf{A})$
 - (g) If $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $s \in h \cap h'$, then for some h'' with $s \in h''$, and $E(s, h'', \mathbf{A}) \subseteq E(s, h, \mathbf{A})$ and $E(s, h'', \mathbf{B}) \subseteq E(s, h', \mathbf{B})$.²

¹The rationale behind lub and glb is to put the terminology in line with that from order theory.

²The condition g) in Broersen and Meyer (2011) is different in that it only requires that $E(s, h, \mathbf{A}) \cap E(s, h', \mathbf{B}) \neq \emptyset$, and that doesn't seem to suffice.

If $\mathfrak{F} = \langle S, H, E \rangle$ is an xstit frame, then each history-static state pair (s, h) is called a dynamic state, and the domain of \mathfrak{F} , denoted $|\mathfrak{F}|$, is the set of dynamic states such that $s \in h$.

We will pause to explain the conditions in this definition. The first, 1, condition is standard in modal logic. Condition 2a says that we can order each history like \mathbb{Z} : $\dots s_{-2}, s_{-1}, s_0, s_1, s_2, \dots$ the set of integers. 2b says that if two histories share a static state in common they, they have each previous static state in common as well. This also means that once two histories diverge, they will not join other histories. This means the histories form a forest; the histories can be a collection of trees. We call $\mathcal{L}_{\text{xstit}}$ frames/models that are made up of trees *regular* frames/models. In what follows, we will assume that the histories form just one tree, and call such a frame/model a regular, *universal* frame/model. This should conjure images of S5 being a logic complete for the class of equivalence relations, and universal relations. The logic of xstit is complete with respect to the class of regular models and the class of regular universal models.

The notion of effectivity functions comes from coalition logic (cf. Pauly (2001)), and in general represents what a group of agents is capable of bringing about. $E(s, h, \mathbf{A})$ represents \mathbf{A} 's choice at the static state s relative to the history h ; $E(s, h, \mathbf{A})$ is the set of next possible states that the world might evolve into given that \mathbf{A} acts in accord with history h . Since $E(s, h, \mathbf{A}) \subseteq \text{lub}(s)$, by condition 3b, for each $s \in S$, the effectivity function selects a set of states from the totality of possible continuations given the current static state. If $s' \notin E(s, h, \mathbf{A})$ but $s' \in \text{lub}(s)$, then s' isn't one of the next states given that \mathbf{A} has chosen relative to h .

Each effectivity function is evaluated at a dynamic state (s, h) : a history-static state pair. Condition 3a states that it only makes sense for agents to be effective for anything at dynamic states where the static state is in the history, i.e., $s \in h$. This condition is mainly to make E a *function*, rather than a partial function. 3b says that the only states that agents can be effective for bringing about are those that follow in some history running through the current static state. Essentially, agents can only constrain the outcomes, not create new ones.

3c says that any set of agents is effective to constrain the outcomes to at least the immediate successor—relative to h —of the current static state. 3d says that the effectivity of the empty set

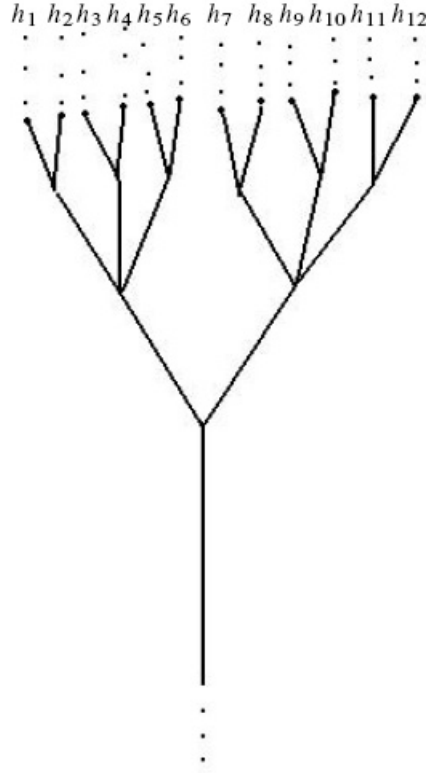


Figure 5.1: Universal, Regular \mathcal{L} -model

of agents is all of the possible continuations from a static state. The empty set is considered to be Nature's effectiveness; Nature sets the range of possible outcomes.

3e requires that the total set of agents, \mathbf{Ag} , determines the successor state of s at h for each $(s, h) \in |\mathfrak{H}|$. The next state in a history is completely determined by the whole set of agents. Broersen and Meyer point out that although the next static state is determined by the set of all agents, static states are only half of the auxiliary parameters in the evaluation of formulas. The set of agents doesn't determine the next *dynamic state*.

3f states that the more choices that are made, the more the outcomes are constrained. That results in the anti-monotonicity of effectivity functions. Finally, 3g states that the choices of agents never determine what other agents can choose. This is referred to as *independence of agency*. In figure 5.1 we have an image of the tree structures that act as frames for $\mathcal{L}_{\text{xstit}}$ models.

The models of xstit are given as follows:

Definition 5.1.2. An xstit model \mathfrak{M} is an xstit frame \mathfrak{F} with a valuation $v : \mathbf{At} \rightarrow \mathcal{P}(S)$.

We can then give the semantics for the language $\mathcal{L}_{\text{xstit}}$:

Definition 5.1.3. Truth or satisfaction of a formula in $\mathcal{L}_{\text{xstit}}$ relative to a model \mathfrak{M} and $(s, h) \in |\mathfrak{M}|$ is defined by:

1. $(s, h) \models \mathbf{p}$ iff $s \in v(\mathbf{p})$
2. $(s, h) \models \neg\varphi$ iff $(s, h) \not\models \varphi$
3. $(s, h) \models \varphi \wedge \psi$ iff $(s, h) \models \varphi$ and $(s, h) \models \psi$
4. $(s, h) \models \Box\varphi$ iff for all h' with $s \in h'$, $(s, h') \models \varphi$
5. $(s, h) \models X\varphi$ iff $(lub(s, h), h) \models \varphi$
6. $(s, h) \models [\mathbf{A} \text{ xstit}] \varphi$ iff $s' \in E(s, h, \mathbf{A})$ and $h' \ni s'$ only if $(s', h') \models \varphi$

Satisfiability of a set of formulas Γ is defined as: there is some model \mathfrak{M} and $(s, h) \in |\mathfrak{M}|$, such that $\mathfrak{M}, (s, h) \models \varphi$ for each $\varphi \in \Gamma$. In short we write $\mathfrak{M}, (s, h) \models \Gamma$. A set Γ *xstit entails* a formula φ ($\Gamma \models_X \varphi$) iff for each xstit model \mathfrak{M} and $(s, h) \in |\mathfrak{M}|$ such that $\mathfrak{M}, (s, h) \models \Gamma$, $\mathfrak{M}, (s, h) \models \varphi$.

Something to note about this semantics is that when a non-modal formula, that is a formula without xstit operators, X or \Box , is true at a dynamic state, then it is true relative to all dynamic states that share that static state. Formally, this means if $\mathfrak{M}, (s, h) \models \varphi$, and φ is a non-modal formula, then $\mathfrak{M}, (s, h') \models \varphi$ for all h' with $s \in h'$. This can be captured by a simple formula $\mathbf{p} \supset \Box\mathbf{p}$. This logic can be axiomatized by the following set of axiom schema for a Hilbert style proof theory.

Definition 5.1.4. Assume that $\mathbf{A}, \mathbf{B} \subseteq \mathbf{Ag}$, $\mathbf{p} \in \mathbf{At}$ and $\varphi, \psi \in \mathcal{L}_{\text{xstit}}$,

(p) $\mathbf{p} \supset \Box\mathbf{p}$

S5 for \Box :

$$\mathbf{K} \quad \Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$$

$$\mathbf{T} \quad \Box\varphi \supset \varphi$$

$$\mathbf{4} \quad \Box\varphi \supset \Box\Box\varphi$$

$$\mathbf{B} \quad \varphi \supset \Box\neg\Box\neg\varphi$$

KD for each $[\mathbf{A} \text{ xstit}] \varphi$ and X :

$$\mathbf{KA} \quad [\mathbf{A} \text{ xstit}](\varphi \supset \psi) \supset ([\mathbf{A} \text{ xstit}]\varphi \supset [\mathbf{A} \text{ xstit}]\psi)$$

$$\mathbf{DA} \quad [\mathbf{A} \text{ xstit}]\varphi \supset \neg[\mathbf{A} \text{ xstit}]\neg\varphi$$

$$\mathbf{KX} \quad X(\varphi \supset \psi) \supset (X\varphi \supset X\psi)$$

$$\mathbf{DX} \quad X\varphi \supset \neg X\neg\varphi$$

$$(\mathbf{DetX}) \quad \neg X\neg\varphi \supset X\varphi$$

$$(\mathbf{\emptyset=SettX}) \quad [\mathbf{\emptyset} \text{ xstit}]\varphi \equiv \Box X\varphi$$

$$(\mathbf{Ag=XSett}) \quad [\mathbf{Ag} \text{ xstit}]\varphi \equiv X\Box\varphi$$

$$(\mathbf{C-mon}) \quad [\mathbf{A} \text{ xstit}]\varphi \supset [\mathbf{A} \cup \mathbf{B} \text{ xstit}]\varphi$$

$$(\mathbf{Indep-G}) \quad \Diamond[\mathbf{A} \text{ xstit}]\varphi \wedge \Diamond[\mathbf{B} \text{ xstit}]\psi \supset \Diamond([\mathbf{A} \text{ xstit}]\varphi \wedge [\mathbf{B} \text{ xstit}]\psi) \text{ where } \mathbf{A} \cap \mathbf{B} = \emptyset.$$

For the rules we have modus ponens (MP) and the necessitation rule: If $\vdash \varphi$, then $\vdash \clubsuit\varphi$ for $\clubsuit \in \{\Box, X, P\} \cup \{[\mathbf{A} \text{ xstit}] : \mathbf{A} \subseteq \mathbf{Ag}\}$. Note that $\Diamond = \neg\Box\neg$. In Broersen and Meyer (2011) a proof sketch is provided that the axioms from definition 5.1.4 are complete with respect to the semantics from definitions 5.1.3 and 5.1.2. This provides a basic logic of action that we use in the sequel. For a comparison of this logic with the usual stit logic of Belnap and Perloff (1992) see Broersen and Meyer (2011). We choose xstit logic instead of regular stit logic because of the existence of the above recursive axiomatization. The stit logic for groups of agents of Belnap and Perloff (1992) has been shown not to be finitely axiomatizable, and the satisfiability problem is not decidable, see Herzig and Schwarzentruher (2008). Although we don't give a finite axiomatization, we conjecture that xstit logic is decidable, and a certain sublogic is finitely axiomatizable.

Stit-type logics have a long history. They start really with Kanger and Kanger (1966) and Pörn (1977), but have been the object of thorough study in the past two decades. The particular logic that we have chosen accords with certain intuitions about action in a non-deterministic world. Each choice that a set of agents makes results in moving the world in a certain direction, i.e., favouring certain histories over others.

We don't want to engage in a defense of stit logic in general as an appropriate logic of action, that has been done in Belnap and Perloff (1992, ch. 1–3), and in Horty (2001, ch. 1–2). We do have to argue that xstit satisfies the same conditions of the logics of those previous works. Since stit operators $[A \text{ stit} : \varphi]$ say that A sees to φ *now*, and $[A \text{ xstit}] \varphi$ says that A sees to φ in the next state, we have to reassure ourselves that things are roughly the same in the xstit cases. We will briefly explain the major difference in the models of stit and xstit.

In a model for stit, choice is instantaneous, so a set of agents' choices is represented by partitioning the set of histories present at a static state, denoted $C(A, s)$. Each possible choice available to an agent or set of agents is represented by a cell in that partition. Agents choose in accord with histories, and the effect of an agent's choice in accord with a history h is represented by the cell of the partition in which h occurs, denoted $C(A, s, h)$. There are three crucial conditions on $C(A, s)$. First, if histories h and h' do not separate at s (i.e., $\text{lub}(s, h) = \text{lub}(s, h')$), then they must be in the same cell of $C(A, s)$, i.e., $C(A, s, h) = C(A, s, h')$. Second, the choices of agents are independent, meaning that for any choice open to an agent, there is some way for that agent to realize that choice regardless of what the other agents choose. Formally, this is expressed as: whenever we take a set of cells X , one from each $C(\{a\}, s)$ for $a \in Ag$, $\cap X \neq \emptyset$. Finally, the effect of the choices of a group distributes evenly. Formally we can say this as $C(A, s, h) = \cap_{a \in A} C(\{a\}, s, h) \neq \emptyset$.

In the xstit case things are slightly different. The first condition above is irrelevant since $E(s, h, A)$ is a set of static states rather than histories so there is no way to separate histories using the effectivity function. The second condition is mimicked in condition g. The crucial differences between xstit and stit are in the third condition above and the way atoms are eval-

uated. In an xstit model $E(s, h, \mathbf{A})$ may not be identical with $\cap_{\mathbf{a} \in \mathbf{A}} E(s, h, \{\mathbf{a}\})$, so the third condition doesn't translate to xstit. In a stit model the truth of an atom is history dependent, so \mathbf{p} may be true relative to (s, h) , and false relative to (s, h') although both $s \in h \cap h'$. The truth condition for $[\mathbf{A} \text{ stit} : \varphi]$ is given by

$$(s, h) \models [\mathbf{A} \text{ stit} : \varphi] \iff \forall h' \in C(\mathbf{A}, s, h), (s, h') \models \varphi.$$

This definition is similar to that for $[\mathbf{A} \text{ xstit}] \varphi$, but uses $C(\mathbf{A}, s, h)$ of course. Note that the instantaneousness of choice is represented because φ is evaluated at (s, h') which is the same static state that $[\mathbf{A} \text{ stit} : \varphi]$ is evaluated at. Now that we have seen the similarities and differences between the formalisms of stit and xstit, we will move onto our initial discussion and look at four things: parsimony of ontological commitment, ability, refraining, and responsibility. The xstit formalism is adequate on all of these considerations. We will deal with each in that order.

In other work offering logical foundations for institutions, i.e., Castañeda (1975) and Lorini et al. (2009), there is a great deal of philosophical discussion in the former case, and little in the latter case. However, Castañeda introduces a rather complicated ontology of language to deal with his logic of action and obligation. There are new kinds of proposition-like entities that are the senses of imperatives, and other semantics entities, but our ontology isn't so diverse. Using Searle's foundation of institutions requires only propositions, and acts involving those propositions, and we will need special kinds of propositions that are under institutional control, but those are still among the same category of semantic entity. To represent our philosophical foundations, then, it is best to do it with no more than the entities we are committed to in our philosophical theory. The xstit language allows us that conservative sentiment.

In the xstit theory there are only agents, choices, and propositions. The effectivity functions are a very abstract representation of action and choice and so are ideal for not begging any questions about the ontology of actions. The xstit theory also takes group action as primary. So the action of an individual \mathbf{a} at a dynamic state (s, h) is given by the action $E(s, h, \{\mathbf{a}\})$ of the degenerate group $\{\mathbf{a}\}$. The conditions on group action are also rather meager. Condition 3f represents that the larger the group, the more they are capable of constraining the outcome

of the future. And 3g ensures that non-overlapping groups are free to realize their potential, meaning that individual agents have a partial independence of action. Each individual is able to realize at least some of each choice regardless of the choices of others. If groups share agents in common, however, one group may not be independent from another group's actions. So we will notice that this allows us to represent the persons S from Searle's schema above. So xstit only requires the ontological commitments that are needed, and is more parsimonious than previous theories.

Representing ability is straightforward: $\Diamond [A \text{ xstit}] \varphi$ is read as 'it is possible for A to see to it that φ in the next state. Harty (2001) has an extended discussion of what the logic of ability should be like (see pp. 19–33). However, many of the principles suggested there require that the stit operator be instantaneous, i.e., the result applies to the moment of evaluation. As a result the stit operators can be iterated in a way that carries a different meaning from iterating xstit operators. $[A \text{ xstit}] [A \text{ xstit}] \varphi$ is very different in meaning from $[A \text{ stit} : [A \text{ stit} : \varphi]]$. The former says 'A can see to it that in the next state that A can see to it in the next state that φ ', whereas the latter says 'A can see to it that A sees to it that φ '. $[A \text{ xstit}] [A \text{ xstit}] \varphi$ makes reference to A's ability two states into the future. But $[A \text{ stit} : \varphi]$ implies both $[A \text{ stit} : \varphi]$ and φ ; the former implies neither.

But this expressive disability isn't a concern since $\Diamond [A \text{ xstit}] \varphi$ still gives a proper sense of ability. $\Diamond [A \text{ xstit}] \varphi$ represents that A's choice now leads to the truth of φ in the next state. Whether A is responsible for the truth of φ is another matter, and we will deal with that in a moment. $\Diamond [A \text{ xstit}] \varphi$ is enough to capture ability.

Correlated with action is *refraining*.³ There is a sense in which simply not going to class and refraining from going to class are different, although the second implies the first. The sentence $\neg [A \text{ xstit}] \varphi$, says that A doesn't see to it that φ in the next state, but that is simply not seeing to φ . Whereas to refrain requires some possibility of doing the thing, even though it isn't done. Harty characterizes this notion as follows: $[A \text{ stit} : \neg [A \text{ stit} : \varphi]]$, i.e., A sees to it that A doesn't

³Harty (2001, p. 25) points out that von Wright (1963, p. 45) characterized refraining in the way he describes and the way we represent in this essay.

see to it that φ . \mathbf{A} is active in ensuring that \mathbf{A} doesn't see to it that φ . But as we noted above, iterating xstit operators has a very different meaning than iterating stit operators. However, Horty offers another account of refraining in terms of not seeing to it that φ , but being able to do so: $\neg[\mathbf{A} \text{ stit} : \varphi] \wedge \Diamond[\mathbf{A} \text{ stit} : \varphi]$. This is a direct translation of this interpretation of refraining. Since we can't iterate xstit operators without referring to what is happening two states into the future, the representation of refraining as $\neg[\mathbf{A} \text{ xstit}] \varphi \wedge \Diamond[\mathbf{A} \text{ xstit}] \varphi$ is preferred.

Finally we come to representing responsibility. Whether a tautology is true has nothing to do with what an agent does, usually. But tautologies hold at every moment-history pair, so they will hold at all the (s, h) such that $s \in \text{lub}(s')$ and $s \in h$, i.e., all the state/history pairs that follow s' , for any $s' \in S$. Of course $E(s', h, \mathbf{A}) \subseteq \text{lub}(s)$, so any tautology will be true at all history/state pairs in $E(s', h, \mathbf{A})$. But it seems odd to say that \mathbf{A} is responsible for the truth of the tautology. To capture responsibility we use the xdstit operator.

For these reasons, the deliberative xstit (or xdstit) captures the idea of being able to bring something about, but had that choice not been made, that something might not have happened. The xdstit operator is defined as $[\mathbf{A} \text{ xdstit}] \varphi \iff [\mathbf{A} \text{ xstit}] \varphi \wedge \neg \Box X\varphi$, we can give a semantic condition for this operator in the metalanguage as follows:

Definition 5.1.5. *[Deliberate xstit]*

1. $(s, h) \models [\mathbf{A} \text{ xdstit}] \varphi$ iff
- [P] 1) $\forall (s', h'), s' \in E(s, h, \mathbf{A}) \ \& \ s' \in h' \Rightarrow (s', h') \models \varphi$, and
- [N] 2) $\exists (s', h')$ with $s \in h', s' \in h'$ and $s' \in \text{lub}(s)$ such that $(s', h') \not\models \varphi$.

Regarding the problem of tautologies above, tautologies can never be false in these models, so $\neg \Box X\varphi$ couldn't be true if φ is a tautology since $X\varphi$ is true everywhere. Thus xdstit gives us a model of deliberate choice or action, in so far as deliberate action is something that merely could have been otherwise and our choice limits decisions.

The xdstit condition represents responsibility, but it might be suggested that $[\mathbf{A} \text{ xstit}]$ doesn't properly represent ability since \mathbf{A} doesn't really have the ability to make tautologies true. Even

though we have said that certain institutional actions bring about new tautologies, e.g., promulgations, those kinds of actions are not represented in the object language of our formal system. Nonetheless, ultimately we will use $[\cdot \text{ xstit}]$ operators for defining ability for institutional facts rather than $[\cdot \text{ xdstit}]$.

Now that we have our account of action and ability, we need a way of explaining the difference between social, institutional facts and, so-called, ‘brute’ facts.

5.1.2 Institutional Facts and Roles

As we noted in section 3.1.2 and in section 4.2, there are special propositional content conditions for status function declarations, and the contents of those declarations are what constitute institutional facts. So institutional facts have special propositional content. This special content is what institutional authorities have control over. We will call that kind of content *institutional content*. The kind of content that institutional authorities do not have control over we will call, following Anscombe (1958), *brute content*.

The distinction between these two kinds of content allows us to account for a few aspects of institutional facts. First is that there are special facts that just wouldn’t exist without being part of an institution, e.g., that Jeff *owns* savings bonds. That fact doesn’t make sense without institutions of private property, money, trade, et cetera. The formal language we will be using later is based on a *propositional* language, i.e., we will not deal with quantifiers or predicates. So in the current work we stick to a coarse grained level of idealized representation.

We will treat the fact that Jeff owns savings bonds as an atomic institutional fact. However, there are other brute facts as well. Consider ‘Dave has cancer’: that sentence is a basic fact, but it isn’t institutional (although it may have ramifications since it may entitle Dave to certain things like being given money to buy special medication to treat the cancer).

The formal model that we suggest to capture this distinction is to interpret the institutional content with a separate language from the brute language. Thus we introduce two distinct sets of atomic formulas: \mathbf{At}_B and \mathbf{At}_I (B for brute and I for institutional). We will distinguish these

by using $\mathbf{p} \in \mathbf{At}_B$ and $\mathbf{q} \in \mathbf{At}_I$. Thus we have given a meaning to the special institutional situations when we consider formulas constructed from \mathbf{At}_I and the boolean connectives, i.e.:

$$\varphi := \mathbf{q} \mid \varphi \wedge \varphi \mid \neg \varphi \mid \varphi \supset \varphi \mid \varphi \vee \varphi \mid \varphi \equiv \varphi$$

we will call this language \mathcal{L}_I , *the institutional language*. This language is really a fragment of a larger language that includes the brute facts.

The other aspect of institutional facts that we need to capture is what Searle describes as institutional facts being “intensional-with-an-s”(Searle, 2010, p. 119). Understanding this characteristic will allow us to represent the relations in Searle’s schema via institutional roles. Let’s consider Searle’s example:

1. As the winner of the 2008 presidential election, Barack Obama counts as the present president of the United States.
2. The president of the United States is identical with Michelle’s husband.
3. As the winner of the 2008 election, Barack Obama counts as Michelle’s husband. (Ibid.)

It is clear that (1) and (2) do not entail (3). Although our language doesn’t include an identity predicate, so we are unable to represent this argument, we just want to use it to make a point. Our point is that there is an intuitive distinction that we can represent, and helps with representing institutional codes. One might diagnose the problem with the argument above as attempting to make predication based on non-equivalent descriptions of the same object, i.e., Michelle’s husband and President of The United States. These two descriptions are accurate since there is one object holding two different *institutional roles*, i.e., Barack Obama. So we introduce into the object language a set of institutional roles.

Thus we introduce a new set of terms: the roles $\mathbf{Rol} = \{\mathbf{r}_1, \mathbf{r}_2, \dots\}$. These terms can be combined with xstit operators in the same way that we combine agent terms with xstit operators.

If $\mathbf{r} \in \mathbf{Rol}$, and φ is some formula, $[\mathbf{r} \text{ xstit}] \varphi$ is also a formula.⁴

⁴A very similar idea is used in Hansson (2001), but his rules contain predicates and variables that can be filled in with constant symbols. Here we are not using predicates, but the roles will act in a similar manner to variables.

But we will not add **Rol** to $\mathcal{L}_{\text{xstit}}$. There will be two separate languages, one that contains only the roles, and the other that only contains agents. We must make a brief aside to explain the rationale behind this separation. Institutional facts, according to Searle have a logical form that looks like: X counts-as Y in C. But in these count-as schema the Y terms are—as in the case of The President of The United States—spelled out with a deontology. This deontology would have to be given in general terms. It is only the imposition of the deontology that ties the institution, the institutional role, to the world. So the institutional facts are specified in abstraction from the “real world”. It is only upon imposition via declarations that they become connected to the world, i.e., individuals become affected by the institution’s deontology. So we specify the *contents* of institutional facts in general terms, and those general terms are the institutional roles involved. Now we can return to what is represented in our formalism.

Now we have to identify the different ways that roles can be held. 1) A role can be held by a group. Committees are like that. 2) One role can be held by many different agents. Think of the role of home owner. Although home owners will often own different homes, it is one and the same role held by many individuals. Also, multiple collective agents can fulfill the same role. Think of the role of basketball team in the NBA; that role is fulfilled by many different groups at the same time. 3) Certain roles are sometimes fulfilled and sometimes not. Think of the legal role of driver. Sometimes people are operating motor vehicles, sometimes not. When people are not driving, they are not drivers. Similarly, sometimes a police officer is off duty, sometimes on duty. 4) Some roles are held continuously. Even when Barack Obama is asleep he is still the president of The United States, and Michelle’s husband. This kind of role isn’t a permanent role, but it is more stable than the role of driver.

We will focus on representing the ways 1, 2, and 4 that roles can be held. We will leave 3 for future research. Ultimately we will interpret institutional roles by by assigning role terms to agent terms: see section 6.6. Representing 1 is straightforward; all that is needed is to allow roles to be interpreted as sets of individual agent terms. To represent 2, we allow the assignment of agent terms to roles to be a relation. So many agent terms can be assigned to a single role

term. To represent all that is needed is to allow the assignment of agent terms to role terms to persist over time. But there are a few worries to deal with before moving on.

We should make another brief remark on role and agent terms. Given two individual agent terms **a**, **b** each is supposed to be interpreted as representing *different* agents. This makes the formalism a bit simpler since we don't have to include an identity predicate. In some work, e.g., Sauro et al. (2006), the agents from the model are part of the language as well. Here we chose to use terms to stand in for agents rather than make the agents part of the language. Given our treatment of agents, we treat roles in the same way. When we use distinct role terms they represent different roles.

From our remarks in the previous paragraph, it should be clear that our formalism can only express things in the following manner: $[\mathbf{r} \text{ xstit}] \varphi \supset [\mathbf{r}' \text{ xstit}] \psi$. If this sentence is supposed to specify something that a citizen **r**, a kind of role, must do for *another* citizen **r'**, and there is only one role term for citizen, we can't use **r'** to specify another, distinct citizen. But reformulating it as $[\mathbf{r} \text{ xstit}] \varphi \supset [\mathbf{r} \text{ xstit}] \psi$ would only say something about one and the same agent that fulfills **r**.

To overcome this problem we will distinguish a *role term* from a *general role*. Each role term must be interpreted as a unique agent (collective or singular), but each role term can be conceived of as an instantiation of a *general role*, i.e., a role that can be held by many agents at a time. Although we can't express the sameness of the roles in the object language, we can overcome the problem by imposing conditions on the metalanguage and on the interpretation of the object language. The way to handle this barrier to interpretation is to partition **Rol** so that each cell of the partition represents a *General Role* (GR). The idea behind introducing this formal characterization via a partition of **Rol** is so that we can have many agents/groups holding the same role, but not being interpreted as the same members of **Rol**. Some cells in the partition may be unit sets. Those cells represent degenerate GRs, i.e., GRs that can't be held by many agents/groups at a time. How this will play out formally is discussed in section 6.6.

But we will not represent the assignment of agent terms to role terms in the object language.

There is no need to represent these assignments in the object language because the object language is used to represent the abstract version of the institution, and agent terms have not been assigned to the roles in the abstract version. We will, however, have a use for a representation of these assignments in the object language when we look at how an institution is imposed on a representation of the brute world.

Now that we have discussed institutional roles, we have completed expressing what we needed to be able to express about agents.

5.1.3 A Special Institutional Fact: The Violation Constant

Deontic logic was first developed in Ernst Mally (1926), and conceived rather differently later in von Wright (1951), but even later developments justified deontic logic largely by analogy with alethic modal logic. This analogy uses a language as follows:

$$\varphi := \mathbf{p} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \supset \varphi \mid \mathbf{O}\varphi$$

Where \mathbf{p} is an atomic sentence, and $\mathbf{O}\varphi$ is read as ‘it is obligatory that φ ’. So the language is interpreted as saying what ought to *be* rather than what someone ought to *do*. The orthodox deontic logic, called SDL for ‘standard deontic logic’, is given by a Hilbert style calculus with axioms $\mathbf{K} \mathbf{O}(\varphi \supset \psi) \supset (\mathbf{O}\varphi \supset \mathbf{O}\psi)$, and $\mathbf{D} \mathbf{O}\varphi \supset \neg\mathbf{O}\neg\varphi$ and closed under the rules of modus ponens and necessitation for \mathbf{O} , i.e., if $\vdash \varphi$, then $\vdash \mathbf{O}\varphi$. This logic is widely seen as inadequate for representing the intuitive notions of moral obligation, or other kinds of obligation, but we won’t be concerned with that at the moment.

SDL can be given an intuitive semantics in terms of possible worlds. $\mathbf{O}\varphi$ is true at a world w , i.e., φ is obligatory, iff φ is true at all deontically ideal worlds relative to w . This can be modelled mathematically in terms of the standard Kripke semantics⁵ for modal logic. SDL is sound and complete with respect to the class of Kripke models where the frame relation R is serial ($\forall x \exists y Rxy$). That means that every argument that is validated by the semantics is also validated by the proof theory and vice versa.

⁵For a reminder of Kripke semantics, i.e., Kripke models see appendix B.1.

Another interpretation of SDL can be derived from Anderson (1958). Anderson's idea is that if the rules of some code have been transgressed, i.e., something prohibited is the case, then there is a violation, or there is liability to sanction. Also, conversely, if there is liability to sanction or some violation of a code, then something prohibited must be the case. So Anderson suggested that deontic logic could be reduced to alethic modal logic, but within that logic a special atomic sentence is singled out as a *violation constant*. We will call this violation constant V .

Anderson's reduction is formalized as follows. The language is given by

$$\varphi := \mathbf{p} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \supset \varphi \mid \Box\varphi \mid V$$

Where $\Box\varphi$ is 'it is necessary that φ ' and V is 'there is a violation'. Although Anderson (1958) uses a more complex logic, a very simple logic can be given for this that is adequate to express the deontic relations of SDL. The logic is given by a Hilbert style calculus

$$\mathbf{K} \quad \Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi),$$

$$\mathbf{V} \quad \neg\Box V,$$

and is closed under the rules of modus ponens and necessitation for \Box , i.e., if $\vdash \varphi$, then $\vdash \Box\varphi$.

We can call this system $\mathbf{K}+\mathbf{V}$. The sentence $\mathbf{O}\varphi$ can be expressed intuitively in this language as 'if not φ , then there is a violation'. Formally that is expressed as $\Box(\neg\varphi \supset V)$. Using that logic and the translation t from SDL to Anderson's language given by:

- $t(\mathbf{p}) = \mathbf{p}$
- $t(\varphi * \psi) = t(\varphi) * t(\psi)$ for $*$ $\in \{ \wedge, \vee, \supset, \equiv \}$
- $t(\neg\varphi) = \neg t(\varphi)$, and
- $t(\mathbf{O}\varphi) = \Box(\neg t(\varphi) \supset V)$

It can be shown that $t(O(\varphi \supset \psi)) \supset (t(O\varphi) \supset t(O\psi))$, and $D\ t(O\varphi) \supset \neg t(O\neg\varphi)$, are both theorems of K+V.

Semantically, we can interpret the language on Kripke frames again. K+V is interpreted onto Kripke models $\langle W, R, v \rangle$ in the usual way for where for \Box , but they must satisfy the following condition. We need to require that the valuation v of the models is such that for all $x \in W$, $\{y \in W : Rxy\} \not\subseteq v(V)$. This means that in every state/world it is possible that there isn't a violation, i.e., there is a related state/world that is a non-violation state/world. It is easily seen that V is validated iff the model satisfies: for all $x \in W$, $\{y \in W : Rxy\} \not\subseteq v(V)$. Through a standard completeness proof we can see that K+V is complete with respect to the class of Kripke models that satisfy for all $x \in W$, $\{y \in W : Rxy\} \not\subseteq v(V)$.

Using Anderson's reduction we can represent other deontic concepts:

1. φ is permitted: $\Diamond(\varphi \wedge \neg V)$
2. φ is forbidden: $\Box(\varphi \supset V)$
3. Ought implies can: $O\varphi \supset \Diamond\varphi$ (Kant's thesis)

In some ways the ability of Anderson's reduction to express Kant's thesis makes the reduction a better expression of deontic concepts than pure SDL.

The violation constant V allows us to represent when a violation has occurred, and allows the classification of φ -states as violation states, i.e., V -states. The idea of interpreting Anderson's reduction as a way of classifying states as violation states we attribute to Grossi (2007). How to represent permission and obligation we leave until section 6.4. The introduction of V into language of xstit we allows us to represent all of the necessary deontic relationships for specifying institutional roles.

So now we have discussed all of the parts of Searle's general schema of status function declarations. The assignment of roles is what allows use to represent the relations. Power is captured by ability via xstit, but we will deal with that fully in section 6.4. We will show that we

can express all that needs to be expressed in the content of those status function declarations, i.e., deontology. Now we will offer a formal model of Vanderveken’s account of strong implication.

5.2 Illocutionary Entailment

Strong implication is the restriction of classical consequence to entailment between propositions where the content of the conclusion is contained in the content of the premise(s). We need to explain a bit more about Vanderveken’s formal semantics of propositions in order to explain strong implication. For this we will follow Vanderveken (1991) and Vanderveken and Nowak (1995): since the former has more to do with our project, but the latter focuses on a more specialized version of the logic suitable for our purposes.

It isn’t clear what propositions are, ontologically speaking. But we don’t want to take a view on the constituents of propositions to represent them in a way that captures our intuitions, and is extensionally adequate in that respect. We are following Vanderveken’s ideas, so we want to develop a way to talk about the propositions that results in them having an equivalent structure to that in Vanderveken’s work.

What consequence relation in a formal language strong implication corresponds to depends on what view of propositional containment is represented in that formal system. Vanderveken’s thought is that the propositional content of a proposition $\llbracket \varphi \rrbracket$ corresponds to the set of atomic propositions that are sub-propositions of $\llbracket \varphi \rrbracket$. In terms of formulas in a formal language, the content of φ is the set of atomic sentences that occur as sub-formulas of φ . But what the content of a proposition “actually is” is not something we want to take a view on. All that we need is a way to represent the content so that it meets up with Vanderveken’s basic idea.

We are on the way to representing strong implication, but we need to introduce a new formal language. Let the language \mathcal{L}_S be defined by the following grammar:

$$\varphi := \perp \mid \mathbf{p} \mid (\varphi \wedge \varphi) \mid \neg(\varphi) \mid \varphi \supset \varphi \mid (\varphi \vee \varphi) \mid (\varphi \equiv \varphi) \mid (A \Subset B)$$

Where A, B are pure Boolean formulas (from the language \mathcal{L}_P) and $\mathbf{p} \in \mathbf{At}$ the set of atomic

formulas. In what follows we will refer to pure Boolean formulas by $A, B, C, A_1, B_2 \dots$. A formula of the form $\varphi \subseteq \psi$ is interpreted as saying that the propositional content of φ is contained in the propositional content of ψ . The reason that we restrict \subseteq in the way that we do is because it is not clear what the propositional content of $A \subseteq B$ would be. Would it be contained in the propositional content of $A \wedge B$? Or just in A (or B)? The issue is one that we simply wish to avoid since the discussion would take us away from the point.

We interpret the formulas of \mathcal{L}_S relative to an interpretation \mathcal{I} which is a combination of an *SI-frame* and a valuation v .

Definition 5.2.1. $\langle \mathcal{D}, I \rangle$ is an *SI-frame* for \mathcal{L}_S where

1. I is a set of possible worlds such that $I \neq \emptyset$,
2. $\mathcal{D} = \langle D, \lesssim, \vee, \odot \rangle$ where $\langle D, \lesssim \rangle$ is a partial order⁶ with D non-empty, such that
 - (a) for all $d_1, d_2 \in D$, there is $d_1 \vee d_2 \in D$ that is a supremum of d_1 and d_2 . Explicitly, $d_1, d_2 \lesssim d_1 \vee d_2$, and for any $d_3 \in D$ such that $d_1 \lesssim d_3$ and $d_2 \lesssim d_3$, $d_1 \vee d_2 \lesssim d_3$.
 - (b) an element \odot in D that is a global infimum. Formally, for all $d \in D$, $\odot \lesssim d$,
 - (c) there is a (non-empty) set of atoms of \mathcal{D} , D_A which have the property that if $d \in D_A$, then $\forall d' \in D (d' \neq \odot \ \& \ d' \lesssim d \text{ only if } d = d')$ (These are the members of D that are almost at the bottom), and
 - (d) the members of D_A also satisfy the following property: for $d \in D_A$ and any $d_1, d_2 \in D$, if $d \lesssim d_1 \vee d_2$, then either $d \lesssim d_1$ or $d \lesssim d_2$.

Definition 5.2.2. An *SI-model* $\mathcal{I} = \langle \mathcal{D}, I, v \rangle$ is an *SI-frame* together with a valuation $v : \mathbf{At} \rightarrow D_A \times \mathcal{P}(I)$. As a stylistic variant we will also call $\mathcal{I} = \langle \mathcal{D}, I, v \rangle$ an *SI-interpretation*.

For these interpretations \mathcal{I} we get a semantic value for each A in \mathcal{L}_S denoted by $\llbracket A \rrbracket_{\mathcal{I}}$ which is a member of $D \times \mathcal{P}(I)$. We will usually omit the subscript \mathcal{I} when there is no ambiguity.

⁶A partial order is one where the order relation is reflexive, transitive and antisymmetric.

These elements D and the subsets of I represent propositional contents and what we call *informational content*, respectively. The subsets of I are the set of worlds where the sentences are true which is the informational content of a sentence (that is how informational content is modelled in Dretske (1981)). In $\llbracket A \rrbracket$ will refer to the first coordinate as $\llbracket A \rrbracket_1$ and the second coordinate $\llbracket A \rrbracket_2$. Now we can extend v to $\llbracket \cdot \rrbracket$ as follows: Let $A, B \in \mathcal{L}_S$, and for $X \subseteq I$, $X^c = I \setminus X$,

1. for $\mathbf{p} \in \mathbf{At}$, $\llbracket \mathbf{p} \rrbracket = v(\mathbf{p})$
2. $\llbracket \perp \rrbracket = \langle \odot, \emptyset \rangle$
3. $\llbracket \top \rrbracket = \langle \odot, I \rangle$
4. $\llbracket \neg A \rrbracket = \langle \llbracket A \rrbracket_1, I \setminus \llbracket A \rrbracket_2 \rangle$
5. $\llbracket A \wedge B \rrbracket = \langle \llbracket A \rrbracket_1 \vee \llbracket B \rrbracket_1, \llbracket A \rrbracket_2 \cap \llbracket B \rrbracket_2 \rangle$
6. $\llbracket A \vee B \rrbracket = \langle \llbracket A \rrbracket_1 \vee \llbracket B \rrbracket_1, \llbracket A \rrbracket_2 \cup \llbracket B \rrbracket_2 \rangle$
7. $\llbracket A \supset B \rrbracket = \langle \llbracket A \rrbracket_1 \vee \llbracket B \rrbracket_1, \llbracket A \rrbracket_2^c \cup \llbracket B \rrbracket_2 \rangle$
8. $\llbracket A \equiv B \rrbracket = \langle \llbracket A \rrbracket_1 \vee \llbracket B \rrbracket_1, (\llbracket A \rrbracket_2^c \cap \llbracket B \rrbracket_2^c) \cup (\llbracket A \rrbracket_2 \cap \llbracket B \rrbracket_2) \rangle$

In the definitions above we can see that only atomic propositional contents, i.e., members of D_A are assigned to atomic sentences, and negation and conjunction are defined as one might think: the propositional content is cumulative and the other conditions correspond to the set-theoretic analogs of the respective logical operations. \perp and \top have null propositional content. But notice that $\varphi \wedge \neg\varphi$ will have the propositional content $\llbracket \varphi \rrbracket_1$, so $\llbracket \perp \rrbracket_1 \neq \llbracket \varphi \wedge \neg\varphi \rrbracket_1$.

Truth at a possible world i is defined recursively for all formulas φ, ψ of \mathcal{L}_S as follows:

1. for $\mathbf{p} \in \mathbf{At}$, $\mathcal{I}, i \models \mathbf{p}$ iff $i \in \llbracket \mathbf{p} \rrbracket_2$
2. $\mathcal{I}, i \not\models \perp$
3. $\mathcal{I}, i \models \top$

4. $\mathcal{I}, i \Vdash \neg\psi$ iff $\mathcal{I}, i \nVdash \psi$.
5. $\mathcal{I}, i \Vdash \varphi \wedge \psi$ iff $\mathcal{I}, i \Vdash \varphi$ and $\mathcal{I}, i \Vdash \psi$
6. $\mathcal{I}, i \Vdash \varphi \vee \psi$ iff $\mathcal{I}, i \Vdash \varphi$ or $\mathcal{I}, i \Vdash \psi$
7. $\mathcal{I}, i \Vdash \varphi \supset \psi$ iff $\mathcal{I}, i \nVdash \varphi$ or $\mathcal{I}, i \Vdash \psi$
8. $\mathcal{I}, i \Vdash \varphi \equiv \psi$ iff $(\mathcal{I}, i \Vdash \varphi$ only if $\mathcal{I}, i \Vdash \psi)$ and $(\mathcal{I}, i \Vdash \varphi$ if, $\mathcal{I}, i \Vdash \psi)$
9. $\mathcal{I}, i \Vdash A \in B$ iff $\llbracket A \rrbracket_1 \lesssim \llbracket B \rrbracket_1$

Recall that A and B are *pure Boolean* formulas. We can then define semantic *SI*-entailment \models_{SI} between a set of sentences Γ and a sentence φ as follows: $\Gamma \models_{SI} \varphi$ iff for all *SI*-models $\mathcal{I} = \langle \mathcal{D}, I, v \rangle$, and $i \in I$, if $\mathcal{I}, i \Vdash \gamma$ for all $\gamma \in \Gamma$, then $\mathcal{I}, i \Vdash \varphi$. When Γ is empty we write $\models_{SI} \varphi$ and say φ is a logical *SI*-truth. Moving from the semantics to the proof theory we give the following axiomatization which is amended from Vanderveken and Nowak (1995, p. 397–8).

Definition 5.2.3. Axioms for the Logic *SI*:

Axioms for Classical Propositional Logic (*CL*)

- CL1 $\varphi \supset (\psi \supset \varphi)$
- CL2 $(\varphi \supset (\psi \supset \theta)) \supset ((\varphi \supset \psi) \supset (\psi \supset \theta))$
- CL3 $(\varphi \wedge \psi) \supset \psi$
- CL4 $(\varphi \wedge \psi) \supset \varphi$
- CL5 $(\varphi \supset \psi) \supset ((\varphi \supset \theta) \supset (\varphi \supset \psi \wedge \theta))$
- CL6 $\varphi \supset (\varphi \vee \psi)$
- CL7 $\psi \supset (\varphi \vee \psi)$
- CL8 $(\varphi \supset \psi) \supset ((\theta \supset \psi) \supset (\varphi \vee \theta \supset \psi))$

$$\text{CL9 } (\psi \supset \neg\varphi) \supset (\varphi \supset \neg\psi)$$

$$\text{CL10 } \neg(\psi \supset \psi) \supset \varphi$$

$$\text{CL11 } \varphi \vee \neg\varphi$$

$$\text{CL12 } (\varphi \wedge \neg\varphi) \supset \perp$$

Axioms for Propositional Containment

$$\text{PC1 } A \in A$$

$$\text{PC2 } (B \in A) \supset ((C \in B) \supset (C \in A))$$

$$\text{PC3 } (\mathbf{p}_i \in \mathbf{p}_j) \supset (\mathbf{p}_j \in \mathbf{p}_i)$$

$$\text{PC4 } A \in (A \wedge B)$$

$$\text{PC5 } B \in (A \wedge B)$$

$$\text{PC6 } (B \in A) \supset ((C \in A) \supset ((C \wedge B) \in A))$$

$$\text{PC7 } A \in \neg A$$

$$\text{PC8 } \neg A \in A$$

$$\text{PC9 } (\mathbf{p}_i \in (A \wedge B)) \supset ((\mathbf{p}_i \in A) \vee (\mathbf{p}_i \in B))$$

$$\text{PC10 } \perp \in A$$

$$\text{PC11 } A \in (A \vee B)$$

$$\text{PC12 } B \in (A \vee B)$$

$$\text{PC12A } (A \vee B) \in (A \wedge B)$$

$$\text{PC13 } A \in (A \supset B)$$

$$\text{PC14 } B \in (A \supset B)$$

$$\text{PC14A } (A \supset B) \in (A \wedge B)$$

$$\text{PC15 } A \in (A \equiv B)$$

$$\text{PC16 } B \in (A \equiv B)$$

PC16A $(A \equiv B) \in (A \wedge B)$

PC17 $\top \in A$

Rules

MP If $\vdash_{SI} \varphi \supset \psi$ and $\vdash_{SI} \varphi$, then $\vdash_{SI} \psi$

An *SI*-proof is a sequence of \mathcal{L}_S sentences $\varphi_1, \dots, \varphi_n$ such that for each $1 \leq i \leq n$, either φ_i is an axiom, or there are sentences φ_j and φ_k with $j, k < i$ such that:

1. $\varphi_j = \varphi_k \supset \varphi_i$ and so we can use MP to get φ_i , or

A set of sentences Γ *SI*-proves a conclusion φ iff there are $\{\gamma_1, \dots, \gamma_n\} \subseteq \Gamma$ and $\vdash_{SI} \gamma_1 \wedge \dots \wedge \gamma_n \supset \varphi$. An *SI*-consistent set Γ is a set of sentences such that $\Gamma \not\vdash_{SI} \perp$. We will omit the subscript *SI* from the \vdash when no confusion will occur, and ‘sentence’ will mean *SI*-sentence unless otherwise specified. In the next section we provide a proof of completeness of \vdash_{SI} with respect to the semantics. Now we can formally define strong implication.

This logic provides us with a complete characterization of how \in acts and interacts with the Boolean connectives. However, $\models_{SI} / \vdash_{SI}$ does not, by itself, give a formal account of the notion of strong implication. But we can use $\models_{SI} / \vdash_{SI}$ to account for strong implication. There are two ways that we can use the logic of *SI* to define strong implication. The intuition underlying strong implication is that for a set of sentences Γ to strongly imply φ , $\Gamma \vdash_{CL} \varphi$ and the content of φ must be contained in the content of Γ . We can interpret the connection between the contents of φ and Γ either relative to a single interpretation \mathcal{I} , or relative to every interpretation. We choose the following definition for its generality.

Definition 5.2.4 (Strong Implication). A set of sentences Γ strongly implies φ ($\Gamma \vdash_S \varphi$) iff there are sentences $\gamma_1, \dots, \gamma_n$ in Γ such that $\vdash_{CL} \gamma_1 \wedge \dots \wedge \gamma_n \supset \varphi$, and there are $\gamma'_1, \dots, \gamma'_m$ in Γ such that $\vdash_{SI} \gamma'_1 \wedge \dots \wedge \gamma'_m \supset (\varphi \in \gamma_1 \wedge \dots \wedge \gamma_n)$

The set of sentences Γ is just a general set of *SI*-sentences. So it may contain propositional content sentences of the form $A \in B$ which are not theorems of *SI*. This means that Γ may

contain auxiliary facts about propositional containment that can be used in making inferences from Γ about what follows. This is in contrast to ignoring these auxiliary facts about content contained in Γ . If we ignored the extra facts about containment, then we would say that $\Gamma \vdash_S \varphi$ iff $\Gamma \vdash_{CL}$ and $\vdash_{SI} \varphi \in \bigwedge_{i=1}^n \gamma_i$ for some $\gamma_1, \dots, \gamma_n \in \Gamma$.

But if Γ consists of pure Boolean formulas, then there are no auxiliary facts about containment to consult. Therefore, the general definition is reduced to the more restrictive definition of strong implication. In the next section we provide a completeness proof for \vdash_{SI} relative to \models_{SI} .

5.3 Completeness of \vdash_{SI}

First we show that the axioms from definition 5.2.3 are sound for the class of models we have defined.

Proposition 5.3.1. *The axioms are sound for any SI-model \mathcal{I} .*

Proof. The CL axioms are those for classical logic and the semantics is defined classically so we will do the proof for the PC axioms only. If in any model $\mathcal{I} = \langle \mathcal{D}, I, v \rangle$ we have that \mathcal{I} assigns to each atom \mathbf{p} a member of D , and any pure Boolean formula A will also be assigned a member of D according to the rules above. For any A , $\llbracket A \rrbracket_1 \lesssim \llbracket A \rrbracket_1$ by the conditions on \mathcal{D} , and so for any $i \in I$, $\mathcal{I}, i \Vdash A \in A$. We can use the conditions on \lesssim to handle PC2 and conditions on \vee to handle PC4-6. The **Not** condition above gives PC7 and PC8. PC3 and PC9 are the only interesting cases. For PC3 assume that $\mathcal{I}, i \Vdash \mathbf{p}_i \in \mathbf{p}_j$. All atomic sentences are assigned atoms of D , and $\llbracket \mathbf{p}_i \rrbracket_1 \lesssim \llbracket \mathbf{p}_j \rrbracket_1$ implies that $\llbracket \mathbf{p}_j \rrbracket_1 \lesssim \llbracket \mathbf{p}_i \rrbracket_1$ by definition 5.2.1–2(c). Thus, $\mathcal{I}, i \Vdash \mathbf{p}_j \in \mathbf{p}_i$ and so $\mathcal{I}, i \Vdash (\mathbf{p}_i \in \mathbf{p}_j) \supset (\mathbf{p}_j \in \mathbf{p}_i)$. For PC9 we note that the condition that all members $d \in D_A$ must be such that if $d \lesssim d_1 \vee d_2$, then either $d \lesssim d_1$ or $d \lesssim d_2$ makes the condition go through. \square

A set of sentences Γ is an *SI-theory* if it is closed under \vdash_{SI} , i.e., if $\Gamma \vdash_{SI} \varphi$, then $\varphi \in \Gamma$. When no confusion will result we will call *SI-theories*, simply ‘theories’. Maximally consistent

theories are theories Γ such that if $\varphi \notin \Gamma$, $\Gamma; \varphi \vdash \perp$. From here we will prove the completeness of the axioms in definition 5.2.3 for the semantics of SI above by using a modified Lindenbaum construction coupled with the canonical model constructions of Scott, Lemmon and Makinson. In what follows we take \vdash to mean \vdash_{SI} .

Proposition 5.3.2. *Each SI -consistent set Γ can be extended to a maximally SI -consistent set Γ^+ .*

Proof. Let Γ be a consistent set. Let $\{\varphi_n : n \in \mathbb{N}\}$ be an enumeration of all the formulas of \mathcal{L}_S . Define a sequence of sets as follows:

$$\Sigma_0 = \Gamma$$

$$\Sigma_n = \begin{cases} \Sigma_{n-1} \cup \{\varphi_{n-1}\} & \text{if } \Sigma_{n-1}; \varphi_{n-1} \not\vdash \perp. \\ \Sigma_{n-1} & \text{otherwise} \end{cases}$$

Then we define $\Gamma^+ = \bigcup_{n \in \mathbb{N}} \Sigma_n$. Γ^+ is consistent since, if $\Gamma^+ \vdash \perp$, there would be a finite $\Gamma_0^+ \subseteq \Gamma^+$ such that $\Gamma_0^+ \vdash \perp$. But $\Gamma_0^+ \subseteq \Sigma_k$ for some k , and hence $\Sigma_k \vdash \perp$. But by definition each Σ_k is consistent.

To show that Γ^+ is maximal suppose that $\varphi \notin \Gamma^+$. $\varphi = \varphi_k$ for some $k \in \mathbb{N}$, so at stage $k + 1$ φ_k was considered, but φ_k wasn't added. Thus, at stage $k + 1$, $\Sigma_k; \varphi_k \vdash \perp$. Since $\Sigma_k \subseteq \Gamma^+$, $\Gamma^+; \varphi_k \vdash \perp$. Hence no extension of Γ^+ is consistent. \square

So we can extend any consistent set to a maximally consistent set, it is also trivial to show that maximally consistent sets are theories. For each maximal set there is a canonical model, and many maximal sets will have the same canonical model. But the canonical model *depends* on which maximal set is chosen as a starting point. Let Δ be a maximal set, then let $I_\Delta^* = \{\Delta' : A, B \in \mathcal{L}_P, A \subseteq B \in \Delta \text{ iff } A \subseteq B \in \Delta'\}$. Also, we define an equivalence relation between sentences of \mathcal{L}_P relative to Δ , \sim_{SI}^Δ , as follows: let $A, B \in \mathcal{L}_P$

$$A \sim_{SI}^\Delta B \text{ iff } (A \subseteq B) \wedge (B \subseteq A) \in \Delta \quad (5.1)$$

We then define $[A] = \{ B \in \mathcal{L}_P : A \sim_{SI}^\Delta B \}$ for $A \in \mathcal{L}_P$. Clearly \sim_{SI}^Δ is an equivalence relation from axioms PC1, PC2, and because both $A \subseteq B$ and $B \subseteq A$ are in Δ . Also, if $\Delta' \in I_\Delta^*$, then $\sim_{SI}^\Delta = \sim_{SI}^{\Delta'}$ since all of the sets in I_Δ^* agree with Δ on the \subseteq -sentences. Before moving on, we have to notice something rather important:

Observation 5.3.3. *For all $A \in \mathcal{L}_P$, $[A] = [\wedge_i \mathbf{p}_i]$ such that $\wedge_i \mathbf{p}_i$ is the conjunction of all the atoms which are subformulas of A .*

Proof. The proof is by induction on the complexity of A . The basis case is trivial. Now by axioms PC2, and PC4,5,6 we can show that if $[A] = [A_1]$ and $[B] = [B_1]$, then $[A \wedge B] = [A_1 \wedge B_1]$. Now by axioms PC11–PC16A, $[A \vee B] = [A \supset B] = [A \equiv B] = [A \wedge B]$. So if $[A] = [\wedge \mathbf{p}_i]$ such that \mathbf{p}_i is a subformula of A , and $[B] = [\wedge \mathbf{q}_j]$ such that \mathbf{q}_j is a subformula of B by the induction hypothesis, then $[(\wedge \mathbf{p}_i) \wedge (\wedge \mathbf{q}_j)] = [A \wedge B] = [A * B]$ for $*$ $\in \{ \supset, \vee, \equiv \}$. But the \mathbf{p}_i and \mathbf{q}_j are the subformulas of A and B , respectively. That completes the induction. \square

Now we can define the canonical model for Δ .

Definition 5.3.1. Let Δ be a maximally SI-consistent set. Define the canonical model for Δ as $\mathcal{I}_\Delta^* = \langle \mathcal{D}_\Delta^*, I_\Delta^*, v_\Delta^* \rangle$ as follows:

1. I_Δ^* is as we have defined it above.
2. $\mathcal{D}_\Delta^* = \langle D^*, \lesssim^*, \Upsilon^*, \odot^* \rangle$ where $D^* = \mathcal{L}_P / \sim_{SI}^\Delta$, and for all $A, B \in \mathcal{L}_P$ $[A] \lesssim^* [B]$ iff $A \subseteq B \in \Delta$ or $[A] = [B]$, and $[A] \Upsilon^* [B] = [A \wedge B]$. Finally, $\odot^* = [\perp]$.
3. $v_\Delta^*(\mathbf{p}) = \langle [\mathbf{p}], \{ \Delta \in I^* : \mathbf{p} \in \Delta \} \rangle$

Now we have to ensure that \mathcal{I}_Δ^* is a model for SI . For this we need to note the following lemma.

Lemma 5.3.4. *For $A \in \mathcal{L}_P$, $\llbracket A \rrbracket_1 = [A]$.*

Proof. The proof is by induction on the complexity of A . The basis case holds by definition. Assume for all C of less complexity than D , $[D] = \llbracket D \rrbracket_1$. By definition $[A \wedge B] = [A] \vee^* [B]$, so by IH $\llbracket A \rrbracket_1 \vee \llbracket B \rrbracket_1 = \llbracket A \wedge B \rrbracket_1$, by definition.

By axioms PC11–PC16A $[A \vee B] = [A \supset B] = [A \equiv B] = [A \wedge B]$. Thus, $[A * B] = [A] \vee^* [B]$ for $*$ $\in \{\supset, \vee, \equiv\}$. So by IH, $[A] \vee^* [B] = \llbracket A \rrbracket_1 \vee^* \llbracket B \rrbracket_1 = \llbracket A * B \rrbracket_1$.

Finally, $[\neg A] = [A]$ by axioms PC7 and PC8, so by IH, $[A] = \llbracket A \rrbracket_1$, and $\llbracket A \rrbracket_1 = \llbracket \neg A \rrbracket_1$ by definition. That completes the induction. \square

Now we can see that \mathcal{I}_Δ^* is an SI -model. Clearly, $I_\Delta^* \neq \emptyset$ since $\Delta \in I_\Delta^*$. $\langle D^*, \lesssim^* \rangle$ is a partial order. The transitive and reflexive axioms for \subseteq also make \lesssim^* transitive and reflexive. It is antisymmetric since if $A \subseteq B$ and $B \subseteq A$ are both in all of the sets in I_Δ^* , $[A] = [B]$. Inspection of the PC axioms guarantees the other restrictions on \mathcal{D} . If $[A], [B] \in D^*$, then $[A \wedge B] \in D^*$ by definition, and by PC4,5 $[A], [B] \lesssim^* [A \wedge B]$. Also, by PC6 $[A \wedge B]$ is a supremum of $[A]$ and $[B]$. Finally, by PC10 $[\perp]$ is a global infimum. Each $[\mathbf{p}]$ is an atom since, first, if $[B] \neq [\perp]$, then suppose that $[B] \lesssim^* [\mathbf{p}]$. That means $B \subseteq \mathbf{p} \in \Delta$. From the observation above we know that $[B] = [\wedge \mathbf{p}_i]$ for all of the atoms in B , but then $[\wedge \mathbf{p}_i] \lesssim^* [\mathbf{p}]$. So $\wedge \mathbf{p}_i \subseteq \mathbf{p} \in \Delta$, and so by PC2,4 and 5, $\mathbf{p}_i \subseteq \mathbf{p}$ for each \mathbf{p}_i from B . But then we know that $\mathbf{p} \subseteq \mathbf{p}_i \in \Delta$ by axiom PC3. Thus $[\mathbf{p}_i] = [\mathbf{p}]$ for each \mathbf{p}_i in B . Thus, $[\mathbf{p}] = [\wedge \mathbf{p}_i]$ by PC2,4 and 6 and the observation above. Yet $[\wedge \mathbf{p}_i] = [B]$, so $[\mathbf{p}] = [B]$. Next we check definition 5.2.1 2(d). If $[\mathbf{p}] \lesssim^* [A] \vee^* [B]$, then since $[A] \vee^* [B] = [A \wedge B]$, $\mathbf{p} \subseteq A \wedge B \in \Delta$. But then by PC9 $(\mathbf{p} \subseteq A) \vee (\mathbf{p} \subseteq B) \in \Delta$. Since Δ is maximal, either $\mathbf{p} \subseteq A \in \Delta$ or $\mathbf{p} \subseteq B \in \Delta$, so either $[\mathbf{p}] \lesssim^* [A]$ or $[\mathbf{p}] \lesssim^* [B]$. Finally, by our definition of v_Δ^* , it is a function from \mathbf{At} to $D^* \times \mathcal{P}(I_\Delta^*)$. Thus, \mathcal{I}^* is an SI -model. Now we can show the fundamental theorem for SI . In the next theorem we take the maximal set to be implicit.

Proposition 5.3.5. *For each canonical model \mathcal{I}^* and formula φ , and all $\Delta \in I^*$,*

$$\mathcal{I}^*, \Delta \Vdash \varphi \iff \varphi \in \Delta$$

Proof. The proof is by induction on the complexity of φ . For the atomic case, \mathbf{p} , it is handled

by the definition of \mathcal{I} . The induction hypothesis is that for all ψ of lower complexity than φ , and all $\Delta \in I^*$

$$\mathcal{I}^*, \Delta \Vdash \psi \iff \psi \in \Delta$$

The Boolean cases are standard so we will omit them. If $\mathcal{I}^*, \Delta \Vdash A \subseteq B$, then $\llbracket A \rrbracket_1 \lesssim^* \llbracket B \rrbracket_1$. By the definition of the canonical model we have $\llbracket A \rrbracket \lesssim^* \llbracket B \rrbracket$, and that occurs only if $A \subseteq B \in \Delta'$ for all $\Delta' \in I^*$. So, *a fortiori* $A \subseteq B \in \Delta$. In the other direction if $A \subseteq B \in \Delta$, then it must be a member of all $\Delta' \in I^*$ by the definition of I^* . Thus, $\llbracket A \rrbracket \lesssim^* \llbracket B \rrbracket$, so by the previous lemma $\llbracket A \rrbracket_1 \lesssim^* \llbracket B \rrbracket_1$; therefore, $\mathcal{I}^*, \Delta \Vdash A \subseteq B$. \square

Given a consistent set of sentences Γ , we can then extend this set to a Γ^+ . We can then take the set of all of the maximally consistent sets of sentences that agree with Γ^+ on all sentences of the form $A \subseteq B$, we call this set I_Γ^* as before. Note that there could be many canonical models for a set Γ since each will depend on the maximal extension constructed.

Proposition 5.3.6. *If $\Gamma \models_{SI} \varphi$, then $\Gamma \vdash_{SI} \varphi$.*

Proof. Suppose that $\Gamma \not\vdash \varphi$, then $\Gamma; \neg\varphi$ is consistent, and so can be extended to maximally consistent set $(\Gamma; \varphi)^+$. Then we construct a canonical model around $(\Gamma; \varphi)^+$ using $I_{(\Gamma; \varphi)^+}$. Since $(\Gamma; \varphi)^+ \in I_{(\Gamma; \varphi)^+}$, by proposition 5.3.5, we will have $\mathcal{I}_{(\Gamma; \varphi)^+}^*, (\Gamma; \varphi)^+ \not\models \varphi$, yet $\mathcal{I}^*, (\Gamma; \varphi)^+ \Vdash \gamma$ for all $\gamma \in \Gamma$ —since $\Gamma \subseteq (\Gamma; \varphi)^+$. But then $\Gamma \not\models_{SI} \varphi$. \square

The proceeding proposition and proposition 5.3.1, provide the completeness proof that we are after.

Chapter 6

Formalization of Normative Entailment

...a consequence of this analysis is that society has a logical structure. ...Theories about ...parts of nature have logical structures but not the nature itself. But society consists in part of representations and those representations have logical structures. Any adequate theory about such phenomena must contain a logical analysis of their structures.

Searle (2005, p. 22)

6.1 Introduction

In chapter 5, we discussed various pieces necessary to represent institutions. In this chapter we will put those pieces together. We will extend each of the languages of the previous chapter to do so. As the quote from Searle above says, we have to provide a way to represent the logical structure of institutions. Let's recall Searle's general characterization of institutional facts.

We (or I) make it the case by Declaration that a Y status function exists in C and in so doing we (or I) create a relation R between Y and a certain person or persons, S, such that in virtue of SRY, S has the power to perform acts (of type) A. (Searle, 2010, pp. 101–2)

In Searle's characterization there are two parts. There is the declaration that is made to generate the institutional fact, i.e., the status function, and there is the characterization of that status function in terms of its powers, i.e., performing acts of type A. In that sense we really have two levels of reality—for lack of a better term: an institutional reality which is constituted by the powers specific to the institution/status function, and the brute reality on to which those powers are imposed. Since institutions have logical structure, as Searle's quote above indicates, this distinction between an institutional reality and brute reality should be mirrored in the logic as well.

To a certain extent, the status function is independent from its being declared. The institutional facts, which we have been calling *norms*, are represented independently from their imposition. A status function declaration will be the imposition of institutional facts on to a brute reality. We call this imposition an *implementation*.

After we have explained the distinctions between the languages, we will characterize the relation of norm consequence for this language. We call this relation \vdash_N . Norm consequence is strong implication applied to this new language. So although we have the basic relation of strong implication, we have to extend it to this more complex language.

In order to represent the various ‘actions of type A’ we have to show that we can represent, in the new language, the relevant relations. The basic sets of relations that are necessary to accomplish this come from Holfeld (1920). We will show that our language is sufficient to this task after in the penultimate section of this chapter.

Just as a new piece of notation, we will use Ω to refer to institutions. Institutions, on Searle’s view, can be identified with the set of institutional facts. So ‘ Ω ’ will refer to a set of formulas that represents the institutional facts of an institution. However, in mathematical logic there is already a notion called *institution*, but we do not mean that. As a stylistic variation we will also refer to Ω as a *code*.

6.2 Formal Languages for Institutions

As we said there are two realities and so to represent that we will need two languages. But there are a number of fragments of these languages that will play important roles in our discussion. So we start by extending $\mathcal{L}_{\text{xstit}}$ as follows:

We construct the language \mathcal{L} as follows: Let $[*] \in \{ \Box, P, X, [\mathbf{A} \text{ xstit}] \}$, and $\mathbf{p} \in \mathbf{At}_I \cup \mathbf{At}_B$

$$\varphi := \perp \mid \mathbf{p} \mid V \mid (\varphi \wedge \varphi) \mid \neg(\varphi) \mid (\varphi \supset \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \equiv \varphi) \mid [*]\varphi$$

where \mathbf{Ag} is a finite set of singular agent terms and $\mathbf{A} \subseteq \mathbf{Ag}$. We will call $\mathbf{A} \subseteq \mathbf{Ag}$ *agent terms* even though they may be *plural* agents. Note that we have added an operator P to this language.

It is the backward looking counterpart to the X operator. $P\varphi$ is read as ‘in the previous state φ ’. Also notice the atoms come from either \mathbf{At}_I or \mathbf{At}_B . Thus they are either institutional or brute. The language \mathcal{L} describes both the brute world and the institutional facts. We also have the violation constant V . It is treated as an atomic sentence, but is assigned a special meaning.

The language \mathcal{L}^B is the “brute fragment” of this language. We construct the language \mathcal{L}^B as follows: Let $[*] \in \{\square, P, X, [\mathbf{A} \text{ xstit}]\}$, and $\mathbf{p} \in \mathbf{At}_B$

$$\varphi := \perp \mid \mathbf{p} \mid (\varphi \wedge \varphi) \mid \neg(\varphi) \mid (\varphi \supset \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \equiv \varphi) \mid [*]\varphi$$

where $\mathbf{A} \subseteq \mathbf{Ag}$. So \mathcal{L}^B expresses all of the brute facts. Now we have to have an institutional counterpart to \mathcal{L}^B .

Recall that status functions are formulated as relations between institutional roles. However, some of the brute language is usually necessary in formulating institutional facts. Consider a sign that says ‘don’t walk on the grass’. That sign says that certain brute objects and actions, i.e., grass and walking, are prohibited in that area. So to represent the status functions, we will require a language to contain the institutional atoms, the brute atoms, and the institutional roles. But since norms are formulated in general ways, as we discussed in section 5.1, we don’t include agent terms, i.e., $\mathbf{A} \subseteq \mathbf{Ag}$, only role terms $\mathbf{R} \subseteq \mathbf{Rol}$.

We construct the language \mathcal{L}^I as follows: Let $[*] \in \{\square, P, X, [\mathbf{R} \text{ xstit}]\}$, and $\mathbf{p} \in \mathbf{At}_I \cup \mathbf{At}_B$

$$\varphi := \perp \mid \mathbf{p} \mid V \mid (\varphi \wedge \varphi) \mid \neg(\varphi) \mid (\varphi \supset \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \equiv \varphi) \mid [*]\varphi$$

where $\mathbf{R} \subseteq \mathbf{Rol}$. So \mathcal{L}^I is just like \mathcal{L} with the exception that there are role terms in place of agent terms.

Sometimes we will need to make reference to only *some* institutional facts. To do that we will define a function $at(\cdot)$ which assigns each formula φ the set $at(\varphi)$ of atomic formulas in φ . This generalizes to sets of formulas in the standard way $at(\Gamma) = \bigcup \{at(\varphi) : \varphi \in \Gamma\}$. We will look at the restricted language \mathcal{L}_Ω^B which is defined as follows.

$$\varphi := \perp \mid \mathbf{p} \mid (\varphi \wedge \varphi) \mid \neg(\varphi) \mid (\varphi \supset \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \equiv \varphi) \mid [*]\varphi$$

where $\mathbf{A} \subseteq \mathbf{Ag}$ and $\mathbf{p} \in at(\Omega)$. This language contains all of the institutional facts that can be expressed using the atomic institutional facts in Ω , but not all of the institutional facts in \mathbf{At}_I .

Now that we have the languages which express actions and institutional facts, we will add to the language \mathcal{L}^I the operator that represents propositional containment \Subset . This generates the language \mathcal{L}_{\Subset}^I . Let $[*] \in \{\Box, P, X, [\mathbf{R} \text{ xstit}]\}$, and $\mathbf{p} \in \mathbf{At}_I \cup \mathbf{At}_B$

$$\varphi := \perp \mid \mathbf{p} \mid V \mid (\varphi \wedge \varphi) \mid \neg(\varphi) \mid (\varphi \supset \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \equiv \varphi) \mid [*]\theta \mid A \Subset B$$

where $\mathbf{R} \subseteq \mathbf{Rol}$, $A, B \in \mathcal{P}(\mathbf{Rol}) \cup \mathcal{L}^I$, and $\theta \in \mathcal{L}^I$. This means this new language can contain formulas like $\{\mathbf{r}\} \Subset [\mathbf{R} \text{ xstit}](\neg p \wedge V)$, and $(\mathbf{p} \wedge \mathbf{q} \Subset [\mathbf{R} \text{ xstit}](\mathbf{p})) \supset (\mathbf{p} \vee [\mathbf{R}' \text{ xstit}]\mathbf{q})$. However, the language *will not* contain sentences like $[\mathbf{R} \text{ xstit}](\mathbf{p} \wedge V) \Subset (\mathbf{p} \Subset \mathbf{q})$, or $[\mathbf{R} \text{ xstit}](\mathbf{p} \Subset \mathbf{r})$. I.e., the operator \Subset cannot be iterated, nor can the operators $\{\Box, P, X, [\mathbf{R} \text{ xstit}]\}$ take \Subset -formulas as their complements.

There are two important restrictions that we must impose. First, as we have mentioned already, \mathbf{Rol} and \mathbf{Ag} are required to be finite. Second, \mathbf{At}_I is finite. This means that the languages we have defined should really be displayed as $\mathcal{L}^{I^{m,k,n}}$ where $|\mathbf{At}_I| = m$, $|\mathbf{Rol}| = n$ and $|\mathbf{Ag}| = k$. The reason for these restrictions have to do with the characterization of norm consequence. To fully characterize norm consequence we need to make sure that the purely institutional language is not only recursively specifiable, but that we can decide whether something is purely institutional in a finite number of steps. The reason for this will be made clear later. These specifications don't affect the logic per se since we will specify the axioms for it schematically, but it is important to be transparent about these matters.

In the next section we will provide a semantics and logic for \mathcal{L}_{\Subset}^I . In section 6.5 we provide a characterization of norm consequence. In the final two sections we formalize the notion of an implementation that we mentioned informally in the introduction to this chapter, then we show that we can represent Holfeld's legal relations.

6.3 Logic of $\mathcal{L}_{\subseteq}^I$

The way that we approach institutional entailment is by constructing something similar to what is called elsewhere¹ the parametrization of the logic of strong entailment by the logic of xstit. In a parametrization one logic L is given, and another logic L' is also given, but in the axiomatization and construction of the language formulas of L' can replace atoms of the logic L . Next we give the semantics for this language $\mathcal{L}_{\subseteq}^I$, leaving the m, k, n implicit.

6.3.1 Semantics for $\mathcal{L}_{\subseteq}^I$

We begin by defining an \mathcal{L}^I -frame. In the next chapter we define frames for \mathcal{L} alone, and show that they match up with the frames and models from definition 5.1.1.

Definition 6.3.1. An \mathcal{L}^I -frame \mathfrak{F} is a pair $\langle \mathcal{D}, \mathfrak{F}_x \rangle$ consisting of

1. \mathcal{D} as in definition 5.2.1, and
2. \mathfrak{F}_x is an xstit frame $\langle S, H, E \rangle$ according to definition 5.1.1 where E satisfies conditions a)–f).²

We can then define models

Definition 6.3.2. A model for \mathcal{L}^I is a frame \mathfrak{F} with a valuation $v : \mathbf{At}_I \cup \mathbf{At}_B \cup \{V\} \cup \mathbf{Rol} \rightarrow (D_A \cup (D_A \times \mathcal{P}(S)))$ such that for each atomic sentence \mathbf{s} ,

1. $v(\mathbf{s}) \in D_A \times \mathcal{P}(S)$, and
2. for each $\mathbf{r} \in \mathbf{Rol}$, $v(\mathbf{r}) \in D_A$ such that
3. if $v(\mathbf{s}) = \langle d, P \rangle$, then $v(\mathbf{r}) \neq d$.

The domain of the model \mathfrak{M} , $|\mathfrak{M}|$ is $\{ (s, h) \mid s \in h \}$.

¹See Caleiro et al. (1999).

²The subscript ‘ x ’ is to differentiate it from the new use of \mathfrak{F} to denote an \mathcal{L}^I frame.

As we saw before, \mathcal{D} is used to interpret the propositional content of a sentence, and \mathfrak{F}_x is used to interpret the xstit fragment of the language. However, there are some differences between this definition and its predecessors. We will start with the differences in \mathcal{D} from definition 5.2.1. The reason the role terms are assigned propositional content is because which role does what has an affect on the proposition expressed by $[\mathbf{R} \text{ xstit}] \varphi$.

Each atomic sentence gets assigned an atomic element in the partial order of \mathcal{D} , and some subset of S . But now we have to interpret the input to propositional content that the role terms will make. Each role term is assigned some atom from D , but no sentence and role term may be assigned the same atom in D . One might suggest that each set in $\mathcal{P}(\mathbf{Rol})$ be assigned an element of D_A , but there is the looming question of the appropriate relationship between a group and its members that we do not wish to beg. We don't want to further complicate matters in the formalism here, so we will simply treat the content that $\mathbf{R} \subseteq \mathbf{Rol}$ contributes to $[\mathbf{R} \text{ xstit}] \varphi$ as given by $\bigvee \{ v(\mathbf{r}') : \mathbf{r}' \in \mathbf{R} \}$.

As before, the domain of the model is determined by all of the dynamic states that are composed of pairs of static states and the histories they are present in. The major difference is that we do not assume that the effectivity functions satisfy the condition (g). In these frames $E : S \times H \times \mathcal{P}(\mathbf{Rol}) \rightarrow \mathcal{P}(S)$ is a function that specifies, relative to a dynamic state, that a group of roles is effective to ensure that the future continuations are among $E(s, h, \mathbf{R})$. The function E must obey only the following conditions in this case:

- (a) if $s \notin h$, then $E(s, h, \mathbf{R}) = \emptyset$
- (b) if $s' \in E(s, h, \mathbf{R})$, then $s' \in lub(s)$
- (c) if $s \in h$, $lub(s, h) \in E(s, h, \mathbf{R})$
- (d) $E(s, h, \emptyset) = lub(s)$
- (e) if $s \in h$, then $E(s, h, \mathbf{Rol}) = \{ lub(s, h) \}$
- (f) if $\mathbf{R} \subsetneq \mathbf{R}'$, then $E(s, h, \mathbf{R}') \subseteq E(s, h, \mathbf{R})$

In this new scenario the logic will be what governs norm consequence. Most of these conditions are acceptable for a logic that is supposed to play this role. Certainly conditions (a)–(c) are unproblematic given their interpretation from section 5.1.1. Condition (d) is still acceptable since it specifies that the empty set of roles, i.e., no one, still can't constrain the evolution of history beyond what is determined by nature. Condition (e) specifies that the total actions of all of the roles according to the same history moves the institution into its next state along that history. That is still acceptable, perhaps even more so given that institutions are simply constituted by its set of roles.

Condition (f) specifies that the larger the group of roles considered, the more the future possibilities are determined. Indeed, the more institutional roles included in the determination of a next state, the more closely predictable the next state will be. But condition (g) gives too much credit to those who construct institutions. Condition (g) says if $\mathbf{R} \cap \mathbf{R}' = \emptyset$ and $s \in h \cap h'$, then there is h'' with $s \in h''$ and $E(s, h'', \mathbf{R})$ and $E(s, h'', \mathbf{R}')$ are contained in $E(s, h, \mathbf{R})$ and $E(s, h', \mathbf{R}')$, respectively. This means that there is no way for two *disjoint* sets of roles to act against each other. Put another way, if \mathbf{R} is capable of ensuring something happens, then \mathbf{R} 's actions cannot frustrate the actions of another, disjoint group \mathbf{R}' . Certainly that has to be false when it comes to modern bureaucracies, so it would be a mistake to resolve that problem by denying its existence. Now we return to the formal descriptions.

Again we extend the valuation to get a semantic value for any sentence. The semantic value of a term or sentence θ in \mathcal{L}^I relative to a model $\mathfrak{M} = \langle \mathcal{D}, \langle S, H, E \rangle, v \rangle$ is referred to as $\llbracket \theta \rrbracket^{\mathfrak{M}}$. We will define $\llbracket \cdot \rrbracket^{\mathfrak{M}}$ for the fragment of $\mathcal{L}_{\subseteq}^I, \mathcal{L}^I$, and leave the superscript ' \mathfrak{M} ' implicit. Recall from section 5.2 that $\llbracket \theta \rrbracket_1$ refers to an element of D , and $\llbracket \theta \rrbracket_2$ refers to a subset of the domain, so in this case to an element of $\mathcal{P}(|\mathfrak{M}|)$. Let $\theta, \theta' \in \mathcal{L}^I$ and $\mathbf{R} \subseteq \mathbf{Rol}$:

At for $s \in \mathbf{At}_B \cup \mathbf{At}_I \cup \{V\}$, and $v(s) = \langle d, P \rangle$,

$$\llbracket s \rrbracket = \langle d, \{ (s, h) \in |\mathfrak{M}| \mid s \in P \} \rangle$$

$$\mathbf{R} \llbracket \mathbf{R} \rrbracket = \bigcap \{ v(\mathbf{r}') : \mathbf{r}' \in \mathbf{R} \}$$

$$\perp \llbracket \perp \rrbracket = \langle \odot, \emptyset \rangle$$

$$\top \llbracket \top \rrbracket = \langle \odot, |\mathfrak{M}| \rangle$$

$$\mathbf{And} \llbracket \theta \wedge \theta' \rrbracket = \langle \llbracket \theta \rrbracket_1 \vee \llbracket \theta' \rrbracket_1, \llbracket \theta \rrbracket_2 \cap \llbracket \theta' \rrbracket_2 \rangle$$

$$\mathbf{Or} \llbracket \theta \wedge \theta' \rrbracket = \langle \llbracket \theta \rrbracket_1 \vee \llbracket \theta' \rrbracket_1, \llbracket \theta \rrbracket_2 \cup \llbracket \theta' \rrbracket_2 \rangle$$

$$\mathbf{If} \llbracket \theta \wedge \theta' \rrbracket = \langle \llbracket \theta \rrbracket_1 \vee \llbracket \theta' \rrbracket_1, (|\mathfrak{M}| \setminus \llbracket \theta \rrbracket_2) \cup \llbracket \theta' \rrbracket_2 \rangle$$

$$\mathbf{Iff} \llbracket \theta \wedge \theta' \rrbracket =$$

$$\langle \llbracket \theta \rrbracket_1 \vee \llbracket \theta' \rrbracket_1, [(|\mathfrak{M}| \setminus \llbracket \theta \rrbracket_2) \cup \llbracket \theta' \rrbracket_2] \cap [(|\mathfrak{M}| \setminus \llbracket \theta' \rrbracket_2) \cup \llbracket \theta \rrbracket_2] \rangle$$

$$\mathbf{Not} \llbracket \neg \theta \rrbracket = \langle \llbracket \theta \rrbracket_1, |\mathfrak{M}| \setminus \llbracket \theta \rrbracket_2 \rangle$$

$$\mathbf{X} \llbracket X\theta \rrbracket = \langle \llbracket \theta \rrbracket_1, \{ (s, h) \mid (lub(s, h), h) \in \llbracket \theta \rrbracket_2 \} \rangle$$

$$\mathbf{P} \llbracket P\theta \rrbracket = \langle \llbracket \theta \rrbracket_1, \{ (s, h) \mid (glb(s, h), h) \in \llbracket \theta \rrbracket_2 \} \rangle$$

$$\Box \llbracket \Box \theta \rrbracket = \langle \llbracket \theta \rrbracket_1, \{ (s, h) \mid \forall h', \text{ s.t. } s \in h', (lub(s, h'), h') \in \llbracket \theta \rrbracket_2 \} \rangle$$

$$\mathbf{xstit} \llbracket [\mathbf{R} \text{ xstit}] \theta \rrbracket =$$

$$\langle \llbracket \mathbf{R} \rrbracket_1 \vee \llbracket \theta \rrbracket_1, \{ (s, h) \mid \forall h', s' \text{ s.t. } s' \in E(s, h, \mathbf{R}), (s', h') \in \llbracket \theta \rrbracket_2 \} \rangle$$

Note that V is treated as any other atomic sentence. Here the propositional content of the atomic sentences are all atoms of \mathcal{D} . The contents of Boolean sentences are just joins of contents, as Vanderveken's theory had it. The difference here is that X , P and \Box don't do anything to the propositional content of the complement θ of $X\theta$, $\Box\theta$ or $P\theta$. Modal operators affect the informational content of the sentence, not its propositional content. The sentence 'the door will be closed in the next state' and 'the door is closed' say different things, but the content is the same. The former just says that the state of affairs of the door being closed happens in the next state, the latter says that the same state of affairs is actual. This difference in informational content is captured in the difference between $\llbracket \theta \rrbracket_2$ and $\llbracket X\theta \rrbracket_2$, but $\llbracket \theta \rrbracket_1 = \llbracket \theta \rrbracket_1$, so they have the same propositional content. Similarly for \Box and P .

For the $[\mathbf{R} \text{ xstit}] \theta$ -case, the complement is affected since the xstit-proposition concerns *who* sees-to-it-that θ is true. So when there are two roles that should be distinct we have to adjoin different atomic elements of \mathcal{D} to $\llbracket \theta \rrbracket_1$. That way the content of $[\{\mathbf{r}\} \text{ xstit}] \theta$ and $[\{\mathbf{r}'\} \text{ xstit}] \theta$ can be different, as long as $v(\mathbf{r}) \neq v(\mathbf{r}')$.

Now we can provide the satisfaction conditions as follows:

Definition 6.3.3. For formulas in $\varphi, \psi \in \mathcal{L}_{\subseteq}^I$, $\theta, \theta' \in \mathcal{L}^I$, and $\mathbf{R} \subseteq \mathbf{Rol}$ with an $\mathcal{L}_{\subseteq}^I$ -model \mathfrak{M} with $s \in S$ and $h \in H$,

- $(s, h) \models \mathbf{p}$ iff $s \in \llbracket \mathbf{p} \rrbracket_2$ where $\mathbf{p} \in \mathbf{At}_I \cup \mathbf{At}_B$
- $(s, h) \models V$ iff $s \in \llbracket V \rrbracket_2$
- $(s, h) \not\models \perp$
- $(s, h) \models \neg\varphi$ iff $(s, h) \not\models \varphi$
- $(s, h) \models \varphi \wedge \psi$ iff $(s, h) \models \varphi$ and $(s, h) \models \psi$
- $(s, h) \models \Box\theta$ iff for all h' with $s \in h'$, $(s, h') \models \theta$
- $(s, h) \models X\theta$ iff $(lub(s, h), h) \models \theta$
- $(s, h) \models P\theta$ iff $(glb(s, h), h) \models \theta$
- $(s, h) \models [\mathbf{R} \text{ xstit}] \theta$ iff for all s', h' , if $s' \in E(s, h, \mathbf{R})$ and $s' \in h'$, then $(s', h') \models \theta$
- $(s, h) \models \theta \subseteq \theta'$ iff $\llbracket \theta \rrbracket_1 \lesssim \llbracket \theta' \rrbracket_1$

Definition 6.3.4. If $\Gamma, \varphi \subseteq \mathcal{L}_{\subseteq}^I$, then $\Gamma \models_{\text{Ixp}} \varphi$ iff for all $\mathcal{L}_{\subseteq}^I$ -models \mathfrak{M} , and $(s, h) \in |\mathfrak{M}|$, if $\mathfrak{M}, (s, h) \models \gamma$ for all $\gamma \in \Gamma$, then $\mathfrak{M}, (s, h) \models \varphi$.

At this point we provide a set of axioms that are complete relative to the semantics.

6.3.2 Proof Theory

Again we provide a Hilbert style proof theory for the logic. The axiom system extends the two previous systems of definitions 5.2.3 and 5.1.4. We will refer to this proof system and its consequence relation as \vdash_{Ixp} .

Definition 6.3.5. Axioms of \vdash_{Exp} .

1. First we include all axioms for classical logic (Group CL axioms) where $\varphi, \psi \in \mathcal{L}_{\subseteq}^I$:

$$\text{CL1 } \varphi \supset (\psi \supset \varphi)$$

$$\text{CL2 } (\varphi \supset (\psi \supset \theta)) \supset ((\varphi \supset \psi) \supset (\psi \supset \theta))$$

$$\text{CL3 } (\varphi \wedge \psi) \supset \psi$$

$$\text{CL4 } (\varphi \wedge \psi) \supset \varphi$$

$$\text{CL5 } (\varphi \supset \psi) \supset ((\varphi \supset \theta) \supset (\varphi \supset \psi \wedge \theta))$$

$$\text{CL6 } \varphi \supset (\varphi \vee \psi)$$

$$\text{CL7 } \psi \supset (\varphi \vee \psi)$$

$$\text{CL8 } (\varphi \supset \psi) \supset ((\theta \supset \psi) \supset (\varphi \vee \theta \supset \psi))$$

$$\text{CL9 } (\psi \supset \neg\varphi) \supset (\varphi \supset \neg\psi)$$

$$\text{CL10 } \neg(\psi \supset \psi) \supset \varphi$$

$$\text{CL11 } \varphi \vee \neg\varphi$$

$$\text{CL12 } (\varphi \wedge \neg\varphi) \supset \perp$$

2. We extend the axioms for propositional containment (group PC axioms) where $A, B, C \in \mathcal{L}_{\subseteq}^I$, $\mathbf{s}, \mathbf{s}' \in \mathbf{At}_I \cup \mathbf{At}_B \cup \{V\}$ and $\mathbf{R} \cup \{\mathbf{r}, \mathbf{r}'\} \subseteq \mathbf{Rol}$:

$$\text{PC1 } A \in A$$

$$\text{PC2 } (B \in A) \supset ((C \in B) \supset (C \in A))$$

$$\text{PC3 } (\mathbf{s} \in \mathbf{s}') \supset (\mathbf{s}' \in \mathbf{s})$$

$$\text{PC4 } A \in (A \wedge B)$$

$$\text{PC5 } B \in (A \wedge B)$$

$$\text{PC6 } (B \in A) \supset ((C \in A) \supset ((C \wedge B) \in A))$$

$$\text{PC7 } A \in \neg A$$

PC8 $\neg A \in A$

PC9 $(s \in (A \wedge B)) \supset ((s \in A) \vee (s \in B))$

PC10 $\perp \in A$

PC11 $A \in (A \vee B)$

PC12 $B \in (A \vee B)$

PC12A $(A \vee B) \in (A \wedge B)$

PC13 $A \in (A \supset B)$

PC14 $B \in (A \supset B)$

PC14A $(A \supset B) \in (A \wedge B)$

PC15 $A \in (A \equiv B)$

PC16 $B \in (A \equiv B)$

PC16A $(A \equiv B) \in (A \wedge B)$

PC17 $\top \in A$

PCX1 $A \in \Box A$

PCX2 $XA \in A$

PCX3 $PA \in XA$

PCX4 $\Box A \in PA$

PCX5 $A \in [\mathbf{r} \text{ xstit}] A$

PCX6 $\mathbf{R} \in [\mathbf{R} \text{ xstit}] A$

PCX7 $\perp \in \mathbf{r}$

PCX8 $(\{\mathbf{r}\} \in \{\mathbf{r}'\}) \supset (\{\mathbf{r}'\} \in \{\mathbf{r}\})$

PCX9 $\{\mathbf{r}\} \in \mathbf{R} \text{ for } \mathbf{r} \in \mathbf{R} \subseteq \mathbf{R} \mathbf{ol}$

PCX10 $\neg(\{\mathbf{r}\} \in \mathbf{p}) \wedge \neg(\mathbf{p} \in \{\mathbf{r}\})$

3. We then extend the axioms for xstit (we call these the XPstit-group) by the following for

$$\theta, \theta' \in \mathcal{L}^I,$$

$$(p) \quad p \supset \Box p$$

S5 for \Box :

$$K \quad \Box(\theta \supset \theta') \supset (\Box\theta \supset \Box\theta')$$

$$T \quad \Box\theta \supset \theta$$

$$4 \quad \Box\theta \supset \Box\Box\theta$$

$$B \quad \theta \supset \Box\neg\Box\neg\theta$$

KD for each $[\mathbf{R} \text{ xstit}] \theta$, $\mathbf{R} \subseteq \mathbf{Rol}$, P and X :

$$KR \quad [\mathbf{R} \text{ xstit}](\theta \supset \theta') \supset ([\mathbf{R} \text{ xstit}] \theta \supset [\mathbf{R} \text{ xstit}] \theta')$$

$$DR \quad [\mathbf{R} \text{ xstit}] \theta \supset \neg [\mathbf{R} \text{ xstit}] \neg\theta$$

$$KX \quad X(\theta \supset \theta') \supset (X\theta \supset X\theta')$$

$$DX \quad X\theta \supset \neg X\neg\theta$$

$$KP \quad P(\theta \supset \theta') \supset (P\theta \supset P\theta')$$

$$DP \quad P\theta \supset \neg P\neg\theta$$

$$(\text{DetX}) \quad \neg X\neg\theta \supset X\theta$$

$$(\text{DetP}) \quad \neg P\neg\theta \supset P\theta$$

$$(XP) \quad XP\theta \equiv \theta$$

$$(PX) \quad \theta \equiv PX\theta$$

$$(NP) \quad P\Box\theta \supset \Box P\theta$$

$$(\text{SettX}) \quad [\emptyset \text{ xstit}] \theta \equiv \Box X\theta$$

$$(\text{XSett}) \quad [\mathbf{Rol} \text{ xstit}] \theta \equiv X\Box\theta$$

$$(\text{C-mon}) \quad [\mathbf{R} \text{ xstit}] \theta \supset [\mathbf{R} \cup \mathbf{R}' \text{ xstit}] \theta \text{ where } \mathbf{R}' \cup \mathbf{R} \subseteq \mathbf{Rol}$$

4. Rules: MP and Nec \clubsuit for $\clubsuit \in \{ \Box, X, P, [\mathbf{R} \text{ xstit}] : \mathbf{R} \subseteq \mathbf{Rol} \}$

It is *very* important to note which axioms apply to which classes of formulas. The addition of the axioms PCX1–5 capture relationships between the contents of the new expressions from \mathcal{L}^I . Also note that the truth of a \subseteq -sentence $A \subseteq B$ depends only on what members of D are assigned to A and B , but that is independent of the xstit part of an $\mathcal{L}_{\subseteq}^I$ -model. However, in the language $\mathcal{L}_{\subseteq}^I$, we cannot say anything that would reflect that fact, e.g., $\Box(A \subseteq B) \equiv (A \subseteq B)$, since that formula isn't well formed.

The XP-stit axioms are largely the same from definition 5.1.4 with the exception that they are in terms of roles \mathbf{R} rather than agents \mathbf{A} . There is the addition of the axioms for P , and those mirror the axioms for X . That is to be expected since P is kind of like an inverse of X , as XP and PX indicate. There are no interaction axioms for P and the $[\mathbf{R} \text{ xstit}]$ -operators since $[\mathbf{R} \text{ xstit}]$ is a forward looking operator, and P is backward looking. The axiom NP is there to indicate that the past is necessary, i.e., the histories do not branch into the past. What should be noted is the absence of Indep-G. The axiom Indep-G corresponds to condition (g) on the effectivity function from the model, which is not required in the semantics of \mathcal{L}^I (see p. 92), so Indep-G fails. From here we sketch a completeness proof for this system. We leave the detailed completeness proof for chapter 7.

6.3.3 Soundness and Completeness

The completeness proof for \vdash_{Ixp} with respect to \models_{Ixp} proceeds in a number of stages. In order to complete the proof we first give a completeness proof for a logic based on the language \mathcal{L} , we refer to that logic—i.e., its consequence relation—as \vdash_{xp} . The logic \vdash_{xp} can be defined as the Hilbert style axiom system consisting of CL-group axioms and the XP-Stit group axioms from definition 6.3.5 above. However, we replace the role terms with agent terms. The axiomatization is as follows:

Definition 6.3.6. Axioms for \vdash_{xp} .

1. Axioms for classical logic

$$(\mathbf{p}) \quad \mathbf{p} \supset \Box \mathbf{p}, \mathbf{p} \in \mathbf{At}_B \cup \mathbf{At}_I$$

S5 for \Box :

$$\mathbf{K} \quad \Box(\theta \supset \theta') \supset (\Box\theta \supset \Box\theta')$$

$$\mathbf{T} \quad \Box\theta \supset \theta$$

$$4 \quad \Box\theta \supset \Box\Box\theta$$

$$\mathbf{B} \quad \theta \supset \Box\neg\Box\neg\theta$$

KD for each $[\mathbf{A} \text{ xstit}] \theta$, $\mathbf{A} \subseteq \mathbf{Ag}$, P and X :

$$\mathbf{KA} \quad [\mathbf{A} \text{ xstit}](\theta \supset \theta') \supset ([\mathbf{A} \text{ xstit}] \theta \supset [\mathbf{A} \text{ xstit}] \theta')$$

$$\mathbf{DA} \quad [\mathbf{A} \text{ xstit}] \theta \supset \neg [\mathbf{A} \text{ xstit}] \neg\theta$$

$$\mathbf{KX} \quad X(\theta \supset \theta') \supset (X\theta \supset X\theta')$$

$$\mathbf{DX} \quad X\theta \supset \neg X\neg\theta$$

$$\mathbf{KP} \quad P(\theta \supset \theta') \supset (P\theta \supset P\theta')$$

$$\mathbf{DP} \quad P\theta \supset \neg P\neg\theta$$

$$[(\text{DetX})] \neg X\neg\theta \supset X\theta$$

$$[(\text{DetP})] \neg P\neg\theta \supset P\theta$$

$$[(\text{XP})] XP\theta \equiv \theta$$

$$[(\text{PX})] \theta \equiv PX\theta$$

$$[(\text{NP})] P\Box\theta \supset \Box P\theta$$

$$[(\text{SettX})] [\emptyset \text{ xstit}] \theta \equiv \Box X\theta$$

$$[(\text{XSett})] [\mathbf{Ag} \text{ xstit}] \theta \equiv X\Box\theta$$

$$[(\text{C-mon})] [\mathbf{A} \text{ xstit}] \theta \supset [\mathbf{A} \cup \mathbf{B} \text{ xstit}] \theta$$

$$[(\text{Indep-G})] \Diamond [\mathbf{A} \text{ xstit}] \theta \wedge \Diamond [\mathbf{B} \text{ xstit}] \theta' \supset \Diamond ([\mathbf{A} \text{ xstit}] \theta \wedge [\mathbf{B} \text{ xstit}] \theta') \text{ where } \mathbf{A} \cap \mathbf{B} = \emptyset.$$

2. Rules: MP and Nec \clubsuit for $\clubsuit \in \{ \Box, X, P, [\mathbf{A} \text{ xstit}] : \mathbf{A} \subseteq \mathbf{Ag} \}$

In chapter 7 we don't distinguish between \mathbf{At}_B and \mathbf{At}_I since they don't play a role in the logic, just in the relation to norm consequence as we will discuss in section 6.5. We then show that if condition g for effectivity functions fails in the class of models, then we can invalidate Indep-G. This provides a completeness proof for the set of axioms for \vdash_{xp}^I which are given as

Definition 6.3.7. Axioms for \vdash_{xp}^I

1. Axioms for classical logic

(p) $p \supset \Box p, p \in \mathbf{At}_B \cup \mathbf{At}_I$

S5 for \Box :

$$K \quad \Box(\theta \supset \theta') \supset (\Box\theta \supset \Box\theta')$$

$$T \quad \Box\theta \supset \theta$$

$$4 \quad \Box\theta \supset \Box\Box\theta$$

$$B \quad \theta \supset \Box\neg\Box\neg\theta$$

KD for each $[\mathbf{R} \text{ xstit}] \theta, \mathbf{R} \subseteq \mathbf{Rol}, P$ and X :

$$KR \quad [\mathbf{R} \text{ xstit}](\theta \supset \theta') \supset ([\mathbf{R} \text{ xstit}] \theta \supset [\mathbf{R} \text{ xstit}] \theta')$$

$$DR \quad [\mathbf{R} \text{ xstit}] \theta \supset \neg[\mathbf{R} \text{ xstit}] \neg\theta$$

$$KX \quad X(\theta \supset \theta') \supset (X\theta \supset X\theta')$$

$$DX \quad X\theta \supset \neg X\neg\theta$$

$$KP \quad P(\theta \supset \theta') \supset (P\theta \supset P\theta')$$

$$DP \quad P\theta \supset \neg P\neg\theta$$

$$[(\text{DetX})] \neg X\neg\theta \supset X\theta$$

$$[(\text{DetP})] \neg P\neg\theta \supset P\theta$$

$$[(\text{XP})] XP\theta \equiv \theta$$

$$[(\text{PX})] \theta \equiv PX\theta$$

$$[(\text{NP})] P \Box \theta \supset \Box P \theta$$

$$[(\text{SettX})] [\emptyset \text{ xstit}] \theta \equiv \Box X \theta$$

$$[(\text{XSett})] [\mathbf{R} \text{ol xstit}] \theta \equiv X \Box \theta$$

$$[(\text{C-mon})] [\mathbf{R} \text{ xstit}] \theta \supset [\mathbf{R} \cup \mathbf{R}' \text{ xstit}] \theta$$

2. Rules: MP and Nec \clubsuit for $\clubsuit \in \{ \Box, X, P, [\mathbf{R} \text{ xstit}] : \mathbf{R} \subseteq \mathbf{R} \text{ol} \}$

That gives us completeness of the logics \vdash_{xp} and \vdash_{xp}^I . That completeness proof allows us to prove that we can then build canonical models for \mathcal{L}_{\in}^I by, essentially, combining an \mathcal{L} frame with an extension of an SI-frame (see definition 5.2.1). This is, as far as we can tell, the first attempt to construct a canonical model for the xstit logic. There haven't been detailed completeness proofs like this in print, and particularly not for these models where the frames are constructed from histories that are copies of \mathbb{Z} from Broersen and Meyer (2011), what we call *regular* models.

An interesting point to note is that the canonical model based on the logic of definition 6.3.6 isn't a model/frame in the class of models from definitions 5.1.1 and 5.1.2. It is even worse since the models are based on what Thomason (1984) calls a Neutral Frame. But adding the P operator and doing some fancy footwork allows us to construct a model that is in the right class to invalidate all of the unprovable \vdash_{xp} -arguments. I.e., what we show is that every model like those in definitions 5.1.1 and 5.1.2 is a neutral model, and for any model like that in definitions 5.1.1 and 5.1.2, we can find a model based on a neutral frame that refutes all of the \vdash_{xp} -theorems.

There is a final consequence relation that we will make use of: $\vdash_{\text{xp}}^{\Omega}$. This consequence relation is defined by the set of axioms for \vdash_{xp} , however, it is defined over the language \mathcal{L}_{Ω}^B . This relation is used to restrict the notion of \vdash_{xp} -consequence to the language that occurs in the code Ω . It simply is \vdash_{xp} ; however, there will only be atoms from $at(\Omega)$ included in the expansion of the language.

Before we move on we will say some things by way of comparison with previous work. The

semantics from definition 5.1.1 are called various things in the literature, e.g., bundled trees or Kamp frames.³ The major difference between our work and the other work is that the language \mathcal{L} doesn't have an operator that looks at *all* the future or past, but only *one step* ahead and behind. The closest that we have seen is that of Ciuni and Zanardo (2010) and Zanardo and Carmo (1993). But these do not give up the “all future times” operators. Here we do. But not without good reason. If a language has a “next state” operator, and an “all future states” operator, then the semantic consequence relation won't be compact. Using the standard notation of ' $G\varphi$ ' for 'in all future states φ is true', the set $\{X^n\varphi : n \in \mathbb{N}\}$ will entail $G\varphi$, but no finite subset will entail $G\varphi$. However, in the sequel (section 8.3) compactness is important for our system. Thus we will take the road into the future one step/state at a time.

6.4 How to Say Things Without Words: Expressing Legal Relations

Now that we have the language $\mathcal{L}_{\subseteq}^I$ we can look at how to say certain things that are involved in constructing a system of norms, in Searle's phrase: a deontology. We want to show how to express sentences like 'employees must submit their time sheets', and 'managers are prohibited from using company cars for non-work related purposes' and 'the president of the company has the ability to purchase new equipment for the production plant'. We will do this by exploiting terminology from Holfeld (1920), and focusing on representing his legal relations.

The first step is to recognize that we are working within a *propositional* language, so some elements of sentences will go unanalysed. So a phrase like 'employees must submit their time sheets' must be paraphrased like 'employees must see to it that *the employee's time sheet is submitted*'. In general we would like to capture the expressions in table 6.1 where x is a role term, and p is a proposition.⁴

³See Zanardo (1996) for an in-depth overview and Thomason (1984), for an equally broad but less detailed overview. Such structures are like Ockamist frames but there may be some slight differences.

⁴There is a long literature on formal representations of legal relations starting with Holfeld (1920), and extending to the new millennium, See Kanger and Kanger (1966), Fitch (1967), Makinson (1986) and Hansson (2001, Ch. 13). These authors have extended and criticized Holfeld's work, but due to lack of space we will avoid an in depth discussion of the subtle points in this literature.

$\text{power}(x, p)$	x has the power to bring about p
$\text{disability}(x, p)$	x is unable to bring about p
$\text{duty}(x, p)$	x has a duty to bring about p
$\text{prohibition}(x, p)$	x is prohibited from bringing about p
$\text{exemption}(x, p)$	x is exempted from bringing about p
$\text{privilege}(x, p)$	x has a privilege to bring about p
$\text{right}(x, p)$	x has a right to bring about p

Table 6.1: Target Expressions

We start by showing how to express $\text{power}(x, p)$, and treat each in turn. These relations are evaluated in this framework relative to dynamic states: (s, h) . If we are considering what powers $\mathbf{R} \subseteq \mathbf{Rol}$ has, we are really asking what \mathbf{R} is able to do. So \mathbf{R} 's powers depend on what \mathbf{R} is effective for relative to all of the histories coincident with h at s , i.e., all h' such that $s \in h'$. Thus $\text{power}(\mathbf{R}, p)$ can be expressed as $\Diamond [\mathbf{R} \text{ xstit}] p$. Similarly, disability is a lack of power, so $\text{disability}(\mathbf{R}, p)$ can be expressed as $\neg \Diamond [\mathbf{R} \text{ xstit}] p$.

Duty and prohibition are trickier items. Since our language \mathcal{L}^I contains the violation constant V , and according to Searle's view, institutions define things like duties, we should extend Anderson's reduction—recall section 5.1.3—to define institutional duties and prohibitions. A standard way⁵ of expressing obligation and forbiddance, i.e., duty and prohibition, in stit logic using Anderson's reduction is via a sentence like $\neg[\mathbf{R} \text{ stit} : \varphi] \supset V$, which is read as 'refraining from seeing to φ results in a violation' for duty; and $[\mathbf{R} \text{ stit} : \varphi] \supset V$ read as 'seeing to φ is a violation' for prohibition. The regular stit operator from Belnap and Perloff (1992), Bartha (1993) and Horty (2001) acts on the current state of evaluation, whereas the $[\mathbf{R} \text{ xstit}]$ operator is future looking. So the question is: does the violation occur now or later? If we express $\text{prohibition}(\mathbf{R}, \varphi)$ as $\neg[\mathbf{R} \text{ xstit}] \varphi \supset V$, then it means that failing to make a choice that leads to φ is a violation now. Saying something like $\neg[\mathbf{R} \text{ xstit}] \varphi \supset XV$, is to say that failing to make a choice that leads to φ leads to a violation in the next state of the current history. To remain consistent with Searle, and Grossi, we will formalize: \mathbf{R} 's failure to see to φ *counts as* a violation.

⁵Cf. Bartha (1993).

What we will argue next is that the best way to represent that is by: $\neg [\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V$. This is read as: \mathbf{R} failing to see to φ in the next states results in \mathbf{R} seeing to a violation in the next state.

We give one positive argument and one negative argument for the view that the violation should be in the future and represented in the way we suggest. These arguments focus on prohibition, but duty and prohibition are developed in analogy with one another. The standard option for prohibition is $\Box([\mathbf{R} \text{ xstit}] \varphi \supset V)$. But in this formalism, since V is treated like an atomic sentence, if $(s, h) \models V$, then $(s, h') \models V$ for all h' with $s \in h'$. So if it is prohibited to see to φ , then it is a violation to choose to see to φ . However, any choice that doesn't lead to φ would intuitively not be in violation of the norms. But suppose that $\Diamond([\mathbf{R} \text{ xstit}] \varphi)$, and $\Box([\mathbf{R} \text{ xstit}] \varphi \supset V)$ are true at (s, h) . Then there is h' such that $(s, h') \models [\mathbf{R} \text{ xstit}] \varphi$, so $(s, h') \models V$. But that means $(s, h) \models V$. So if \mathbf{R} is capable of doing something wrong, then \mathbf{R} is already in violation. But that can't be right.

The positive argument has to do with reasoning about prohibitions. In our system there can be classifications of states as violations. Intuitively, if φ is classified as a violation, then it should be prohibited to bring about φ . However, interpreting prohibition as $\Box([\mathbf{R} \text{ xstit}] \varphi \supset V)$ doesn't guarantee that prohibitions against φ arise from classifying φ as a violation. Suppose that φ is classified as a violation, then $\varphi \supset V$ is always true, i.e., $(s, h) \models \Box X^n(\varphi \supset V)$ for $n \in \mathbb{N}$. But $[\mathbf{R} \text{ xstit}] \varphi$ doesn't imply that φ is true, so V doesn't have to be true even though $[\mathbf{R} \text{ xstit}] \varphi$ is. However, and this is the positive argument, if prohibition is interpreted as $[\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V$, then we get the result: assume $(s, h) \models \Box X^n(\varphi \supset V)$ for $n \in \mathbb{N}$, then $(s, h) \models \Box X(\varphi \supset V)$, so then $(s, h) \models [\emptyset \text{ xstit}](\varphi \supset V)$ by settX , so $(s, h) \models [\mathbf{R} \text{ xstit}](\varphi \supset V)$ is by C-mon, so by K for $[\mathbf{R} \text{ xstit}]$, we get $(s, h) \models [\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V$. Whereas interpreting 'prohibited(\mathbf{R}, φ)' as ' $[\mathbf{R} \text{ xstit}] \varphi \supset V$ ' results in two counterintuitive results within the formalism, $[\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V$ gives us exactly what we would expect.

This means that prohibition is put in terms of making choices that lead to violations. Of course when we want to express that something is a duty or prohibited always, that means it is

so relative to any dynamic state. Thus to express prohibition we need to say $\Box([\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$. Duty is similar to prohibition except that it is the failure to see to something that is classified as seeing to a violation, in all cases, i.e., $\Box(\neg [\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$.

Now we come to another two correlated concepts *exemption* and *privilege*. A role \mathbf{R} has a privilege when it is possible for the role to do something without violation. The way that we interpret this is by saying that it is possible for \mathbf{R} to see to φ , and be guaranteed not to see to a violation, i.e., $\Diamond([\mathbf{R} \text{ xstit}] \varphi \wedge \neg \langle \mathbf{R} \text{ xstit} \rangle V)$. Likewise, exemption is not doing something and avoiding violation: $\Diamond(\neg [\mathbf{R} \text{ xstit}] \varphi \wedge \neg \langle \mathbf{R} \text{ xstit} \rangle V)$. Of course privilege and exemption usually have a connotation that \mathbf{R} is doing something that it usually wouldn't be allowed to do. Usually exemptions are equivalent to the negation of a duty, and a privilege is the negation of a prohibition. However, because of axiom DR we get the weaker:

Observation 6.4.1. For $\mathbf{R} \subseteq \mathbf{Rol}$ and $\varphi \in \mathcal{L}_{\subseteq}^I$,

1. $\Diamond([\mathbf{R} \text{ xstit}] \varphi \wedge \neg \langle \mathbf{R} \text{ xstit} \rangle V) \vdash_{\text{Exp}} \neg \Box([\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$.
2. $\Diamond(\neg [\mathbf{R} \text{ xstit}] \varphi \wedge \neg \langle \mathbf{R} \text{ xstit} \rangle V) \vdash_{\text{Exp}} \neg \Box(\neg [\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$

So a privilege implies the negation of a prohibition, and an exemption implies the negation of a duty.

Finally we come to a right. There is a long literature on how to represent rights that we don't want to engage with here. That can be displaced to future work. We will be to focus on just one interpretation of what a right is, and admit that it may not be the best nor the state of the art. Intuitively, \mathbf{R} has a right to bring about φ , when no one is permitted to interfere with \mathbf{R} 's ability to bring φ about. Another way to put that is to say that everyone else is prohibited from interfering with \mathbf{R} 's ability to bring φ about. This latter paraphrase is close to something we have already discussed. The major obstacle is the 'everyone else' in the paraphrase.

For instance if we were attributing a right to walk dogs to a role \mathbf{r} , then we could say that any role \mathbf{r}' interfering with \mathbf{r} 's seeing to it that \mathbf{r} walks a dog, then \mathbf{r}' is seeing to a violation.

$\text{power}(\mathbf{R}, \varphi)$	$\Diamond [\mathbf{R} \text{ xstit}] \varphi$
$\text{disability}(\mathbf{R}, \varphi)$	$\neg \Diamond [\mathbf{R} \text{ xstit}] \varphi$
$\text{duty}(\mathbf{R}, \varphi)$	$\Box (\neg [\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$
$\text{prohibition}(\mathbf{R}, \varphi)$	$\Box ([\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$
$\text{exemption}(\mathbf{R}, \varphi)$	$\Diamond ([\mathbf{R} \text{ xstit}] \varphi \wedge \neg [\mathbf{R} \text{ xstit}] V)$
$\text{privilege}(\mathbf{R}, \varphi)$	$\Diamond (\neg [\mathbf{R} \text{ xstit}] \varphi \wedge \neg [\mathbf{R} \text{ xstit}] V)$
$\text{right}(\mathbf{R}, \varphi)$	$\bigwedge_{\mathbf{R}' \subseteq \mathbf{Rol} \setminus \{\mathbf{R}\}} \Box ([\mathbf{R}' \text{ xstit}] \neg [\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R}' \text{ xstit}] V)$

Table 6.2: Institutional Relations

But also since \mathbf{Rol} is finite $\mathcal{P}(\mathbf{Rol})$ is finite. So generally we can express \mathbf{R} s right to see to φ as

$$\bigwedge_{\mathbf{R} \subseteq \mathbf{Rol} \setminus \{\mathbf{r}\}} \Box ([\mathbf{R} \text{ xstit}] \neg [\mathbf{r} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V) \quad (6.1)$$

This formula takes the conjunction of every formula that says \mathbf{R} seeing to it that \mathbf{r} doesn't bring about φ , is bringing about a violation. However we have restricted the \mathbf{R} s so that they don't include \mathbf{r} . This is an odd point of debate whether an individual can interfere with his/her own rights. We have just taken a side and said no. So now we have represented all of the relations in table 6.1. We collect our findings about how to express these institutional relation together in table 6.2.

6.5 Normative Entailment

Using the formal machinery developed so far we will characterize the relation of normative entailment for the language \mathcal{L}^I . Normative entailment is the relation that holds between institutional facts, i.e., holds between the contents of norms that are in force. In chapter 4, we had a detailed philosophical discussion to defend the view that the relation of norm consequence was, at least on Searle's view, strong implication. In that discussion, however, the language that the consequence relation was defined over was assumed to be a natural language. Then in section 6 we introduced a formal language to characterize strong implication, but that was for a merely propositional language. In the current chapter we have introduced a formal language in which we can formulate norms, as we have seen in the previous section, and so we can, using

definition 5.2.4, define a formal relation for normative consequence.

Let's recall how we characterized norm consequence in chapter 4. There we said

The relation \vdash_S represents norm entailment if and only if for all Γ, φ , that are under institutional control, $\Gamma \vdash_S \varphi$ iff $\mathbf{D}[\Gamma] \models_I \mathbf{D}(\varphi)$.

What we went on to argue was that \vdash_S is the relation of strong implication. The formulas of \mathcal{L}^I represent possible contents of norms that are declared by authorities. Thus codes, strictly speaking, are sets of formulas from \mathcal{L}^I . And as we said in the informal characterization of norm consequence, Γ and φ must be under institutional control. So to have a total characterization we also have to represent in a formal manner the idea of a formal sentence being under institutional control. Obviously atomic sentences from \mathbf{At}_B , i.e., the brute sentences, are not under institutional control. But clearly members of \mathbf{At}_I are. But there are cases where sentences can combine atoms from both the brute and institutional primitive vocabulary. For instance, when $\mathbf{p} \supset V$, i.e., classifying \mathbf{p} -states as violation states.

We characterize the set of formulas that is under institutional control via a recursive definition as follows:

Definition 6.5.1. The *institutional control* function $Ic : \mathcal{L} \rightarrow \{0, 1\}$ is defined recursively as follows:

1. $Ic(\mathbf{s}) = 1$, iff $\mathbf{s} \in \mathbf{At}_I \cup \{V\}$,
2. $Ic(\perp) = 0$
3. $Ic(\neg\varphi) = 1$ iff $Ic(\varphi) = 0$.
4. $Ic(\varphi \wedge \psi) = 1$ iff $Ic(\varphi) = 1$ and $Ic(\psi) = 1$
5. $Ic(\varphi \vee \psi) = 1$ iff $Ic(\varphi) = 1$ or $Ic(\psi) = 1$
6. $Ic(\varphi \supset \psi) = 1$ iff $Ic(\varphi) = 0$ or $Ic(\psi) = 1$
7. $Ic(\varphi \equiv \psi) = 1$ iff $Ic(\varphi) = 1$ and $Ic(\psi) = 1$
8. $Ic([\mathbf{R} \text{ xstit}] \varphi) = 1$ for $\mathbf{R} \subseteq \mathbf{Rol}$

$$9. Ic(\Box\varphi) = 1 \text{ iff } Ic(\varphi) = 1$$

$$10. Ic(X\varphi) = 1 \text{ iff } Ic(\varphi) = 1$$

This function represents the set of formulas of \mathcal{L}^I that are under institutional control as follows:

Definition 6.5.2. $\varphi \in IC(\mathcal{L}^I)$ iff $Ic(\varphi) = 1$.

We can explain our rationale for the recursive clauses as follows. The right propositional content conditions for status function declarations, e.g., promulgations, is that the truth of the content be *manipulable* by the authority making the declaration. But by ‘manipulable’ we mean that the truth (or falsity) of sentence φ depends on whether some of its content has been introduced into the institutional vocabulary. Without being introduced that sentence isn’t comprehensible within the institution.

Let’s first consider the atoms. Clearly, it is the atoms in \mathbf{At}_I that are controlled by the authority. Now it isn’t that the truth or falsity depends on whether an authority says the atom is true or false. But the authority lays down the conditions under which that atom can be true or false, the authority connects the atom to the world. Whether Frank is the legal guardian of Jesse may be the case because of a purely biological accident: Frank is Jesse’s father. But without the legal authority stipulating that ‘legal guardian’ applies when there is a relationship of biological fatherhood is up to the authority. But whether there is biological fatherhood isn’t up to the authority. So atoms in \mathbf{At}_B are not under institutional control.

For the Boolean cases, we must consider both the truth conditions and the falsity conditions. The sentence $\neg\varphi$ is going to be under institutional control when φ is. For $\varphi \wedge \psi$, we see that if one conjunct wasn’t under institutional control, then the truth of the sentence wouldn’t be under institutional control, although the falsity of the sentence would be. Of course we have the dual situation in the case of $\varphi \vee \psi$, if only one of the disjuncts is under institutional control, then the truth of the sentence is under institutional control, but the falsity of the sentence isn’t.

For the sentence $\varphi \supset \psi$, we are in a similar situation as with atoms. These institutions

function by introducing ways of classifying previously existing entities (i.e., states) under newly introduced vocabulary, and specifying deontologies. By having $\varphi \supset \psi$ under institutional control iff ψ is, conditionals can be used by authorities to introduce new classifications onto pre-existing vocabulary. But they can't introduce new vocabulary onto old, think of trying to classify money as a mineral, it just wouldn't work. For a similar reason, $\varphi \equiv \psi$ will be under institutional control when either equivalent is.

What about sentences of the form $X\varphi$, $P\varphi$ and $\Box\varphi$? Here, it is a mode of truth, whether φ is true in the next stage, or true relative to every history at a static state. So these sentences will be under institutional control when φ is. Finally, we have $[\mathbf{r} \text{ xstit}] \varphi$ (or $[\mathbf{R} \text{ xstit}] \varphi$ where $\mathbf{R} \subseteq \mathbf{Rol}$). Does this sentence being under institutional control depend on whether φ is? Surprisingly the answer is no. The reason is, institutional authorities have complete control. Whether \mathbf{R} has certain abilities depends on what other abilities that the authority gives that set of roles. This means that whether $\Diamond [\mathbf{R} \text{ xstit}] \mathbf{p}$ for some $\mathbf{p} \in \mathbf{At}_B$ is true, is under the institutional authority's control. It might be objected that an authority is not able to give a baby the power to see to it that a 1000kg of cement is lifted. Indeed, but the authority could reassign whatever role was given to the baby to something or someone that can see to that task. Or consider an example where some role \mathbf{r} is under an obligation to cut down a tree. But suppose \mathbf{r} can order \mathbf{r}' to cut down the tree, and \mathbf{r}' does so. If that happens the authority can count \mathbf{r}' 's cutting the tree down as \mathbf{r} seeing-to-it-that the tree is cut down. And that latter sentence should be under institutional control. But if whether $[\mathbf{r} \text{ xstit}] \varphi$ is under institutional control depended on φ being under institutional control, that sentence wouldn't be under institutional control. Thus $[\mathbf{r} \text{ xstit}] \varphi$ (or $[\mathbf{R} \text{ xstit}] \varphi$) is under institutional control regardless of whether φ is.

The set $IC(\mathcal{L}^I)$ is defined by Ic , but we can notice another property that this set has. Since \mathbf{At}_I is finite, and Ic is defined recursively, whether a sentence in \mathcal{L}^I is a member of $IC(\mathcal{L}^I)$ is decidable. This is very easy to see since each formula is finite, and there are only finitely many steps to go through to get to atoms, and only finitely many atoms in $at(\varphi)$ and only finitely many elements of \mathbf{At}_I to check any $\mathbf{p} \in at(\varphi)$ against. Thus there are only finitely many steps

to check whether $Ic(\varphi) = 1$. So now we have a formal characterization of institutional control for \mathcal{L}^I .

We can informally characterize normative entailment as follows: $\Gamma \vdash_N \varphi$ iff 1) $\Gamma \vdash_S \varphi$, and 2) $\Gamma; \varphi \subseteq IC(\mathcal{L}^I)$. All that remains is extending strong implication from definition 5.2.4 to the language \mathcal{L}^I .

As we mentioned at the end of section 5.2, we allowed strong implication to hold between a set of sentences that might contain sentences of the form $\theta \in \theta'$. Strictly speaking a sentence about propositional containment isn't the content of a declaration. However, in a certain sense, sentences about containment are under institutional control. Of course some aren't as well. The point is that there are different levels of normative entailment, each specific to a set of background facts.

Indeed, we see this kind of specificity manifested in the law. We have argued that the law can't decide what is true scientifically speaking, but there is a kind of background scientific framework that is in use when making legal judgments. A theory of norm consequence may represent this level of specificity, but it could be argued that normative consequences derived in relation to a particular background theory aren't the strictly "logical" consequences.

Finally, we come to give a definition of norm consequence or norm entailment, its converse.

Definition 6.5.3 (Ξ -Relative Norm Consequence). Let Ω be a code, i.e., $\Omega \subseteq IC(\mathcal{L}^I)$, and Ξ a set of formulas from $\mathcal{L}_{\subseteq}^I$, then

$\Omega \Xi$ -Norm Entails φ ($\Omega \vdash_N^{\Xi} \varphi$) iff there are $\delta_1, \dots, \delta_n \in \Omega \cup \Xi$, $\psi_1, \dots, \psi_k \in \Xi$, and $\psi'_1, \dots, \psi'_m \in \Omega$ s.t.

$$\Xi\text{-NC1: } \vdash_{\text{Exp}} \delta_1 \wedge \dots \wedge \delta_n \supset \varphi,$$

$$\Xi\text{-NC2: } \vdash_{\text{Exp}} (\psi_1 \wedge \dots \wedge \psi_k) \supset (\varphi \in (\psi'_1 \wedge \dots \wedge \psi'_m)), \text{ and}$$

$$\Xi\text{-NC3: } \varphi \in IC(\mathcal{L}^I).$$

So we can state this more succinctly as follows:

$$\Xi\text{-NC1: } \Omega \cup \Xi \vdash_{\text{Exp}} \varphi$$

Ξ -NC2: $\Omega \cup \Xi \vdash_{\text{Ixp}} \varphi \in (\psi'_1 \wedge \dots \wedge \psi'_m)$ for some $\psi'_1 \dots \psi'_m \in \Omega$, and

Ξ -NC3: $\varphi \in IC(\mathcal{L}^I)$.

So the definition says that φ is a normative consequence of Ω when it is a logical consequence, in the language of \mathcal{L}^I_{Ξ} , of Ω along with other background facts from Ξ , and the content of φ is contained in the content of Ω , possibly given other facts about propositional containment in Ξ . But it is also required that φ be under institutional control. Thus we can see how this connects to the conditions on promulgations discussed in section 4.4.

This definition might be called ‘ Ξ ence⁶ relative norm consequence’. Intuitively as well, we would restrict Ξ to contain only facts about nature, i.e., brute facts, and positional containment (\in)-sentences. The special case of this, which we will use in the sequel, we will call *Norm Consequence*. Norm consequence is \emptyset -Relative Norm Consequence, and defined as

Definition 6.5.4 (Norm Consequence). Ω *Norm Entails* φ ($\Omega \vdash_N \varphi$) iff there are $\psi'_1, \dots, \psi'_m \in \Omega$ s.t.

NC1: $\Omega \vdash_{\text{Ixp}} \varphi$,

NC2: $\vdash_{\text{Ixp}} \varphi \in (\psi'_1 \wedge \dots \wedge \psi'_m)$, and

NC3: $\varphi \in IC(\mathcal{L}^I)$.

This account of normative consequence is most like the account of illocutionary entailment since its relations of propositional containment will hold regardless of which semantic interpretation is given to the sentences. Of course we would hold the semantic interpretation of \mathcal{L}^I fixed in a certain respect since the elements of \mathbf{At}_I and \mathbf{At}_B would always be interpreted as institutional or brute atomic sentences. This relates to the most heretical position considered in this essay. We have made a notion of logical consequence, i.e., the logic of institutions/normative consequence, dependent on content and not strictly on form. We think, however, that the notion of content used is misnamed. We are, after all, still working with a formal language. Any

⁶Pronounced ‘Xi-ence’.

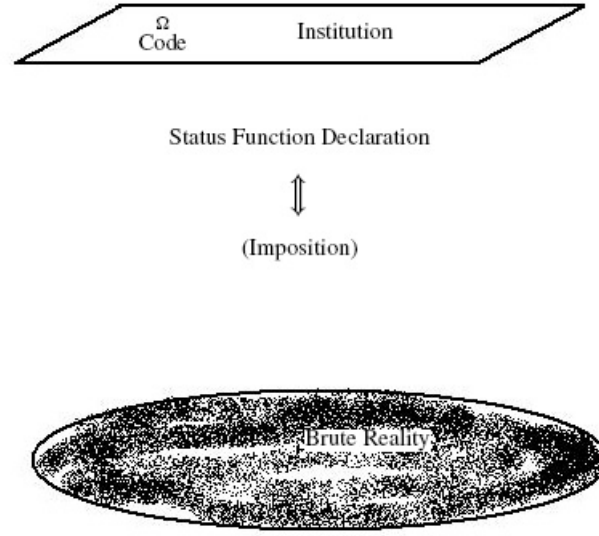


Figure 6.1: The Relation Between Institutions and Reality

consequence relation that we construct will be formal in the sense that the interpretations of the basic atoms in \mathbf{At}_B and \mathbf{At}_I are left unspecified. So that charge won't worry us.

6.6 Institutions and “The World”

Social reality is, on Searle's model, independent from brute reality. We can picture this as in figure 6.1. The things that make the brute language true are not the same as those which make the institutional language true; that is why we offered distinct basic languages to represent atomic brute and institutional facts. It is also why we have roles (**Rol**) as distinct from agents (**Ag**). But the connection between the brute and institutional reality is given in the form of the status function declaration. The institutional deontology is imposed on the brute reality when agents are assigned roles, and the various *count as* conditionals are brought into force.

In this section we want to model, formally, the structure of this imposition. We are not looking to model the mechanisms by which social reality is imposed or sustained, just what happens after the institution is imposed. Our formal model of this imposition is what we will call an *implementation*.

In our logic of institutions we have two languages, one of which is used to represent the brute world, and the other represents the social world. These languages are \mathcal{L}^B and \mathcal{L}^I , respectively. The language \mathcal{L}^I allows us to represent the combinations of the two realities, and \vdash_{Exp} provides us with its logic. We required this separation partly because it makes better sense of Searle's account of institutions, but, second, it will also allow us to model the fact that one norm applies to many agents.

Recall that \mathcal{L}^I contains \mathbf{At}_B and \mathbf{At}_I , whereas \mathcal{L}^B only contains \mathbf{At}_B . Also, \mathcal{L}^I only has role terms, whereas \mathcal{L}^B has agent terms. As per our notion of norm consequence, i.e., definition 6.5.4, a code Ω is completely contained in $IC(\mathcal{L}^I)$ thus it doesn't contain any agent terms in its representations. An *implementation* is an assignment of agent terms to roles in the sentences of Ω . Recall that a code Ω is our formal representation of an institution.

We will also provide a way to connect a code to a model of the brute world, i.e., an \mathcal{L}^B -model. Models of \mathcal{L}^B are representations of ways the world might be independently of an institution. And if Searle is right that there is independence between the social and brute reality, then the range of possibilities of the brute world is all there is to the total range of possibilities. Of course this seems wrong because there are many ways an institution might be imposed on the world. Once we have the notion of an implementation we can resolve this putative conflict. So now we come to our definition.

Definition 6.6.1. Let Ω be a code. An *implementation* of Ω is a triple $\mathfrak{I} = \langle \text{holds}, \pi, \mathfrak{I}(\Omega) \rangle$ such that

1. $\text{holds} \subseteq \mathcal{P}(\mathbf{Ag}) \times \mathbf{Rol}$, such that for each $\mathbf{a} \in \mathbf{Ag}$ there is $\mathbf{A} \subseteq \mathbf{Ag}$ and $\mathbf{r} \in \mathbf{Rol}$ such that $\mathbf{a} \in \mathbf{A}$ and $\langle \mathbf{A}, \mathbf{r} \rangle \in \text{holds}$.
2. π is a partition of \mathbf{Rol} , where $\pi_{\mathbf{r}}$ is the cell of the partition containing \mathbf{r} , such that
3. if $\langle \mathbf{A}, \mathbf{r} \rangle \in \text{holds}$, then for all $\mathbf{r}' \in \pi_{\mathbf{r}}$, $\langle \mathbf{A}, \mathbf{r}' \rangle \in \text{holds}$, and
4. $\mathfrak{I}(\Omega) \subseteq \mathcal{L}^B_{\Omega}$ such that each $\delta \in \mathfrak{I}(\Omega)$ is a substitution instance of some $\varphi \in \Omega$, where each role term \mathbf{r} mentioned in φ is replaced uniformly in φ by an agent term \mathbf{A} such that

$\langle \mathbf{A}, \mathbf{r} \rangle \in \text{holds}$.

We will now explain the conditions in definition 6.6.1. holds is a relation between agent terms and singular role terms. As we said above, an implementation is a specification of who plays what roles, and that is exactly what holds does. But notice that since holds is a relation, multiple agents (singular or plural) can hold a single role, and vice versa.

The partition of \mathbf{Rol} by π provides what we called a *general role* in section 5.1.2. The rationale behind the general role is that since we don't have quantifiers, we can still represent multiple types of the same role in one formula. Why is this important? In the formula $[\mathbf{r} \text{ xstit}] \varphi \supset ([\mathbf{r} \text{ xstit}] \psi \supset V)$, we have to interpret r by the same agent, in formula $[\mathbf{r} \text{ xstit}] \varphi \supset ([\mathbf{r}' \text{ xstit}] \psi \supset V)$, \mathbf{r} and \mathbf{r}' can be interpreted by different agents. But when we want to interpret natural language expressions like 'citizens must not see to it that other citizens see to fires' we have to represent it with something like $\Box(\neg [\mathbf{r} \text{ xstit}] \neg [\mathbf{r}' \text{ xstit}] \varphi \supset [\mathbf{r} \text{ xstit}] V)$ as per our discussion in section 6.4. We can then interpret \mathbf{r} and \mathbf{r}' both as citizens when $\mathbf{r}, \mathbf{r}' \in \pi_{\text{Citizen}}$. The cells in π have another role to play in connecting \mathcal{L}^I to \mathcal{L}^B .

Condition 3 specifies that if $\mathbf{r}, \mathbf{r}' \in \pi_{\mathbf{r}''}$, then any agent that holds \mathbf{r} must also hold \mathbf{r}' (and \mathbf{r}'' as well). This makes sense since if, say, \mathbf{A} is a citizen, then \mathbf{A} must have all of the responsibilities, powers, and privileges that any citizen has simply in virtue of being a citizen.

As we said, the implementation is to bridge the gap between the institutional language and the brute language. So far holds provides a way to connect the institutional roles to agents, but we need to do more. To fully assign the deontology of an institution to agents it must be the relevant agents that hold the roles are also given the powers, and bear the burden of the responsibilities that come with those roles. In our language, we represent that \mathbf{r} has a duty to φ , as per section 6.4, by $\Box(\neg [\mathbf{r} \text{ xstit}] \varphi \supset [\mathbf{r} \text{ xstit}] V)$. To impose the institutional fact that \mathbf{A} now fulfills the role \mathbf{r} under this implementation, it must be that $\Box(\neg [\mathbf{A} \text{ xstit}] \varphi \supset [\mathbf{A} \text{ xstit}] V)$ is true. Indeed, that must be the case for each member of Ω , \mathbf{r} and \mathbf{A} . So we take all of the substitution instances of the formulas ψ in Ω , where if $\langle \mathbf{A}, \mathbf{r} \rangle \in \text{holds}$, then $\mathfrak{F}(\varphi)$ replaces \mathbf{r} with \mathbf{A} uniformly in ψ .

So for instance say $\{ \Box(\neg [\mathbf{r} \text{ xstit}] \neg [\mathbf{r}' \text{ xstit}] \varphi \supset [\mathbf{r} \text{ xstit}] V) \} = \Omega$, and say that the implementation \mathfrak{F} is given by

1. $\text{holds} = \{ \langle \mathbf{A}, \mathbf{r} \rangle, \langle \mathbf{B}, \mathbf{r} \rangle, \langle \mathbf{B}, \mathbf{r}' \rangle \}$ and
2. $\pi = \{ \{ \mathbf{r} \}, \{ \mathbf{r}' \} \}$.

For $\mathbf{Ag} = \{ \mathbf{a}, \mathbf{b} \}$ and $\mathbf{A} = \{ \mathbf{a} \}$ and $\mathbf{B} = \{ \mathbf{b} \}$. Then $\mathfrak{F}(\Omega)$ must contain at least the following formulas:

$$\Box(\neg [\mathbf{A} \text{ xstit}] \neg [\mathbf{B} \text{ xstit}] \varphi \supset [\mathbf{A} \text{ xstit}] V)$$

$$\Box(\neg [\mathbf{A} \text{ xstit}] \neg [\mathbf{B} \text{ xstit}] \varphi \supset [\mathbf{B} \text{ xstit}] V)$$

$$\Box(\neg [\mathbf{B} \text{ xstit}] \neg [\mathbf{B} \text{ xstit}] \varphi \supset [\mathbf{B} \text{ xstit}] V)$$

Just one more point about the relationship between the general roles and $\mathfrak{F}(\Omega)$. Condition 3 also says that anywhere that an agent term \mathbf{A} can appear for a role \mathbf{r} , any other role term in that general role \mathbf{A} can be, and must be substituted there as well. Finally, we should also notice that condition 1 requires that each agent hold *some* role, or be part of some role. This assumption is for technical reasons, but it has an intuitive grounding as well. The implementation of a code specifies who plays which roles in an institution. If someone is accorded absolutely no role, then it is reasonable to assume that agent isn't part of that institution. So from a technical standpoint that agent can simply be removed from \mathbf{Ag} without any loss since none of the institutional facts applies to that agent.

The implementation allows us to bridge the gap between \mathcal{L}^I and \mathcal{L}^B . The set $\mathfrak{F}(\Omega)$ represents what has to be true for the institutional norms to be in force relative to a particular imposition. But it doesn't represent a model of "the world", in which the code Ω is in force. For that we need an xstit model, but not just any xstit model.

Our argument in section 4.4 for why normative entailment was strong implication turned, in the end, on the fact that institutional facts that don't exist relative to a particular society shouldn't be included as normative consequences of declarations of authorities in that society, even when

propositions including those institutional facts are classical consequences of declarations. That is enshrined in condition NC2 of definition 6.5.4. So if an \mathcal{L}^B -model is a model of “the world”, then it shouldn’t interpret any institutional language that isn’t in Ω . Thus, we need a model of the language \mathcal{L}_Ω^B , to represent a world in which Ω is in force.

But again we want to maintain the Searlean⁷ point that the brute world is all that is or could possibly be the case. More importantly, an institution can be imposed on the brute world. To reflect these points we define models of \mathcal{L}_Ω^B as expansions of \mathcal{L}^B -models. Of course only some expansions will be appropriate, i.e., only some expansions will be models of the code Ω . Let’s recall the definition of a model for \mathcal{L}^B .

An \mathcal{L}^B model \mathfrak{M} is a regular, universal model $\langle S, H, E, v \rangle$ where

1. $S \neq \emptyset$, are the static states.
2. $H \neq \emptyset$ is a set of orders $\langle h, <_h \rangle$ such that for each $h \in H$
 - (a) $h \subseteq S$ and $\langle h, <_h \rangle$ is isomorphic to \mathbb{Z} with its usual order, and
 - (b) if $s \in h \cap h'$, then $\{s' : s' <_h s\} = \{s' : s' <_{h'} s\}$. Since each order is isomorphic with \mathbb{Z} , there is a unique successor and predecessor for each $s \in h$, we refer to these by $\text{lub}(s, h)$ and $\text{glb}(s, h)$ respectively.
3. $E : S \times H \times \mathcal{P}(\mathbf{Ag}) \rightarrow \mathcal{P}(S)$, the h -effectivity function, assigns a set of static states to each triple (s, h, \mathbf{A}) . It must obey the following conditions:
 - (a) if $s \notin h$, then $E(s, h, \mathbf{A}) = \emptyset$
 - (b) if $s' \in E(s, h, \mathbf{A})$, then $s' \in \text{lub}(s)$
 - (c) if $s \in h$, $\text{lub}(s, h) \in E(s, h, \mathbf{A})$
 - (d) $E(s, h, \emptyset) = \text{lub}(s)$, if $s \in h$
 - (e) if $s \in h$, then $E(s, h, \mathbf{Ag}) = \{\text{lub}(s, h)\}$

⁷Although it could be a Wittgensteinian point too.

- (f) if $\mathbf{A} \subsetneq \mathbf{B}$, then $E(s, h, \mathbf{B}) \subseteq E(s, h, \mathbf{A})$
 - (g) if $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $s \in h \cap h'$, then there is h'' with $s \in h''$ and $E(s, h'', \mathbf{A})$ and $E(s, h'', \mathbf{B})$ are contained in $E(s, h, \mathbf{A})$ and $E(s, h', \mathbf{B})$, respectively.
4. $v : \mathbf{At}_B \rightarrow \mathcal{P}(S)$. And interprets the language \mathcal{L}^B as follows:

v gives rise to a truth relation \models as follows:

1. $(s, h) \models \mathbf{p} \in \mathbf{At}_B$ iff $s \in v(\mathbf{p})$
2. $(s, h) \models \neg\varphi$ iff $(s, h) \not\models \varphi$
3. $(s, h) \models \varphi \wedge \psi$ iff $(s, h) \models \varphi$ and $(s, h) \models \psi$
4. $(s, h) \models \varphi \vee \psi$ iff $(s, h) \models \varphi$ or $(s, h) \models \psi$
5. $(s, h) \models \varphi \supset \psi$ iff $(s, h) \not\models \varphi$ or $(s, h) \models \psi$
6. $(s, h) \models \Box\theta$ iff for all h' with $s \in h'$, $(s, h') \models \theta$
7. $(s, h) \models X\theta$ iff for all $(lub(s, h), h) \models \theta$
8. $(s, h) \models P\theta$ iff for all $(glb(s, h), h) \models \theta$
9. $(s, h) \models [\mathbf{A} \text{ xstit}] \theta$ iff for all s', h' , if $s' \in E(s, h, \mathbf{A})$ and $s' \in h'$, then $(s', h') \models \theta$

To extend \mathfrak{M} to a model of \mathcal{L}_Ω^B , \mathfrak{M}_Ω , v is extended to a function v_Ω from $\mathbf{At}_B \cup at(\Omega)$ to $\mathcal{P}(S)$ such that if $\mathbf{p} \in \mathbf{At}_B$, then $v_\Omega(\mathbf{p}) = v(\mathbf{p})$. We will say that a particular \mathcal{L}_Ω^B -model \mathfrak{M}_Ω realizes Ω when it models an implementation of Ω . Put formally,

Definition 6.6.2. Let Ω be a code, and \mathfrak{I} an implementation of Ω . Let \mathfrak{M} be an \mathcal{L}^B model, \mathfrak{M}_Ω an extension of \mathfrak{M} to a model of \mathcal{L}_Ω^B , and $(s, h) \in |\mathfrak{M}_\Omega| = |\mathfrak{M}|$. Then $\mathfrak{M}_\Omega, (s, h)$ realizes Ω relative to \mathfrak{I} iff $\mathfrak{M}_\Omega, (s, h) \models \mathfrak{I}(\Omega)$.

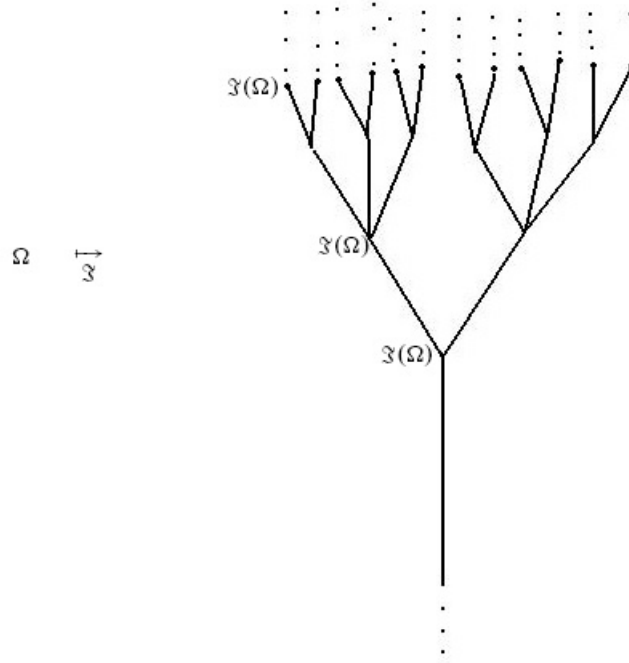


Figure 6.2: Realization of Implementation

So the relation of ‘realization’ is triadic between a model, a code and an implementation. However, our notions of implementation and realization (relative to an implementation) allow us to talk about codes, i.e., subsets of $IC(\mathcal{L}^I)$, as subsets of \mathcal{L}_Ω^B which are satisfied (in the usual logical sense) in models of the “real” world—modulo some institutional facts. So we have provided a bridge from the institutional reality represented in Ω to its imposition onto a brute reality through the realization of implementations of those codes. We picture a realization of an implementation of Ω in figure 6.2.

But there is a pressing question. Since Ω is a subset of $IC(\mathcal{L}^I)$, we take it to represent the set of explicitly promulgated norms of some authority. The implicitly promulgated norms are the norm consequences of Ω , i.e., the implicit norms are $\mathbb{C}_N(\Omega) = \{\varphi \in \mathcal{L}^I : \Omega \vdash_N \varphi\}$. But what, if any, relationship is there between implicit norms, and substitution instances of those implicit norms in an implementation? The worrisome scenario is if $\Omega \vdash_N \varphi$, but there was a realization of Ω that didn’t realize φ . Fortunately that can’t happen, and that is shown in the

following observation (see section 7.4 for the proof).

Proposition 6.6.1. *For any implementation \mathfrak{I} , if $\Omega \vdash_N \varphi$, $\Omega \not\vdash_{\text{Exp}} \perp$ and $\varphi \in \mathcal{L}^I$, then $\mathfrak{I}(\Omega) \vdash_{\text{xp}}^{\Omega} \delta$ for all $\delta \in \mathfrak{I}(\varphi)$.*

So not only does every consequence of Ω get reflected, but every instantiation under an implementation gets reflected in any realization of an implementation. This is very good news.

6.7 Responses to Some Objections

We will pause to respond to some objections in the literature. Sven Ove Hansson has levelled a couple of criticisms on our way of expressing legal relations like power. The first is that the *legal* relation of having a power should not be confused with the idea of a power in general, i.e., a physical power. Hansson's objection can be interpreted as claiming that stit-like proposals mistakes physical power for legal power. In this system there is no such confusion. Agents have physical powers, while institutional roles have legal (or institutional) powers *as well*. Agents have institutional powers in virtue of holding a particular role, i.e., that role being assigned to the agent via an implementation. But a general may have the power to destroy a city simply because of their institutional power of command of the military. Legal power and physical power are not confused.

Second, Kanger and Kanger (1966) express a legal power as the permission to see to a proposition. This again, claims Hansson, mistakes the power to do something with permission to do that something, i.e., it confuses permission with ability. The current system doesn't make that mistake. Power is the *possibility* of seeing to something, not the permission. And from the previous paragraph, we do not conflate institutional power with physical power.

The next issue we would like to address is a common worry to do with Anderson-like reductions of deontic logic.⁸ In an Ixstit model, when $(s, h) \models \Box XV$, then $(s', h') \models V$ for any h' with $s \in h'$ and $\text{lub}(s, h') = s'$. That means $(s, h) \models \Box [\mathbf{R} \text{xstit}] V$ for any $\mathbf{R} \subseteq \mathbf{Rol}$.

⁸A form of this objection is in one of Anderson's papers on the topic, cf. Anderson (1967).

Thus, when $(s, h) \models \Box XV$, then for any φ , $(s, h) \models \Box([\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$ as well as $(s, h) \models \Box(\neg[\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$. So when there is a violation in all the next states *everything is a duty and prohibited!*

This is not the case. Indeed, when $\Box XV$ is true, all duty and prohibition sentences are true, but that does not a duty or prohibition make. For there to be a prohibition the sentence $\Box([\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$ must *follow* from the code, i.e., $\Omega \vdash_N \Box([\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V)$. Mere truths in a model don't, necessarily, mean anything. Also, once a code has been implemented, there are no longer roles expressed in the language. An implementation takes Ω from \mathcal{L}^I to \mathcal{L}_{Ω}^B , and the latter language doesn't contain role terms, only agent terms. Thus duties for an agent are given in virtue of an implementation of a code on an \mathcal{L}_{Ω}^B -model. Again, what happens in *one* particular model isn't an issue. Indeed, how we interpret *obligations* relative to an institution is distinct from institutional duties, and is treated briefly in section 9.1.2.

One final objection has to do with \vdash_N .⁹ As we have defined it, if $\Omega \vdash_N \varphi$, then $\varphi \in Ic(\mathcal{L}^I)$. But this implies that although $\varphi \supset \psi \in Ic(\mathcal{L}^I)$, and $\vdash_{\text{Exp}} \neg\psi \supset \neg\varphi \in \varphi \supset \psi$, and $\varphi \supset \psi \vdash_{\text{Exp}} \neg\psi \supset \neg\varphi$, it is not the case that $\varphi \supset \psi \vdash_N \neg\psi \supset \neg\varphi$, since $\neg\psi \supset \neg\varphi$ is not under institutional control when φ isn't. Contraposition fails for norm consequence. But this isn't a problem for our system.

Indeed, if φ isn't under institutional control, then an institutional authority doesn't want to say that non- ψ states are to be classified as non- φ states. As we discussed in our explanation of Ic , allowing $\theta \supset \theta'$ to be under institutional control when θ' isn't supposes that non-institutional vocabulary can be hijacked by institutional vocabulary. But it will be true that a non- ψ state is a non- φ states when φ -states are already classified as ψ -states in the institution. So what is expressible by $\neg\psi \supset \neg\varphi$ is expressible by a sentence that is under institutional control, viz. $\varphi \supset \psi$. Any conclusion under institutional control that is \vdash_{Exp} -derivable from $\neg\psi \supset \neg\varphi$, when $\Omega \vdash_{\text{Exp}} \neg\psi \supset \neg\varphi$, and whose content is contained in Ω will be a norm consequence of Ω .

⁹Thank you to Allen Hazen for this criticism.

Chapter 7

Completeness of \vdash_{Ixp}

“I’m a metalogician”, Bron said. “I define and redefine the relation between P and Not-P five hours a day, four days a week.”

From Trouble on Triton, by Samuel Delany, 1976.

This chapter acts as an appendix to chapter 6. In this chapter we prove various important facts about the logical systems presented in the previous chapter. Our goal is to prove that \vdash_{Ixp} is complete with respect to \models_{Ixp} . To accomplish that, as we mentioned in the last chapter, we first prove completeness of \vdash_{xp} with respect to the frames/models from definition 5.1.1. We then show that if condition g on effectivity functions fails, then we can invalidate Indep-G. This provides a completeness proof for \vdash_{xp}^I . Then we argue that we can combine that result with the one from section 5.3 for SI-validity. Those will give us a completeness proof for \vdash_{Ixp} .

7.1 Completeness of \vdash_{xp}

The proof of completeness for \vdash_{xp} is rather roundabout. Recall that in definition 5.1.1 we called that kind of model a universal, regular frame. A universal, regular frame for \mathcal{L} has the special properties that each of its histories look like \mathbb{Z} , and that when $s \in h \cap h'$, then all of the past static states are shared, i.e., no backwards branching.

It became evident that the usual way of proving completeness didn’t function with respect to regular models. The problem is that the usual canonical model construction doesn’t generate a regular model. In the canonical model construction we take the maximally consistent sets—which we call maxi sets—as the static states, and the histories to be lines of maxi sets. But since the successor and predecessors in those lines must be unique there can be no branching. But that means the canonical model doesn’t provide enough countermodels for non-theorems.

But it didn’t seem that the canonical model could easily be made into a regular model either.

The “canonical” way of imposing a branching structure on the canonical model is to take the static states to be equivalence classes of maxi sets. But doing that would cause states to loop which would disrupt the \mathbb{Z} -like structure of histories in regular models. So what we decided to do was provide a completeness proof in the usual way to see what kind of models the language \mathcal{L} is directly talking about. These irregular models for \mathcal{L} we call *neutral* models after Thomason (1984). We show that completeness holds for the irregular models, but we can also show that regular models can be generated from the irregular models so that completeness holds for the regular models as well. We will start with a review of regular models.

7.1.1 Regular Models of \mathcal{L}

We will recall the definitions of a regular model here, as well as the semantics, for clarity. An \mathcal{L} -model $\mathfrak{M} = \langle S, H, E, v \rangle$ is a regular model when

1. $S \neq \emptyset$, are the static states.
2. $H \neq \emptyset$ is a set of orders $\langle h, <_h \rangle$ such that for each $h \in H$
 - (a) $h \subseteq S$ and $\langle h, <_h \rangle$ is isomorphic to \mathbb{Z} with its usual order, and
 - (b) if $s \in h \cap h'$, then $\{s' : s' <_h s\} = \{s' : s' <_{h'} s\}$. Since each order is isomorphic with \mathbb{Z} , there is a unique successor and predecessor for each $s \in h$, we refer to these by $\text{lub}(s, h)$ and $\text{glb}(s, h)$ respectively. We can generalize these concepts in the following way: $\text{glb}(s) = \{s' : \exists h \text{ glb}(s, h) = s'\}$ and $\text{lub}(s) = \{s' : \exists h \text{ lub}(s, h) = s'\}$. These give the set of successors and predecessors of s , respectively.
3. $E : S \times H \times \mathcal{P}(\mathbf{Ag}) \rightarrow \mathcal{P}(S)$ is called an h -effectivity function. The effectivity function provides a set of states that a group of agents is effective in ensuring from a given state s , relative to a history h . The function E must obey the following conditions:
 - (a) if $s \notin h$, then $E(s, h, \mathbf{A}) = \emptyset$

- (b) if $s' \in E(s, h, \mathbf{A})$, then $s' \in \text{lub}(s)$
- (c) if $s \in h$, $\text{lub}(s, h) \in E(s, h, \mathbf{A})$
- (d) $E(s, h, \emptyset) = \text{lub}(s)$, if $s \in h$
- (e) if $s \in h$, then $E(s, h, \mathbf{Ag}) = \{ \text{lub}(s, h) \}$
- (f) if $\mathbf{A} \subsetneq \mathbf{B}$, then $E(s, h, \mathbf{B}) \subseteq E(s, h, \mathbf{A})$
- (g) if $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $s \in h \cap h'$, then there is h'' with $s \in h''$ and $E(s, h'', \mathbf{A})$ and $E(s, h'', \mathbf{B})$ are contained in $E(s, h, \mathbf{A})$ and $E(s, h', \mathbf{B})$, respectively.

And v is a function $\mathbf{At} \rightarrow \mathcal{P}(S)$. We interpret the language \mathcal{L} as follows:

1. $(s, h) \models \mathbf{p} \in \mathbf{At}$ iff $s \in v(\mathbf{p})$
2. $(s, h) \models \neg\varphi$ iff $(s, h) \not\models \varphi$
3. $(s, h) \models \varphi \wedge \psi$ iff $(s, h) \models \varphi$ and $(s, h) \models \psi$
4. $(s, h) \models \varphi \vee \psi$ iff $(s, h) \models \varphi$ or $(s, h) \models \psi$
5. $(s, h) \models \varphi \supset \psi$ iff $(s, h) \not\models \varphi$ or $(s, h) \models \psi$
6. $(s, h) \models \Box\theta$ iff for all h' with $s \in h'$, $(s, h') \models \theta$
7. $(s, h) \models X\theta$ iff for all $(\text{lub}(s, h), h) \models \theta$
8. $(s, h) \models P\theta$ iff for all $(\text{glb}(s, h), h) \models \theta$
9. $(s, h) \models [\mathbf{A} \text{ xstit}] \theta$ iff for all s', h' , if $s' \in E(s, h, \mathbf{A})$ and $s' \in h'$, then $(s', h') \models \theta$

After introducing regular frame and models in definition 5.1.1, we said that we would focus on *universal* regular \mathcal{L} -models. The distinction between a universal and a non-universal, regular model can be explained as follows: in a universal regular model $\bigcap H \neq \emptyset$ whereas in a non-universal regular model $\bigcap H = \emptyset$. We use the term ‘regular’ for these models since they are the models that are used most in applications, and so we will use them to define validity for \mathcal{L} and denote entailment relative to these models with \models_{xp} .

Definition 7.1.1. $\Gamma \models_{\text{xp}} \varphi$ iff For any regular \mathcal{L} -model \mathfrak{M} , and $(s, h) \in |\mathfrak{M}|$, $\mathfrak{M}, (s, h) \models \Gamma$ only if $\mathfrak{M}, (s, h) \models \varphi$.

We have to be clear about calling something a regular \mathcal{L} -model since it encourages the thought that there could be irregular \mathcal{L} -models. By ‘irregular’ we do not mean non-universal, we mean that it is a general form of model for \mathcal{L} than the regular models. Indeed, our proof of completeness will depend on these irregular models for \mathcal{L} .

7.1.2 Neutral Models of \mathcal{L}

To define a neural model, we have to define some mathematical structures that have a similar structure to \mathbb{Z} . So we will start with some general notation.

Suppose that $R \subseteq S \times S$ is a function (i.e., For each $s \in S$ there is unique $s' \in S$ such that sRs' , hence we could write sRs' as $R(s) = s'$). A function that is from a set to itself is called an *endomorphism* and one can iterate endomorphisms: $R(R(R(s))) = R^3(s)$, for instance. Whenever an endomorphism satisfies $R(s) = R(s')$ only if $s = s'$ we say that it is injective. Usually we will refer to functions with lower case letters: f, g, b, \dots

If R is a relation on S , then the transitive closure of R , denoted R^+ , is constructed recursively: aR^0b iff aRb , aR^1b iff there is z such that aRz and zRb , $aR^{n+1}b$ iff there is z such that aR^nz and zRb . Then $R^+ = \bigcup_{n=0}^{\infty} R^n$. R^+ is a transitive relation. If f is a function on S , then we can define the transitive closure of the function, denoted f^+ , recursively as $f^1(s) = s'$ iff $f(s) = s'$, and $f^n(s) = s'$ iff there is z such that $f^{n-1}(s) = z$ and $f(z) = s'$. s and s' are related by the transitive closure of f is denoted as $f^+(s) = s'$, i.e., for some $n \in \mathbb{N}$, with $n > 0$ $f^n(s) = s'$. Notice that $f^0(s) = s$.

An *almost injective endomorphism* f is an endomorphism that obeys the following property:

$$\forall s, s' \in S [f(s) \neq s \ \& \ f(s') \neq s'] \Rightarrow [f(s) = f(s') \Rightarrow s = s'] \quad (7.1)$$

This says that as long as arguments are not fixed points of f , f acts injectively on the arguments. It is also important to note that injective endomorphisms also satisfy this property. As

we noted each history in a regular model looks like \mathbb{Z} , and our goal is to generalize this structure to provide an irregular class of models for \mathcal{L} .

Definition 7.1.2. A Discrete Line Function (DLF) on a non-empty set S is an almost injective endomorphism $f : S \rightarrow S$ such that

[TRI] its transitive closure obeys trichotomy: for every $s, s' \in S$: either $f^+(s) = s'$ or $f^+(s') = s$ or $s = s'$, and

[UFIX] if $s, s' \in S$, with $f(s) = s$ and $f(s') = s'$, then $s = s'$.

The first condition TRI says that every two distinct members of S are related by f by some finite distance, i.e., there is $n \in \mathbb{N}$ such that either $f^n(s) = s'$ or $f^n(s') = s$. The UFIX condition states that if there is a fixed point, it is unique. The structures in this class are all manner of objects: finite and infinite lines, loops, lines with loops at a point on the end, reverse trees, among others. But we want to focus on structures like $\langle \mathbb{N}, x + 1 \rangle$ or $\langle \mathbb{Z}, x + 1 \rangle$. We call the class of discrete line functions **DLF**.

To properly mimic the structure of \mathbb{Z} we need to place an inverse-like function to match f on the DLF, which we will call b , which is also a DLF, but in the opposite direction. However, we want to require that it relates to f in a certain way. We will call such a structure a Double DLF (DDLDF).

Definition 7.1.3. A Double DLF is a triple $\langle S, f, b \rangle$ where f and b are both DLFs from $S \neq \emptyset$ to S such that

[CONV1] For any $s \in S$, $\exists s' \neq s$ with $f(s) = s'$ only if $b(f(s)) = s$,

[CONV2] For any $s \in S$, $\exists s' \neq s$ with $b(s) = s'$ only if $f(b(s)) = s$, and

[SIZE] $|S| \geq 2$ only if for any $s \in S$, if $f(s) = s$, then $b(s) \neq s$.

So b is almost the inverse of f . At the end points the functions must stop being inverses. DDLFs of size 1 are unique up to isomorphism, it is not so with DDLFs of larger finite cardinality. Intuitively, DDLFs of finite cardinality are either loops or lines with loops on either end:

$$\mathcal{U} \cdot \leftrightarrow \cdot \leftrightarrow \cdot \leftrightarrow \cdot \leftrightarrow \cdot \leftrightarrow \cdot \mathcal{U}$$

A DDLF of infinite cardinality will look like \mathbb{Z} or like \mathbb{N} with a loop at 0. We will call the class of DDLFs **DDL**F. We notice that when $|S| = 1$, both $f(s) = s = b(s)$ by necessity, so in such a case the SIZE condition would fail, but that is why there is the condition that the cardinality must be at least 2.

We make another observation that allows us to classify these objects.

Observation 7.1.1. *Let $\langle S, f, b \rangle$ be a DDLF, then*

1. *f is injective iff for all $s \in S$, $b(f(s)) = s$*
2. *g is injective iff for all $s \in S$, $f(b(s)) = s$*
3. *f and g are injective iff for all $s \in S$ $f(b(s)) = s = b(f(s))$.*

Proof. We will do the first, since the second is symmetric. (ONLY IF) Suppose that f is injective. If $|S| = 1$, then clearly $b(f(s)) = s$. So assume that $|S| \geq 2$. Now what we will show is that for all $s \in S$, $f(s) \neq s$. Suppose for reductio that $f(s) = s$. Then by SIZE, $b(s) \neq s$, which means by CONV2 that $f(b(s)) = s$. But by injectivity we also have that $f(b(s)) \neq s$, a contradiction. So then by CONV1, it follows that $b(f(s)) = s$.

(IF) Suppose that for all $s \in S$, $b(f(s)) = s$, then suppose that $f(s) = f(s')$. Then by our assumption, $s = b(f(s)) = b(f(s')) = s'$. Hence f is injective.

3 follows from 1 and 2. □

We must also notice something about the structures,

Observation 7.1.2. *If $\langle S, f, b \rangle$ is a DDLF with both f and g injective, then either it is isomorphic to $\langle \mathbb{Z}, x + 1, x - 1 \rangle$, or $\langle \mathbb{Z}_n, x + 1, x - 1 \rangle$ for some $n \in \mathbb{N}$.*

Proof. Clearly, $\langle \mathbb{Z}, x + 1, x - 1 \rangle$ and $\langle \mathbb{Z}_n, x + 1, x - 1 \rangle$ are such DDLF structures. If S is infinite, then pick some $s \in S$, and define $g : \mathbb{Z} \rightarrow S$ such that $g(0) = s$, and $g(n + 1) =$

$f(g(n))$ and $g(n-1) = b(g(n))$. The same method can be used in the finite case. But we just pick $\mathbb{Z}_{|S|}$ as the domain of g . \square

From here on, when we call something an ‘injective DDLF’ we will mean that both f and b are injective. We will refer to the class of injective DDLFs as **IDDLF** and abbreviate ‘injective DDLF’ as IDDLF. From these structures we can define *lub* and *glb* functions. What we had in the regular models was a kind of standard ordering on the histories, i.e., the one from \mathbb{Z} . Here things are a bit more liberal, but we can still interpret the language \mathcal{L} . What we do is make each history h a triple $\langle h, f_h, b_h \rangle$ such that $h \subseteq S$ (S the static domain of the model) which is a DDLF.

Now we define a semantics for \mathcal{L} based on a kind of structure that Thomason (1984) calls a *neutral frame*. In general, a neutral frame is a bunch of independent time streams (histories) whose points are related by an equivalence relation. This relation represents what the modal alternatives are. This model is somewhat different from the original conception of definition 5.1.1 because there alternatives share the past, coincident pasts are *identical*. That may seem more intuitive, in a neutral frame the pasts of alternatives are at most *indiscernible*. What we will get are the following:

Definition 7.1.4. A *neutral \mathcal{L} -frame* is a triple $\mathfrak{F} = \langle S, H, E, \approx \rangle$ such that:

1. $S \neq \emptyset$, are the static states.
2. \approx is an equivalence relation on S
3. $H \neq \emptyset$ is a set of triples $h = \langle h, f_h, b_h \rangle$ with $h \subseteq S$ such that

H1 each $\langle h, f_h, b_h \rangle \in H$ is an injective **DDLF**,

H2 if $s \in h$ and $s' \in h'$ with $s \approx s'$, then for each $n \in \mathbb{N}$, $b_h^n(s) \approx b_{h'}^n(s')$.

4. $|\mathfrak{F}| = \{ (s, h) \in S \times H : s \in h \}$ (the domain of \mathfrak{F})
5. $\text{lub}_{\mathfrak{F}}(s, h) = f_h(s)$ and $\text{glb}_{\mathfrak{F}}(s, h) = b_h(s)$ but now
6. $\text{lub}_{\mathfrak{F}}(s) = \{ s^* \in S : \exists h', s' \text{ s.t. } s \approx s' \ \& \ f_{h'}(s') = s^* \}$, and

7. $glb_{\mathfrak{F}}(s) = \{s^* \in S : \exists h', s' \text{ s.t. } s \approx s' \ \& \ b_{h'}(s') = s^* \}$.

8. $E : S \times H \times \mathcal{P}(\mathbf{Ag}) \rightarrow \mathcal{P}(S)$ is called an h -effectivity function. The effectivity function provides a set of states that, relative to a history h a group of agents is effective in ensuring from a given state s . The function E must obey the following conditions:

- (a) if $s \notin h$, then $E(s, h, \mathbf{A}) = \emptyset$
- (b) if $s' \in E(s, h, \mathbf{A})$, then $s' \in lub(s)$
- (c) if $s \in h$, $lub(s, h) \in E(s, h, \mathbf{A})$
- (d) $E(s, h, \emptyset) = lub(s)$
- (e) if $s \in h$, then $E(s, h, \mathbf{Ag}) = \{s' : s' \approx lub(s, h)\}$
- (f) if $\mathbf{A} \subsetneq \mathbf{B}$, then $E(s, h, \mathbf{B}) \subseteq E(s, h, \mathbf{A})$
- (g) For all \mathbf{A}, \mathbf{B} $(s, h), (s', h'), (s'', h'') \in |\mathfrak{F}|$, if $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $s' \approx s \approx s''$, then there is $(s''', h''') \in |\mathfrak{F}|$ such that $s''' \approx s$ with $E(s''', h''', \mathbf{A}) \subseteq E(s', h', \mathbf{A})$ and $E(s''', h''', \mathbf{B}) \subseteq E(s'', h'', \mathbf{B})$.

We will usually omit the subscript \mathfrak{F} on, inter alia, the lub and glb functions, taking it to be understood. We then define a model as follows:

Definition 7.1.5. A *neutral \mathcal{L} -model* \mathfrak{M} , is a neutral \mathcal{L} -frame \mathfrak{F} with a valuation $v : \mathbf{At} \rightarrow \mathcal{P}(S)$ such that if $s \approx s'$ and $s \in v(\mathbf{p})$, then $s' \in v(\mathbf{p})$.

We can picture the neutral models for \mathcal{L} as in figure 7.1. Notice that \approx can hold between static states in the same history, and so successor states of s .

When we have a neutral \mathcal{L} -model, we can provide an interpretation of \mathcal{L} as follows:

Definition 7.1.6. For formulas in \mathcal{L} and $\mathbf{A} \subseteq \mathbf{Ag}$ and neutral model \mathfrak{M} with $s \in S$ and $h \in H$,

- $(s, h) \models \mathbf{p}$ iff $s \in v(\mathbf{p})$ where $\mathbf{p} \in \mathbf{At}$
- $(s, h) \models \neg\varphi$ iff $(s, h) \not\models \varphi$
- $(s, h) \models \varphi \wedge \psi$ iff $(s, h) \models \varphi$ and $(s, h) \models \psi$

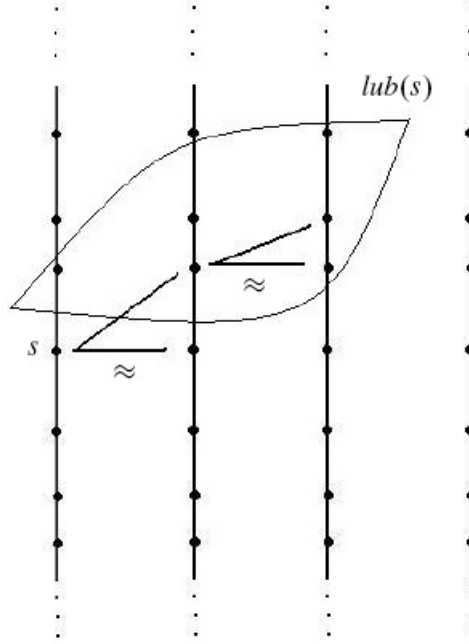


Figure 7.1: $\text{lub}(s)$ in a neutral model

- $(s, h) \models \Box\varphi$ iff for all h' with $s' \in h'$ and $s \approx s'$, $(s', h') \models \varphi$
- $(s, h) \models X\varphi$ iff for all $(\text{lub}(s, h), h) \models \varphi$
- $(s, h) \models P\varphi$ iff for all $(\text{glb}(s, h), h) \models \varphi$
- $(s, h) \models [\mathbf{A} \text{ xstit}] \varphi$ iff for all s', h' , if $s' \in E(s, h, \mathbf{A})$ and $s' \in h'$, then $(s', h') \models \varphi$

We can then define entailment in the usual way:

Definition 7.1.7. $\Gamma \models_{\text{NU}} \varphi$ iff for any neutral \mathcal{L} -model \mathfrak{M} , if $\mathfrak{M}, (s, h) \models \Gamma$, then $\mathfrak{M}, (s, h) \models \varphi$.

Notice that \mathbf{At} isn't broken up into \mathbf{At}_I and \mathbf{At}_B , since those distinctions don't matter to the logic, just the notion of norm consequence.

In Broersen and Meyer (2011) a proof sketch is provided that the axioms from definition 5.1.4 for the logic \vdash_x alone are complete with respect to the semantics from definitions 5.1.3 and 5.1.2 which we will call regular models. One thing to notice immediately is that each regular model is a neutral model, the \approx relation in a regular model is simply identity. That

means after noting that \vdash_{xp} is sound for neutral models, \vdash_{xp} is sound for regular models as well. But we will also show that \vdash_{xp} is complete for the class of neutral models.

7.1.3 Completeness Relative to **NU**

The proof of completeness proceeds in the usual way by constructing a canonical model for the logic. First we prove soundness.

Proposition 7.1.3. *The axioms from definition 6.3.6 are sound with respect to neutral \mathcal{L} -frames.*

Proof. We here provide the cases of the more irregular axioms. Clearly classical logic and MP are valid, as are the Nec rules. DX: If $(s, h) \models X\theta$, then $(lub(s, h), h) \not\models \neg\theta$. So $(s, h) \not\models X\neg\theta$, i.e., $(s, h) \models \neg X\neg\theta$.

DetX: Suppose $(s, h) \models \neg X\neg\theta$. Then $(s, h) \not\models X\neg\theta$, so $(lub(s, h), h) \not\models \neg\theta$. But that means $(lub(s, h), h) \models \theta$, thus $(s, h) \models X\theta$. The argument is similar for the P case except using glb .

XP: Suppose that $(s, h) \models XP\theta$. Then $(lub(s, h), h) \models P\theta$, but that means $(glb(lub(s, h), h), h) \models \theta$. Now in an injective DDLF, for each s , $f_h(b_h(s)) = s$. Thus $glb(lub(s, h), h) = s$ by observation 7.1.1 3. Thus $(s, h) \models \theta$. Suppose that $(s, h) \models \theta$. Again since $glb(lub(s, h), h) = s$, $(glb(lub(s, h), h), h) \models \theta$, so $(lub(s, h), h) \models P\theta$. Finally, $(s, h) \models XP\theta$. The PX case is symmetric.

NP: Suppose $(s, h) \models P\Box\theta$. Then $(glb(s, h), h) \models \Box\theta$. That means that for any h' with $s' \in h'$ such that $s' \approx glb(s, h)$, $(s', h') \models \theta$. Now suppose that $s^* \in h^*$ with $s^* \approx s$, thus $b_h(s) \approx b_{h^*}(s^*)$ by the condition H2, so $glb(s, h) \approx glb(s^*, h^*)$. By what we just noted, we can conclude that $(s^*, h^*) \models P\theta$. Since (s^*, h^*) was arbitrary we can conclude that $(s, h) \models \Box P\theta$.

$\emptyset = SettX$: Suppose $(s, h) \models [\emptyset xstit] \theta$. By condition d $E(s, h, \emptyset) = lub(s)$. Let (s', h') be such that $s \approx s' \in h'$. By the supposition, we have that for each $s' \in lub(s)$, and h' such that $s' \in h'$ $(s', h') \models \theta$. Thus $(lub(s'), h') \models \theta$. Since $s' \in lub(s)$, $glb(s', h') \approx s$ and so $(glb(s', h'), h') \models X\theta$. Since (s', h') was arbitrary, for all (s', h') , with $s' \in lub(s)$, $(glb(s', h'), h') \models X\theta$. Therefore, $(s, h) \models \Box X\theta$.

Suppose $(s, h) \models \Box X\theta$. Let $s' \in E(s, h, \emptyset)$ which is $\text{lub}(s)$ by condition d. Suppose that $s^* \in h'$ so that both $s \approx s^*$ and $\text{lub}(s^*, h') = s'$. By the supposition $(s^*, h') \models X\theta$ and because $\text{lub}(s^*, h') = s'$, $(s', h') \models \theta$. Since s' and h' were arbitrary, $(s, h) \models [\emptyset \text{ xstit}] \theta$.

Ag = XSett: Suppose $(s, h) \models [\mathbf{Ag} \text{ xstit}] \theta$. So by definition for each $s' \in E(s, h, \mathbf{Ag})$ and h' with $s' \in h'$ $(s', h') \models \theta$. By condition e from definition 7.1.4

$E(s, h, \mathbf{Ag}) = \{s' : s' \approx \text{lub}(s, h)\}$. So if we take $s' \approx \text{lub}(s, h)$, $s' \in E(s, h, \mathbf{Ag})$ thus for any $(s'', h'') \in |\mathfrak{M}|$ with $s'' \approx \text{lub}(s, h)$, $(s'', h'') \models \theta$. That means $(\text{lub}(s, h), h) \models \Box\theta$. So finally $(s, h) \models X\Box\theta$.

Suppose $(s, h) \models X\Box\theta$. Then by definition $(\text{lub}(s, h), h) \models \Box\theta$. And for each $s' \in h'$ with $s' \approx \text{lub}(s, h)$, $(s', h') \models \theta$. But since $E(s, h, \mathbf{Ag}) = \{s' : s' \approx \text{lub}(s, h)\}$, for all $s' \in E(s, h, \mathbf{Ag})$ and h' with $s' \in h'$, $(s', h') \models \theta$. Therefore, $(s, h) \models [\mathbf{Ag} \text{ xstit}] \theta$.

C-mon: Suppose $(s, h) \models [\mathbf{A} \text{ xstit}] \theta$. So for all $s' \in E(s, h, \mathbf{A})$ and h' with $s' \in h'$, $(s', h') \models \theta$. There are two possibilities either $\mathbf{A} \cup \mathbf{B} = \mathbf{A}$ or not. If so then $E(s, h, \mathbf{A} \cup \mathbf{B}) = E(s, h, \mathbf{A})$ so $(s, h) \models [\mathbf{A} \cup \mathbf{B} \text{ xstit}] \theta$. If not, then $\mathbf{A} \subsetneq \mathbf{A} \cup \mathbf{B}$, so by condition f, $E(s, h, \mathbf{A} \cup \mathbf{B}) \subseteq E(s, h, \mathbf{A})$. If $s' \in E(s, h, \mathbf{A} \cup \mathbf{B})$, then $s' \in E(s, h, \mathbf{A})$; therefore, for each h' with $s' \in h'$ $(s', h') \models \theta$. Hence $(s, h) \models [\mathbf{A} \cup \mathbf{B} \text{ xstit}] \theta$.

The axiom is **Indep-G**: Suppose that $(s, h) \models \Diamond [\mathbf{A} \text{ xstit}] \theta \wedge \Diamond [\mathbf{B} \text{ xstit}] \theta'$. So for some $(s', h'), (s'', h'')$ with $s' \approx s \approx s''$, $(s', h') \models [\mathbf{A} \text{ xstit}] \theta$ and $(s'', h'') \models [\mathbf{B} \text{ xstit}] \theta'$. That also means that for all $(s_1, h_1), (s_2, h_2)$ such that $s_1 \in E(s', h', \mathbf{A})$ and $s_2 \in E(s'', h'', \mathbf{B})$, $(s_1, h_1) \models \theta$ and $(s_2, h_2) \models \theta'$. By condition g we then have (s^*, h^*) such that $s' \approx s^*$, and both $E(s^*, h^*, \mathbf{A}) \subseteq E(s', h', \mathbf{A})$ and $E(s^*, h^*, \mathbf{B}) \subseteq E(s'', h'', \mathbf{B})$. Let $s_t \in E(s^*, h^*, \mathbf{A})$, then $s_t \in E(s', h', \mathbf{A})$, so $(s_t, h_t) \models \theta$ and the same will go for arbitrary $s_t \in E(s^*, h^*, \mathbf{B})$ in relation to θ' . Thus $(s^*, h^*) \models [\mathbf{A} \text{ xstit}] \theta \wedge [\mathbf{B} \text{ xstit}] \theta'$, and since \approx is an equivalence we have $s \approx s^*$; thus, $(s, h) \models \Diamond ([\mathbf{A} \text{ xstit}] \theta \wedge [\mathbf{B} \text{ xstit}] \theta')$. \square

So \vdash_{xp} is sound for \models_{NU} , and so a fortiori for \models_{xp} . Next we show completeness via a canonical model construction.

Relations for the Canonical Models

Let Δ, Δ' be \vdash_{xp} maximal consistent sets. These can be derived from maximal extensions of consistent sets with the usual Lindenbaum extension. The set of all maximally consistent sets of \mathcal{L} is $\max^+(\mathcal{L})$.

Definition 7.1.8. Define the relations on $\max^+(\mathcal{L})$ as follows:

$$\begin{aligned}\Delta R_{\Box} \Delta' &\iff \forall \theta, \Box \theta \in \Delta \Rightarrow \theta \in \Delta' \\ \Delta R_X \Delta' &\iff \forall \theta, X\theta \in \Delta \Rightarrow \theta \in \Delta' \\ \Delta R_A \Delta' &\iff \forall \theta, [\mathbf{A} \text{ xstit}] \theta \in \Delta \Rightarrow \theta \in \Delta' \\ \Delta R_P \Delta' &\iff \forall \theta, P\theta \in \Delta \Rightarrow \theta \in \Delta' \\ \Delta R_S \Delta' &\iff [\forall \theta, n \in \mathbb{N}, P^n \Box \theta \in \Delta \Leftrightarrow P^n \Box \theta \in \Delta']\end{aligned}$$

where $P^n = \underbrace{P \dots P}_{n\text{-times}}$.

The R_P relation is like the converse of the R_X relation. Finally, let's notice that if $\forall \theta, n \in \mathbb{N}, P^n \Box \theta \in \Delta \Leftrightarrow P^n \Box \theta \in \Delta'$, then for all $\varphi \Box \varphi \in \Delta$ iff $\Box \varphi \in \Delta'$, but since \Box obeys T, $\Delta R_{\Box} \Delta'$. Clearly, R_S is an equivalence relation, thus it partitions $\max^+(\mathcal{L})$.

We will now proceed to construct canonical models for \vdash_{xp} . From these relations we can define the canonical collection of static states. But before we do that let's make some observations

Observation 7.1.4. 1. If $\Delta R_{\Box} \Delta'$ and $\varphi \in \Delta'$, then $\Diamond \varphi \in \Delta$

2. If $\mathbf{A} \subsetneq \mathbf{B}$, then $R_B \subseteq R_A$.

3. For all \mathbf{A} , $R_A \subseteq R_{\emptyset}$.

4. For all \mathbf{A} , $R_{\mathbf{A}g} \subseteq R_A$.

5. If $\Delta R_X \Delta'$ and $\theta \in \Delta'$, then $X\theta \in \Delta$.

6. If $\Delta R_P \Delta'$ and $\theta \in \Delta'$, then $P\theta \in \Delta$.

7. If $\Delta R_X \Delta'$ and $\Delta R_X \Delta''$, then $\Delta'' = \Delta'$.

8. If $\Delta R_P \Delta'$ and $\Delta R_P \Delta''$, then $\Delta'' = \Delta'$.

9. R_X and R_P are functions.

Proof. 1 is a standard point in modal logic. 2 follows from C-mon, and 3 and 4 from 1 and 2. For 5 and 6, Assume $\Delta R_X \Delta'$ and $\theta \in \Delta'$. Then suppose $X\theta \notin \Delta$. But if that is the case, then by maximality and DetX $X\neg\theta \in \Delta$, so $\neg\theta \in \Delta'$ because $\Delta R_X \Delta'$. But that is impossible. Thus $X\theta \in \Delta$. For 6 we just swap P for X . Results 7 and 8 follow immediately from DetX and DetP, respectively.

Finally, for 9, DX and DP imply that R_X and R_P are serial relations. 7 and 8 then imply that R_X and R_P are partial functions. Together with seriality R_X and R_P are total functions. \square

Also we notice that:

Proposition 7.1.5. $\Delta R_X \Delta \iff \neg \exists \Delta' \neq \Delta \text{ s.t. } \Delta R_X^+ \Delta'. \text{ (Same for } R_P.)$

Proof. (IF) Note that R_X is serial by DX, thus there is something that R_X relates Δ to. So if there is no distinct maximal set that Δ relates to, R_X must relate Δ to itself.

(ONLY IF) By contraposition. Suppose that $\Delta R_X^+ \Delta'$ with $\Delta \neq \Delta'$. We show that $\Delta \not R_X \Delta$. There must be a least m , and $\Delta'' \neq \Delta$ such that $\Delta R_X^m \Delta''$. I.e., if $\Delta R_X^k \Delta''$ where $k < m$, then $\Delta = \Delta'$. This follows by the well ordering of \mathbb{N} . Suppose for reductio that $\Delta R_X \Delta$. If $m = 0$, then $\Delta R_X \Delta''$ and $\Delta \neq \Delta''$, but $\Delta R_X \Delta$, so we would have a contradiction since R_X -successors are unique by result 7. So suppose $m > 0$. Then we have that $\Delta R_X^{m-1} \Delta$ (since m was the least natural number) and $\Delta R_X \Delta''$ by definition of R_X^m , with $\Delta \neq \Delta''$. But again that would contradict our assumption that $\Delta R_X \Delta$. Thus, not $\Delta R_X \Delta$. \square

Also let's note that

Observation 7.1.6. (1) If $\Delta, \Delta' \in \max^+(\mathcal{L})$ and $\Delta \neq \Delta'$, then $\Delta R_X \Delta'$ only if $\Delta' R_P \Delta$.

(2) If $\Delta, \Delta' \in \max^+(\mathcal{L})$ and $\Delta \neq \Delta'$, then $\Delta R_P \Delta'$ only if $\Delta' R_X \Delta$.

(3) If $\Delta_1, \dots, \Delta_{n+1}$ are all distinct with $\Delta_i R_X \Delta_{i+1}$ and $\Delta_{i+1} R_P \Delta_i$, $\Delta_1 R_X^n \Delta_n$ iff $\Delta_n R_P^n \Delta_1$.

Proof. Since $\Delta \neq \Delta'$, there is $\theta \in \Delta'$ and $\theta \notin \Delta$. By observation 7.1.4 we know that $X\theta \in \Delta$, and so $X\theta \wedge \neg\theta \in \Delta$. Thus $\neg(X\theta \supset \theta) \in \Delta$. This means that for any φ , $XP\varphi \equiv \varphi \in \Delta$ by XP. Suppose that $P\theta' \in \Delta'$, then $XP\theta' \in \Delta$, again by observation 7.1.4; thus, $\theta' \in \Delta$. Since θ' was arbitrary, $\Delta' R_P \Delta$. (2) Follows by swapping P for X and using PX.

The third observation follows from the first two by induction. \square

Now we want to define a canonical model $\mathfrak{M}^o = \langle S^o, H^o, E^o, \approx^o, v^o \rangle$ from the maximal consistent sets.

Histories

We can define the canonical domain as follows:

Definition 7.1.9 (Canonical Static Domain). Let S^o be the canonical static domain, i.e., $S^o = \max^+(\mathcal{L})$.

We now define the histories in a couple of steps. To define histories we start with the base of a history as a line running through $\max^+(\mathcal{L})$ as follows:

$$h_\Delta =_{Df} \{ \Delta' \in S^o : \Delta R_X^+ \Delta', \Delta R_P^+ \Delta' \text{ or } \Delta' = \Delta \} \quad (7.2)$$

This relates maximal sets that are successors or predecessors of the current set Δ . Some distinct maxi sets will end up in the same lines. But more importantly, from observation 7.1.6, we can argue by induction that $\Delta R_X^n \Delta'$ iff $\Delta' R_P^n \Delta$ as long as $\Delta \neq \Delta'$. So, if we have $\Delta R_X^n \Delta'$, and $\Delta \neq \Delta'$, then $\Delta' R_P^n \Delta$. But that means $\Delta \in h_{\Delta'}$. So when either $\Delta R_X^+ \Delta'$ or $\Delta R_P^+ \Delta'$, then $h_\Delta = h_{\Delta'}$. These lines are also equivalence relations on $\max^+(\mathcal{L})$, and so we will be guaranteed that each maxi set appears in some line. Now we can notice something about these lines through $\max^+(\mathcal{L})$.

Lemma 7.1.7. Let $h_{\Delta''}$ be defined as in 7.2. Now define the functions f and b on $h_{\Delta''}$ as follows:

For $\Delta, \Delta' \in h_{\Delta''}$

1 $f(\Delta) = \Delta'$ iff $\Delta R_X \Delta'$, and

2 $b(\Delta) = \Delta'$ iff $\Delta R_P \Delta'$.

Then $\langle h_{\Delta''}, f, b \rangle$ is an injective DDLF.

Proof. First note that R_X and R_P are functions on the set of maximal sets by observation 7.1.4(9). f^+ and b^+ satisfy TRI by definition of $h_{\Delta''}$ but we relate R_X to f as $xR_X^n y$ iff $f^{n+1}(x) = y$, similarly for R_P and b . Suppose that $f^+(\Delta) \neq \Delta'$, and that $b^+(\Delta) \neq \Delta'$, we want to show that $\Delta = \Delta'$. Since $\Delta, \Delta' \in h_{\Delta''}$, we know that both

$$\Delta R_X^+ \Delta'', \Delta R_P^+ \Delta'' \text{ or } \Delta'' = \Delta$$

and

$$\Delta' R_X^+ \Delta'', \Delta' R_P^+ \Delta'' \text{ or } \Delta'' = \Delta'$$

by definition, so we argue by cases. If both $\Delta'' = \Delta$ and $\Delta'' = \Delta'$, then $\Delta' = \Delta$. On the other hand, if only one of either $\Delta'' = \Delta$ or $\Delta'' = \Delta'$, then that will contradict our assumptions that $f^+(\Delta) \neq \Delta'$, and $b^+(\Delta) \neq \Delta'$. So $\Delta'' \neq \Delta$ or $\Delta'' \neq \Delta'$. But that means by an easy induction from observation 7.1.6 that

$$\Delta R_X^+ \Delta'' \iff \Delta'' R_P^+ \Delta$$

and the same with Δ' . But if $\Delta R_X^+ \Delta'' \Delta R_P^+ \Delta''$, then by transitivity we would have a contradiction with our assumptions that $f^+(\Delta) \neq \Delta'$, and $b^+(\Delta) \neq \Delta'$. Similarly when $\Delta R_P^+ \Delta''$ and $\Delta' R_X^+ \Delta''$. So either $\Delta R_X^+ \Delta''$ and $\Delta' R_X^+ \Delta''$ or $\Delta' R_P^+ \Delta''$ and $\Delta' R_P^+ \Delta''$.

If it is the first case then there are $m, n \in \mathbb{N}$ such that $\Delta R_X^n \Delta''$ and $\Delta' R_X^m \Delta''$. If $m \neq n$, that means without loss of generality that $m < n$, and so $\Delta R_X^{n-(m+1)} \Delta'$, and so $n - (m + 1) \geq 0$, but then $\Delta R_X^+ \Delta'$, contrary to assumption. So $m = n$. But that means $\Delta = \Delta'$. The case is symmetric for the R_P^+ case. Thus, $\Delta = \Delta'$.

Both functions are almost injective for suppose that $b(\Delta') \neq \Delta'$ and $b(\Delta'') \neq \Delta''$. Now suppose that $b(\Delta') = \Delta^*$ and $b(\Delta'') = \Delta^*$. Then we know that $\Delta' \neq \Delta^* \neq \Delta''$, and so by observation 7.1.6(2) and the definition of b , we have that $\Delta^* R_X \Delta'$ and $\Delta^* R_X \Delta''$, thus $\Delta' = \Delta''$ since R_X is a function.

The fixed points of R_X and R_P are unique. If there were two, Δ' and Δ , then either $\Delta R_P^+ \Delta'$ or $\Delta' R_P^+ \Delta$. Either way there would need to some non-identical set R_P -related to the fixed point which is impossible given the assumption that $\Delta R_P \Delta$ and $\Delta' R_P \Delta'$. The same reasoning works for the R_X case.

CONV1 and 2 follow from observation 7.1.4 pretty much immediately. If $h_{\Delta''}$ has more than one member, then suppose that $\Delta', \Delta \in h_{\Delta''}$, and $f(\Delta') = \Delta'$. Now it must be that either $\Delta R_X^+ \Delta'$ or $\Delta R_P^+ \Delta'$. If $\Delta = \Delta'$, then we see that there must be another Δ'' such that $\Delta'' R_X \Delta'$, and $\Delta'' \neq \Delta'$. But then It is not the case that $\Delta' R_P \Delta'$. If $\Delta \neq \Delta'$, there must still be $\Delta'' \neq \Delta'$ such that $\Delta'' R_X \Delta'$, so it is not the case that $\Delta' R_P \Delta'$. Thus SIZE holds.

Now that we know that $\langle h_{\Delta''}, f, b \rangle$ is a DDLF, let's notice that for all $\Delta \neq \Delta' \in h_{\Delta''}$,

$$\Delta R_X \Delta' \iff \Delta' R_P \Delta$$

Suppose (1) $\Delta R_X \Delta'$, and (2) $P\varphi \in \Delta'$. Then $XP\varphi \in \Delta$ because of 1, so that means $\varphi \in \Delta$ by closure, XP and PX. Thus $\Delta' R_P \Delta$. Now suppose (3) $\Delta' R_P \Delta$, and let $X\varphi \in \Delta$. Then $PX\varphi \in \Delta'$ by 3 and so $\varphi \in \Delta'$ by XP and PX. Hence $\Delta R_X \Delta'$. But that means $f(b(\Delta)) = \Delta = b(f(\Delta))$, and by observation 7.1.1, f and b are both injective. \square

Thus, we define histories as $h = h_{\Delta}$ for $\Delta \in \max^+ \mathcal{L}$, and for $\Delta', \Delta'' \in h$,

$$f_h^o(\Delta') = \Delta'' \iff \Delta' R_X \Delta'' \tag{7.3}$$

and

$$b_h^o(\Delta') = \Delta'' \iff \Delta' R_P \Delta'' \tag{7.4}$$

So this gives rise to the definitions of *lub* and *glb* as

$$f_h^o(\Delta') = \text{lub}(s, h) \ \& \ s = \Delta' \tag{7.5}$$

and

$$b_h^o(\Delta') = \text{glb}(s, h) \ \& \ s = \Delta' \tag{7.6}$$

Now we will deal with the \approx^o relation and as one might guess that is going to be handled by R_S :

$$\Delta \approx^o \Delta' \iff \Delta R_S \Delta' \quad (7.7)$$

Now we have to check the conditions governing it, i.e., H2.

Observation 7.1.8. *If $\Delta = s \in h$ and $\Delta' = s' \in h'$ with $s \approx^o s'$, then for each $n \in \mathbb{N}$, $(b_h^o)^n(s) \approx^o (b_{h'}^o)^n(s')$*

Proof. Let $\Delta \in h$ and $\Delta' \in h'$ with $\Delta \approx^o \Delta'$, i.e., $\Delta R_S \Delta'$. Let $n \in \mathbb{N}$, and suppose $\Delta R_P^n \Delta^*$. Assume $P^m \Box \varphi \in \Delta^*$; then $P^n P^m \Box \varphi \in \Delta$. So $P^n P^m \Box \varphi \in \Delta'$, thus $P^m \Box \varphi \in \Delta''$ such that $\Delta' R_P^n \Delta''$. The same goes in the other direction. Thus $(b_h^o)^n(\Delta) = \Delta^* \approx^o \Delta'' = (b_{h'}^o)^n(\Delta')$. So it obeys condition H2. \square

The final thing we have to check is whether we can define effectivity functions.

Effectivity Functions

Now that we have made sure that H^o is constructed properly we must construct the effectivity function: $E^o(s, h, \mathbf{A})$.

Definition 7.1.10. We define $E^o(\Delta, h, \mathbf{A})$ as follows

$$E^o(\Delta, h, \mathbf{A}) = \begin{cases} \{ \Delta' : \Delta R_{\mathbf{A}} \Delta' \} & \text{if } \Delta \in h \\ \emptyset & \text{o.w.} \end{cases}$$

We must now check to make sure that E^o defined in this way meets all of the criteria for an effectivity function from definition 5.1.1. First we mention a lemma.

Lemma 7.1.9. 1. $s' = \text{lub}(s, h)$ iff $\Delta' = s'$, $\Delta = s$ with $\Delta, \Delta' \in h$ and $\Delta R_X \Delta'$.

2. For $\Delta = s \in S^o$, $\text{lub}(s) = \{ \Delta' : \exists \Delta'' \text{ w/ } \Delta R_S \Delta' \ \& \ \Delta'' R_X \Delta' \}$.

3. $\Box \theta \supset [\mathbf{A} \text{ xstit}] P \theta$ is a theorem for each \mathbf{A} .

Proof. The first two follow immediately from the definitions. For the third observation note that $\Box\theta \equiv \Box XP\theta \equiv [\emptyset \text{ xstit}] P\theta$, so it follows that $[\mathbf{A} \text{ xstit}] P\theta$ by C-mon. \square

Proposition 7.1.10. *E^o meets criteria a)–g).*

Proof. Let's assume that $s = \Delta$ and $h = h_\Delta$. (a) If $s \not\in h$, then $E(s, h, \mathbf{A}) = \emptyset$: If $s = \Delta \not\in h$, then $E(s, h, \mathbf{A}) = \emptyset$ by definition.

(b) If $s' \in E(s, h, \mathbf{A})$, then $s' \in \text{lub}(s)$: if $s' = \Delta'$, and $s = \Delta$, with $E^o(\Delta, h, \mathbf{A})$ then by definition $\Delta R_{\mathbf{A}} \Delta'$. We define the history h' as follows: $h' = h_{\Delta'}$. Note that

$$\Phi = \{ \theta : P\theta \in \Delta' \} \cup \{ P^n \Box \varphi : P^n \Box \varphi \in \Delta \}$$

is consistent. If it wasn't, there would be some φ s and θ s, such that $\{ P^n \Box \varphi \} \vdash \neg \theta$, so $\{ \Box P^n \Box \varphi \} \vdash \Box \neg \theta$ (normality of \Box). But $P^n \Box \varphi \vdash \Box P^n \Box \varphi$ (NP, and use of 4 for \Box), so $\Box \neg \theta \in \Delta$. From last observation of lemma 7.1.9, $\Box \neg \theta \vdash [\mathbf{A} \text{ xstit}] P \neg \theta$, thus $[\mathbf{A} \text{ xstit}] P \neg \theta \in \Delta$. That means $P \neg \theta \in \Delta'$ which then means that both $P\theta \in \Delta'$ as well so by DP $\Delta' \vdash \perp$ which is a contradiction. Thus we extend Φ to a maximal set Φ^+ . Clearly, $\Delta R_S \Phi^+$, and if $P\theta \in \Delta'$, then $\theta \in \Phi^+$, so $\Delta' R_P \Phi^+$. But then we can conclude that $\Phi^+ R_X \Delta'$ since R_X and R_P are injective.

(c) If $s \in h$, $\text{lub}(s, h) \in E(s, h, \mathbf{A})$: If Δ' is such that $\Delta R_X \Delta'$, i.e., $\Delta' = \text{lub}(s, h)$, then suppose that $[\mathbf{A} \text{ xstit}] \varphi \in \Delta$. But then by C-mon, $[\mathbf{A} \mathbf{g} \text{ xstit}] \varphi \in \Delta$ and by $\mathbf{A} \mathbf{g} X \Box \varphi \in \Delta$. But that means $\Box \varphi \in \Delta'$ so by T for \Box , $\varphi \in \Delta'$. Thus, $\Delta R_{\mathbf{A}} \Delta'$. Hence, $\Delta' \in E(s, h, \mathbf{A})$.

(d) $E(s, h, \emptyset) = \text{lub}(s)$: From b we have for any $\mathbf{A} \subseteq \mathcal{P}(\mathbf{A} \mathbf{g})$, $E(s, h, \mathbf{A}) \subseteq \text{lub}(s)$ so the same holds for $\mathbf{A} = \emptyset$. For the other direction suppose $s' \in \text{lub}(s)$, so $s' = \Delta'$ and $\Delta'' R_X \Delta'$ with $\Delta R_S \Delta''$. We want to show that $\Delta R_{\emptyset} \Delta'$. Suppose that $[\emptyset \text{ xstit}] \theta \in \Delta$. So $\Box X\theta \in \Delta$ by SettX, and then since $\Delta R_{\Box} \Delta''$, $X\theta \in \Delta''$, thus $\theta \in \Delta'$.

(e) If $s \in h$, then $E(s, h, \mathbf{A} \mathbf{g}) = \{ s' : s' \approx \text{lub}(s, h) \}$: Suppose $s = \Delta$, then by definition $E(\Delta, h, \mathbf{A} \mathbf{g}) = \{ \Delta' : \Delta R_{\mathbf{A} \mathbf{g}} \Delta' \}$. We want to show that $\{ \Delta' : \Delta' \approx^o \text{lub}(s, h) \} = \{ \Delta' : \Delta R_{\mathbf{A} \mathbf{g}} \Delta' \}$. Since $s = \Delta$, by 7.5 we have that $\text{lub}(s, h) = f_h^o(\Delta)$, and by 7.7, $\approx^o = R_S$. What we want to show is that both $\{ \Delta' : \Delta R_{\mathbf{A} \mathbf{g}} \Delta' \} \subseteq \{ \Delta' : f_h^o(\Delta) R_S \Delta' \}$ and vice versa.

Suppose that $\Sigma \in \{ \Delta' : \Delta R_{\mathbf{Ag}} \Delta' \}$, then $\Sigma \in lub(\Delta)$ from b above, so there is $\Psi \approx^o \Delta$ and $\Psi R_X \Sigma$. That means if $P^n \Box \varphi \in \Sigma$, then $P^{n-1} \Box \varphi \in \Psi$ by observation 7.1.4(5), XP and PX. So $P^{n-1} \Box \varphi \in \Delta$, thus $P^n \Box \varphi \in f_h^o(\Delta)$ (since $f_h^o(\Delta) R_P \Delta$). Now suppose that $P^n \Box \varphi \in f_h^o(\Delta)$, then $\Box P^n \Box \varphi \in f_h^o(\Delta)$ and so $X \Box P^n \Box \varphi \in \Delta$, but that means $[\mathbf{Ag} \text{ xstit}] P^n \Box \varphi \in \Delta$; thus, $P^n \Box \varphi \in \Sigma$. So for $n > 0$, $P^n \Box \varphi \in f_h^o(\Delta)$ iff $P^n \Box \varphi \in \Delta'$. Now suppose that $\Box \varphi \in f_h^o(\Delta)$, then $X \Box \varphi \in \Delta$, thus $[\mathbf{Ag} \text{ xstit}] \varphi \in \Delta$ so $\varphi \in \Sigma$. That means $f_h^o(\Delta) R_{\Box} \Sigma$, and so $f_h^o(\Delta) R_S \Sigma$. So $\{ \Delta' : \Delta R_{\mathbf{Ag}} \Delta' \} \subseteq \{ \Delta' : f_h^o(\Delta) R_S \Delta' \}$. Now for the other containment. Assume that $f_h^o(\Delta) R_S \Sigma$, i.e., $\Sigma \in \{ \Delta' : f_h^o(\Delta) R_S \Delta' \}$, then let $[\mathbf{Ag} \text{ xstit}] \varphi \in \Delta$, we have $X \Box \varphi \in \Delta$, so $\Box \varphi \in f_h^o(\Delta)$ and so $\Box \varphi \in \Sigma$, but then by T $\varphi \in \Delta'$. I.e., $\Delta R_{\mathbf{Ag}} \Sigma$. Thus, $\{ \Delta' : f_h^o(\Delta) R_S \Delta' \} \subseteq \{ \Delta' : \Delta R_{\mathbf{Ag}} \Delta' \}$.

(f) If $\mathbf{A} \subsetneq \mathbf{B}$, then $E(s, h, \mathbf{B}) \subseteq E(s, h, \mathbf{A})$: Follows because $R_B \subseteq R_A$ from observation 7.1.4.

(g) For all $\mathbf{A}, \mathbf{B} (s, h), (s', h'), (s'', h'')$, if $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $s' \approx s \approx s''$, then there is (s''', h''') such that $s''' \approx s$ with $E(s''', h''', \mathbf{A}) \subseteq E(s', h', \mathbf{A})$ and $E(s''', h''', \mathbf{B}) \subseteq E(s'', h'', \mathbf{B})$: Suppose $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $\Sigma \approx^o \Delta \approx^o \Psi$. Then the set $\Phi = \{ P^n \Box \varphi \in \Delta : n \in \mathbb{N} \} \cup$

$$\{ [\mathbf{A} \text{ xstit}] \theta \in \Sigma : \Diamond [\mathbf{A} \text{ xstit}] \theta \in \Delta \} \cup \{ [\mathbf{B} \text{ xstit}] \theta' \in \Psi : \Diamond [\mathbf{B} \text{ xstit}] \theta' \in \Delta \}$$

is consistent. If it wasn't there would be θ, θ' and φ such that

$$[\mathbf{A} \text{ xstit}] \theta \wedge [\mathbf{B} \text{ xstit}] \theta' \wedge P^n \Box \varphi \vdash \perp$$

which then means that with a bit of modal logic

$$\Diamond([\mathbf{A} \text{ xstit}] \theta \wedge [\mathbf{B} \text{ xstit}] \theta') \vdash \Diamond \neg P^n \Box \varphi$$

But $\Diamond \neg P^n \Box \varphi \equiv \neg \Box P^n \Box \varphi$ and $P^n \Box \varphi \equiv \Box P^n \Box \varphi$, so

$$\Diamond([\mathbf{A} \text{ xstit}] \theta \wedge [\mathbf{B} \text{ xstit}] \theta') \vdash \neg P^n \Box \varphi$$

Since the xstit operators are normal and each $[\mathbf{A} \text{ xstit}] \theta$ is in Σ and $[\mathbf{B} \text{ xstit}] \theta'$ is in Ψ we will have $\Diamond [\mathbf{A} \text{ xstit}] \theta \in \Delta$ and $\Diamond [\mathbf{B} \text{ xstit}] \theta' \in \Delta$ because $\Delta R_{\Box} \Sigma$ and $\Delta R_{\Box} \Psi$. But that means,

from axiom Indepen-G, $\Diamond([A \text{ xstit}] \theta \wedge [B \text{ xstit}] \theta') \in \Delta$. But that would mean $\neg P^n \Box \varphi \in \Delta$, but it can't be.

Now we extend Φ to Φ^+ . Now for each θ , $[A \text{ xstit}] \theta \in \Phi^+ \iff [A \text{ xstit}] \theta \in \Sigma$, so if Δ^* is such that $\Phi^+ R_A \Delta^*$, then if $[A \text{ xstit}] \theta \in \Sigma$, $\theta' \in \Delta^*$. Thus $\Sigma R_A \Delta^*$. The same goes for R_B . Thus $E(\Phi^+, h_{\Phi^+}, A) \subseteq E(\Sigma, h_\Sigma, A)$ and $E(\Phi^+, h_{\Phi^+}, B) \subseteq E(\Sigma, h_\Sigma, B)$. And it clearly holds that $\Delta \approx^o \Phi^+$.

□

Thus, $\mathcal{F}^o = \langle S^o, H^o, E^o, \approx^o \rangle$ is a \mathcal{L} -frame, we then make the canonical model in the usual way by setting $v^o(\mathbf{p}) = \{ \Delta : \mathbf{p} \in \Delta \}$. The domain $|\mathcal{F}^o|$ is the set of (s, h) such that $s \in h$ (i.e., $\Delta \in h$). Note that if $\mathbf{p} \in \Delta$, then $\mathbf{p} \in \Delta'$ for all Δ' that are \approx^o -related to Δ (They all agree on \Box ed formulas, and $\mathbf{p} \supset \Box \mathbf{p}$). Now we have to show the following:

Theorem 7.1.11 (Fundamental Theorem of \vdash_{xp}). *For all θ , and $s \in S^o$, if $(s, h) = (\Delta, h_\Delta)$, then*

$$\theta \in \Delta \iff (s, h) \models \theta.$$

To show this we must make an observation.

Observation 7.1.12. 1. *If $\Diamond \theta \in \Delta$, then there exists Δ' such that $\Delta R_S \Delta'$ and $\theta \in \Delta'$.*

Proof. If $\Diamond \varphi \in \Delta$, then $\Phi = \{ P^n \Box \theta \mid P^n \Box \theta \in \Delta \ \& \ n \in \mathbb{N} \} \cup \{ \varphi \}$ is consistent. If it wasn't, then there are some θ s, such that $P^n \Box \theta \vdash \neg \varphi$, and so $\Box P^n \Box \theta \vdash \Box \neg \varphi$. But $\{ \Box P^n \Box \theta \} \subseteq \Delta$; therefore, $\Box \neg \varphi \in \Delta$, i.e., $\neg \Diamond \varphi \in \Delta$, a contradiction. If we extend Φ to a maxi set Φ^+ , $\Phi^+ R_S \Delta$. Done. □

proof of theorem 7.1.11. By induction. For atomic sentences \mathbf{p} , $(s, h) \models \mathbf{p}$ iff $\Delta \in v^o(\mathbf{p})$ iff $\mathbf{p} \in \Delta$ by definition of v^o .

The IH is: for all θ' , with less complexity than θ and $s \in S^o$, if $s = \Delta'$ and $s \in h = h_{\Delta'}$, then

$$\theta' \in \Delta' \iff (s, h) \models \theta'.$$

We won't go through all of the cases in detail, just the novel ones. The \Box case uses the observation above. We will give the P case. Suppose $(s, h) \models P\theta$, then since $s = \Delta$, $glb(s, h) = b_h^o(\Delta)$. By the semantics $(glb(s, h), h) \models \theta$. We use the IH and $\theta \in b_h^o(\Delta)$. Of course, $\Delta R_P b_h^o(\Delta)$, (i.e., for all θ , $\theta \in b_h^o(\Delta)$ only if $P\theta \in \Delta$) and so $P\theta \in \Delta$. If $P\theta \in \Delta$, then $\theta \in b_h^o(\Delta) = glb(s, h)$, so $(glb(s, h), h) \models \theta$ by the IH, so then $(s, h) \models P\theta$.

The X case works the same as the P case just using f_h^o , the real trouble one is the $[\mathbf{A} \text{ xstit}]$ case. First we show that $(s, h) \models [\mathbf{A} \text{ xstit}] \theta$ only if $[\mathbf{A} \text{ xstit}] \theta \in \Delta$ where $\Delta = s$. As per usual this is done by contraposition. Suppose $[\mathbf{A} \text{ xstit}] \theta \notin \Delta$. The set

$$\Phi = \{ P^{n+1}\Box\varphi : P^n\Box\varphi \in \Delta \ \& \ n \in \mathbb{N} \} \cup \{ \theta' : [\mathbf{A} \text{ xstit}] \theta' \in \Delta \} \cup \{ \neg\theta \}$$

is consistent. If it wasn't there would be φ s and θ' s such that $P^{n+1}\Box\varphi \wedge \theta' \vdash \theta$. But then $[\mathbf{A} \text{ xstit}] P^{n+1}\Box\varphi \wedge [\mathbf{A} \text{ xstit}] \theta' \vdash [\mathbf{A} \text{ xstit}] \theta$. But from the lemma 7.1.9(3) above, $P^n\Box\varphi \vdash \Box P^n\Box\varphi$ and $\Box P^n\Box\varphi \vdash [\mathbf{A} \text{ xstit}] P P^n\Box\varphi$, so $[\mathbf{A} \text{ xstit}] P^{n+1}\Box\varphi \wedge [\mathbf{A} \text{ xstit}] \theta' \in \Delta$. That means $[\mathbf{A} \text{ xstit}] \theta \in \Delta$, but it isn't.

We extend Φ to Φ^+ . Clearly, $\Delta R_A \Phi^+$, and $\Delta R_S b^o(\Phi^+)$. Thus we have made $s' = \Phi^+ \in E(s, h, \mathbf{A})$, and set $h' = h_{\Phi^+}$ and so $\neg\theta \in \Phi^+$. By the IH $(s', h') \not\models \theta$. By the semantics $(s, h) \not\models [\mathbf{A} \text{ xstit}] \theta$.

For the converse, suppose $[\mathbf{A} \text{ xstit}] \theta \in \Delta$. Suppose $\Delta' = s' \in h'$ and $s' \in E(s, h, \mathbf{A})$. Then $\Delta R_A \Delta'$. So $\theta \in \Delta'$. By the IH $(s', h') \models \theta$. Now s' and h' were arbitrary; therefore, $(s, h) \models [\mathbf{A} \text{ xstit}] \theta$. Also \mathbf{A} is arbitrary. Thus we have completed the inductive step. \square

Now completeness is easily proved. If $\Gamma \not\vdash_{xp} \varphi$, for an arbitrary Γ , then $\Gamma; \neg\varphi$ is consistent, and so we can extend it to a maxi set Γ^+ . It will be in the canonical model, and $\neg\varphi \in \Gamma^+$, so $\varphi \notin \Gamma^+$, but $\Gamma \subseteq \Gamma^+$. Letting $s = \Gamma^+$ and $h = h_{\Gamma^+}$, by the fundamental theorem, $(s, h) \models \Gamma$, but $(s, h) \not\models \varphi$. Therefore, $\Gamma \not\models \varphi$. So we have proved:

Theorem 7.1.13. *The class of neutral \mathcal{L} -models provides a (strongly) complete semantics for the axioms of \vdash_{xp} .*

7.1.4 Different Classes

Now we have seen the following facts

$$\Gamma \vdash_{\text{xp}} \varphi \iff \Gamma \models_{\text{NU}} \varphi$$

and Broersen and Meyer (2011) said that

$$\Gamma \vdash_x \varphi \iff \Gamma \models_{\text{xp}} \varphi$$

What we now want to show is that \vdash_{xp} can be captured by the class of neutral models where each history is like \mathbb{Z} . That is each history in a neutral frame is is an infinite IDDLF. One thing to notice from before is that \vdash_{xp} is complete with respect to a smaller class of models than **NU**. The canonical model is one such that for all $h, h' \in H^o$ $h \cap h' = \emptyset$. So \vdash_{xp} is complete with respect to the class of *disjoint* neutral models.¹ We first have to start with some facts about histories and neutral models.

Lemma 7.1.14. *Let $h = \langle h, f, b \rangle$ be a history in a neutral model $\mathfrak{M} = \langle S, H, E, \approx \rangle$.*

1. *If $s \approx f(s)$ in h , then for any $n \in \mathbb{N}$, $b^n(f(s)) \approx f(s)$.*
2. *If $s \approx s' \in h$ when $s \neq s'$ with $|h| = n$, and $b^m(s) = s'$ and $b^m(s') = s$, then $f(s) \approx f(s')$.*
3. *If $s' \approx s$ in h with $b^m(s) = s'$, $b^{km}(s) \approx s$ for $k \in \mathbb{N}$.*
4. *If $s' \approx s$ in h with $b^m(s) = s'$, and $b^{km}(s') = s$ for some $k \in \mathbb{N}$, then $f(s) \approx f(s')$.*

Proof. 1. Suppose that $s \approx f(s)$ in h . We proceed by induction. For $n = 1$, then $b^1(f(s)) = s$ and $f(s) \approx s$. Now suppose that for all $k < n$, $b^k(f(s)) \approx f(s)$. $b^n(f(s)) = b(b^{n-1}(f(s)))$, so by inductive hypothesis $b^{n-1}(f(s)) \approx f(s)$. That means that $b(b^{n-1}(f(s))) \approx b(f(s))$ by H2, and $b(f(s)) = s \approx f(s)$, so $b^n(f(s)) \approx f(s)$.

2. Now suppose $s \approx s' \in h$ where $s \neq s'$ with $|h| = n$, $b^m(s) = s'$ and $b^m(s') = s$. Since $|h| = n$, $n = 2m$ since we can always get from one element to another either via f or

¹ Actually what Thomason (1984) calls a neutral model is what we call a disjoint neutral model.

b , but since we can get from s to s' and from s' to s with b , we go in a semicircle from s to s' and vice versa m steps each way, so $n = 2m$. But that means that $f(s') = b^{m-1}(s)$ and $b^{m-1}(s') = f(s)$. But we know, since $s \approx s'$, $b^{m-1}(s) \approx b^{m-1}(s')$, i.e., $f(s) \approx f(s')$.

3. Suppose $s' \approx s$ in h with $b^m(s) = s'$. We proceed by induction. For $n = 1$, $b^m(s) = s' \approx s$. Suppose that it is true for $k < l$, i.e., $b^{km}(s) \approx s$. Now, $b^{lm}(s) = b^m(b^{(l-1)m}(s))$, and by inductive hypothesis, $b^{(l-1)m}(s) \approx s$. So $b^m(b^{(l-1)m}(s)) \approx b^m(s)$ by H2, but $b^m(s) = s' \approx s$. Thus $b^{lm}(s) \approx s$.

4. Suppose that $s' \approx s$ in h with $b^m(s) = s'$, and $b^{km}(s') = s$ for some $k \in \mathbb{N}$. Then from our third observation, $b^{(k-1)m}(s') \approx s \approx s'$. Now $b^{m-1}(b^{(k-1)m}(s')) = f(s)$, and $b^{m-1}(s) = f(s')$. But we know that $b^{m-1}(b^{(k-1)m}(s')) \approx b^{m-1}(s)$ by H2, thus $f(s) \approx f(s')$. \square

These facts allow us to prove that

Lemma 7.1.15. *If $|h| = n$ and $s' \approx s$ in h , then $f(s) \approx f(s')$.*

Proof. Suppose $|h| = n$ and $s' \approx s$ in h . Since h is finite, there is some $m \in \mathbb{N}$ such that $b^m(s) = s'$. Then, because h is finite, $b^k(s') = s$ for some k . Now there are three possibilities, 1) $k = m$, $k < m$ or $k > m$. If $k = m$, then we have the situation in lemma 7.1.14 (2) so our conclusion follows. If $k > m$ and $m = 1$, then we have the situation in lemma 7.1.14 (1), and so our conclusion, again, follows. If $m > 1$, then there is some $t \in \mathbb{N}$ such that it is the smallest natural number such that $b^{lt}(s) \approx s$ (well orderedness of \mathbb{N}). It then follows that there is an $l \in \mathbb{N}$, such that $b^{lt}(s) = s$, i.e., $n = lt$. If there weren't then $t < n$, and $n = lt + r$ for some $r < t$. But we know from lemma 7.1.14 (2) that $b^{lt}(s) \approx s$, and $b^r(b^{lt}(s)) = s$, so $b^r(s) \approx b^r(b^{lt}(s)) = s$. But that contradicts t being the smallest such natural number. Now $m = jt + r$ for some $j, r \in \mathbb{N}$. If $r > 0$, and it must be less than t , then we have $b^{jt}(s) \approx s$, and $b^{jt+r}(s) = b^m(s) = s' \approx s$, but that means $s \approx b^r(b^{jt}(s)) \approx b^r(s)$. So t wouldn't be the smallest natural number; thus, $r = 0$. So $m = jt$. But then $b^{(l-1)t}(s) \approx s \approx b^{(j-1)t}(s)$ by lemma 7.1.14 (3). That means $b^{lt-1}(s) = f(s)$, and $b^{jt-1}(s) = f(s')$. But then $f(s') = b^{t-1}(b^{(j-1)t}(s)) = b^{jt-1}(s) \approx b^{t-1}(b^{(l-1)t}(s)) = f(s)$. Again our conclusion follows.

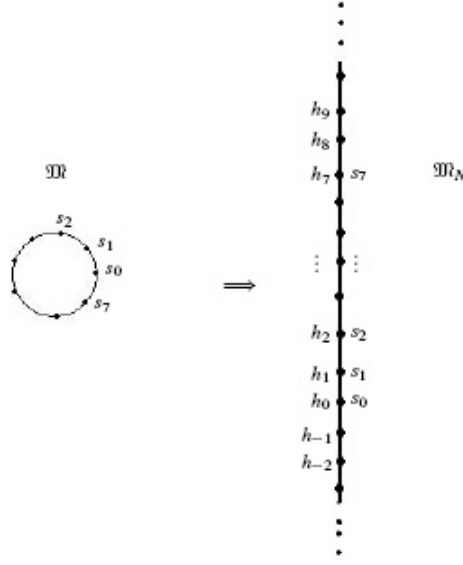


Figure 7.2: Stretching Out h to h_N

Now suppose that $k < m$. In this case we can argue in the same way as before except that we reverse the roles of s and s' . So then our conclusion follows as well. \square

We will use these facts in constructing our new model.

Let $\mathcal{M} = \langle S, H, E, \approx, v \rangle$ be a disjoint neutral model, i.e., one such that for each $h, h' \in H$, $h \cap h' = \emptyset$. We will build a new model which we will call $\mathfrak{M}_N = \langle S_N, H_N, E_N, \approx_N, v_N \rangle$ in a number of stages starting with H_N . Each $h \in H$ is either finite or not. If it is infinite, then $h_N = h = \langle h, f_h, b_h \rangle$, i.e., we just continue to use h . If not, then $|h| = n$, say, and we can pick an ordering of h as $h = \{s_0, \dots, s_{n-1}\}$ and then add to h two denumerable sets h^+ and h^- as $\langle h^-, h, h^+ \rangle$. Any sets that we add must be disjoint from other sets added and from S . So the new infinite h_N will look like: $h_{-2}, h_{-1}, s_0, s_1, \dots, s_{n-1}, h_n, h_{n+1}, \dots$. And we will notice that $h_0 = s_0$, $h_1 = s_1$, etc. So here we are taking up a convention of calling the n th position in h_N , h_n . An image of the transformation is in figure 7.2.

On each of these new sets h_N , if h was infinite, then we keep the old functions f_h and b_h . If h was finite we have to define new functions f_{h_N} and b_{h_N} . We do that as follows: define $f_{h_N}(h_i) = h_{i+1}$ and $b_{h_N}(h_i) = h_{i-1}$. Of course $f_{h_N}(h_i) = f_h(s_i)$ for $0 \leq i < n-2$. The idea

is that the successors of the “original” components of h in h_N are the same, i.e., $f_{h_N}(h_0) = f_{h_N}(s_0) = f_h(s_0)$, the same for s_1 and all up until s_{n-1} , the successor of $h_{n-1} = s_{n-1}$ in h_N has to be h_n , which is “the first” new element from h^+ . An analogous situation will hold for b_N and s_0 with respect to h^- . So $h_N = \langle h_N, f_{h_N}, b_{h_N} \rangle$. These new functions clearly form an IDDLF (it looks like $\langle \mathbb{Z}, z + 1, z - 1 \rangle$). Next we have to define \approx_N .

We also have to define \approx_N in stages. For the moment consider an h_N generated from a finite h from the original model. If $|h| = n$, we then stipulate that for $m < n$, and $k \in \mathbb{Z}$, $h_k \equiv_{h_N} s_m$ where $s_m = h_m$ and $k \equiv m \pmod{n}$. So every m th object (from h_0) in h_N will be \equiv -related to some member from the original set h . Notice that $h_{-1} \equiv_{h_N} s_{n-1}$, as we might expect. Also, $s_m \equiv_{h_N} s_m$ since $h_m = s_m$.

To form \approx_N we want to keep the old relationships in \approx , but add in all of the new ones formed from the \equiv_{h_N} relations. Thus we form the set

$$\bigcup_{h_N \in H_N} (\equiv_{h_N} \cup \approx) \quad (7.8)$$

This relation includes all of the new objects. Now to form \approx_N we take the reflexive, then symmetric, then the transitive closures of the set from 7.8. Clearly that new relation will be an equivalence relation since it will be reflexive, transitive and symmetric. Now this new relation doesn't require that $h_1 \approx_N h'_1$, for instance. The orders according to \mathbb{Z} h_N has are dependant on the ordering given to the h it was built from. However, each h_m will be \approx_N -related to some member of the original h —this is true trivially for infinite histories from H since \approx is reflexive. We will also have that if s' is an n th-predecessor of s in h_N , and if s^* is a member of h that s is \approx_N -related to, s' will be \approx_N -related to an n th-predecessor of s^* from the original order that was on h .

We must also notice that the reflexive and symmetric closure of each $\equiv_{h_N} \cup \approx$ would only relate new elements of h_N with elements of h_N . Any elements of h_N related to objects outside of h_N would have to already be related to those objects by \approx . What that means is that any objects from S that are related by \approx_N are related by \approx . So \approx_N is conservative over \approx . We can see this since if $s, s' \in S$, are related, there must be some chain in the reflexive and symmetric

closures of 7.8 to get from s to s' . We should also notice that no new relations between elements of S will be introduced by the reflexive and symmetric closures. But this chain must include some new element, but if s^* and s'' are new, then they only relate via old elements, so that chain was already there for \approx , and \approx is transitive. Also, we should notice that if $s, b_{h_N}(s)$ are not in h , then $s = h_i$ and $b_{h_N}(s) = h_{i-1}$, and s will relate to s_m such that $i \equiv m(\text{mod } n)$. So $i - 1 \equiv m - 1(\text{mod } n)$, thus $h_{i-1} \approx_N s_{m-1}$. That means the successor/predecessor relationships that holds in h are mirrored in h_N .

A notion that we will be using again and again in the next stage is the notion of a representative of a static state, we will denote this by $rep(s, h_N)$. It is defined as follows:

$$rep(s, h_N) = \begin{cases} s & \text{if } |h| \notin \mathbb{N} \\ s & |h| \in \mathbb{N} \ \& \ s \in h \\ s_m & \text{the least } m \text{ where } s = h_k \ \& \ k \equiv m(\text{mod } n) \text{ o.w.} \end{cases} \quad (7.9)$$

So in each case where $s^* = rep(s, h_N)$, $s^* \approx_N s$.

Now what can be shown, assuming those two facts (the conservativeness of \approx_N over \approx , and that the representative in h of h_{i-1} is the predecessor of the representative in h of h_i), is that condition H2 from definition 7.1.4 is met, i.e., if $s \approx_N s'$ with $s \in h_N$ and $s' \in h'_N$, then for all $n \in \mathbb{N}$, $b_{h_N}^n(s) \approx_N b_{h'_N}^n(s')$. We can show that by induction. The cases are long, but not hard using those two facts so we omit the proof.

Here we can define the effectivity function for the new model. If $s \notin h$, then as we have defined it we have an $s_m \in h$ such that $s = h_k$ with $k \equiv m(\text{mod } n)$ where $n = |h|$. Thus, we define $E_N(s, h_N, \mathbf{A})$ as

$$\begin{cases} \{s^* \in S_N : \exists s' \in E(rep(s, h_N), h, \mathbf{A}) \text{ w/ } s' \approx_N s^*\} & \text{if } s \in h \\ \emptyset & \text{if } s \notin h \end{cases} \quad (7.10)$$

What this function does is generate E_N from E of the representative of s from the original model. The general condition for something, x say, to be in $E_N(s, h_N, \mathbf{A})$ is for there to be

something in $E(s', h, \mathbf{A})$ that \approx_N -relates x where s' is the representative of s from h . We now have to check that this function fulfills its duties. To do that we need some supporting facts:

Lemma 7.1.16. 1 If $|h| = n$ and $s \in h$, then $f_h(s) \approx_N f_{h_N}(s)$.

2 If $|h| \in \mathbb{N}$, then $s_0 \approx_N s$ with $s_0 \in h$ and $s \in h_N$ only if $f_h(s_0) \approx f_{h_N}(s)$.

Proof. For 1, if $s \in h$ and $|h| = n$, then $s = s_m$ for $0 \leq m < n$. If $m < n - 1$, then $f_{h_N}(s_m) = f_h(s_m)$ and \approx_N is reflexive. If $m = n - 1$, then $f_{h_N}(s_m) = h_n$ and $n \equiv 0 \pmod{n}$ so $s_0 \approx_N f_{h_N}(s_m)$. But also $f_h(s_m) = f_h(s_{n-1}) = s_0$. So we have our result.

For 2, suppose that $s_0 \approx_N s$ with $s_0 \in h$ and $s \in h_N$ and h finite. Now there is $s_m \in h$ such that $h_m = s_m$ and $s = h_k$ where $k \equiv m \pmod{n}$, $|h| = n$. Thus $k + 1 \equiv m + 1 \pmod{n}$, and that means $f_{h_N}(s_m) \approx_N f_{h_N}(s) = h_{k+1}$ by definition. But $f_{h_N}(s_m) \approx_N f_h(s_m)$ also by construction. Since $s_0 \approx_N s$ and $s \approx_N s_m$, $s_0 \approx_N s_m$ and by the conservativity of \approx_N over \approx , $s_0 \approx s_m$. Thus, by lemma 7.1.15, $f_h(s_0) \approx f_h(s_m)$, but that means $f_h(s_0) \approx_N f_h(s_m)$, so $f_h(s_0) \approx_N f_{h_N}(s)$, by the transitivity of \approx_N . \square

Now we can show that E_N obeys the conditions. In what follows we make no distinction between $x \approx_N y$ and $y \approx_N x$ since the relation is symmetric. The same goes for \approx .

- (a) if $s \notin h_N$, then $E_N(s, h, \mathbf{A}) = \emptyset$. In order for $E_N(s, h_N, \mathbf{A})$ to be non-empty $E(s', h, \mathbf{A})$ would have to be non-empty for the representative of s , s' , but if $s \notin h_N$ then s has no representative in h_N .
- (b) if $s' \in E_N(s, h_N, \mathbf{A})$, then $s' \in \text{lub}(s)$: Suppose that $s' \in E_N(s, h_N, A)$. So there is $s^* \in h$ (from the original model) s.t. $s^* \approx_N s$ and $s'' \in E(s^*, h, A)$ (in the original model) with $s'' \approx_N s'$. Since $E(s^*, h, A) \subseteq \text{lub}(s^*)$, $b_{h'}(s'') \approx s^*$. So then $b_{h'}(s'') \approx_N s^*$ by the definition of \approx_N , and since \approx_N is transitive, $b_{h'}(s'') \approx_N s$. Now $s'' \approx_N s'$ so $b_{h'_N}(s'') \approx_N b_{h_N}(s')$. Therefore, $b_{h_N}(s') \approx_N s$ and $f_{h_N}(b_{h_N}(s')) = s'$, thus $s' \in \text{lub}(s)$.
- (c) if $s \in h_N$, $\text{lub}(s, h_N) \in E_N(s, h_N, \mathbf{A})$: If $s \in h_N$, then $\text{lub}(s, h_N) = f_{h_N}(s)$. We have the two cases where h is finite or infinite. Let's suppose h is infinite. Then the representative

of s is s , so for any object x , $x \in E_N(s, h_N, \mathbf{A})$ iff there is $y \in E(s, h, \mathbf{A})$ and $y \approx_N x$. Since E obeys c, $\text{lub}(s, h) \in E(s, h, \mathbf{A})$. Because \approx_N is reflexive, $\text{lub}(s, h) \approx_N \text{lub}(s, h) = \text{lub}(s, h_N)$, and so $\text{lub}(s, h_N) \in E_N(s, h_N, \mathbf{A})$.

For the next case assume $|h| = n$. Now we know that there is $s_m \in h$ which is the representative of s such that $s_m \approx_N s$ and $s = h_k$ with $k \equiv m(\text{mod } n)$. We also have for any object x , $x \in E_N(s, h_N, \mathbf{A})$ iff there is $y \in E(s_m, h, \mathbf{A})$ and $y \approx_N x$. Since E obeys c we have that $f_h(s_m) = \text{lub}(s_m, h) \in E(s_m, h, \mathbf{A})$. But we know by lemma 7.1.16(2) $f_h(s_m) \approx_N f_{h_N}(s)$. Thus, $f_{h_N}(s) = \text{lub}(s, h_N) \in E_N(s, h_N, \mathbf{A})$.

(d) $E_N(s, h_N, \emptyset) = \text{lub}(s)$. For this we already have that $E_N(s, h_N, \emptyset) \subseteq \text{lub}(s)$ from condition b, so we assume that $x \in \text{lub}(s)$. That means that there are h'_N and $s' \in h'_N$ such that $s \approx_N s'$ and $f_{h'_N}(s') = x$. From here there are four cases:

Case 1: $|h|, |h'| \notin \mathbb{N}$. Here $h_N = h$ and $h' = h'_N$. So $x \in h'$ and $s \in h$, so $x \in E(s, h, \emptyset)$. So simply from the definition of E_N , $x \in E_N(s, h_N, \emptyset)$.

Case 2: $|h| \in \mathbb{N}, |h'| \notin \mathbb{N}$. Here $h' = h'_N$, so $f_{h'_N}(s') = f_{h'}(s') = x$ and $s' \in h'$. Now there is $s_m \in h$ such that $s \approx_N s_m$ and $s = h_k$ with $k \equiv m(\text{mod } n)$ where $|h| = n$. By the transitivity of \approx_N , $s_m \approx_N s'$ and so by conservativity, $s_m \approx s'$. Since $E(s_m, h, \emptyset) = \text{lub}(s_m)$ in the original model, and $x \in \text{lub}(s_m)$ of the original model, so $x \in E(s_m, h, \emptyset)$. Since \approx_N is reflexive, $x \approx_N x$ so by the second clause of the definition 7.10, we have $x \in E_N(s, h_N, \emptyset)$.

Case 3: $|h| \notin \mathbb{N}, |h'| \in \mathbb{N}$. Here $h = h_N$, so $s \in h$. There must be $s^* \in h'$ such that $s' \approx_N s^*$, so $s^* \approx s$ by conservativity. Now by lemma 7.1.16(2) $f_{h'}(s^*) \approx_N f_{h'_N}(s') = x$. But then, since $E(s, h, \emptyset) = \text{lub}(s)$ in the original model and $f_{h'}(s^*) \in \text{lub}(s)$, $f_{h'}(s^*) \in E(s, h, \emptyset)$. So there is something in $E(s, h, \emptyset)$ whose predecessor relates to s , and that relates to x . Thus, $x \in E_N(s, h_N, \emptyset)$.

Case 4: $|h|, |h'| \in \mathbb{N}$. Since both are finite, $|h| = n$ and $|h'| = n'$ and there must be $s_m \in h$ and $s_l \in h'$ such that $s = h_k$ and $k \equiv m(\text{mod } n)$, while $s' = h'_{k'}$ such

that $k' \equiv l \pmod{n'}$. That means $h'_{k'+1} = x = f_{h'_N}(s')$, with $h'_{k'+1} \approx_N f_{h'_N}(s_l)$. But we know also that $s_m \approx_N s_l$ by transitivity since $s \approx_N s_m$ and $s' \approx_N s_l$, so by conservativity we have $s_m \approx s_l$. Since $E(s_m, h, \emptyset) = \text{lub}(s_m)$, $f_{h'}(s_l) \in E(s_m, h, \emptyset)$, and we know $f_{h'_N}(s_l) \approx_N f_{h'}(s_l)$. So that means there is something in h , namely s_m , that \approx_N -relates to s and that there is something in $E(s_m, h, \emptyset)$ that \approx_N -relates to x . Therefore, $x \in E_N(s, h_N, \emptyset)$.

- (e) if $s \in h_N$, then $E_N(s, h_N, \mathbf{Ag}) = \{s' : s' \approx_N \text{lub}(s, h_N)\}$. From c we know that $\text{lub}(s, h_N) \in E_N(s, h_N, \mathbf{Ag})$. Now we have two cases, one where h is finite, and the other where it isn't.

Suppose $|h| = n$, then we know that there is $h_k = s$ and $s_m \in h$ such that $k \equiv m \pmod{n}$. So $h_{k+1} = \text{lub}(s, h_N) = f_{h_N}(s)$. We also know that by construction $f_h(s_m) \approx_N f_{h_N}(s_m)$, and $f_{h_N}(s_m) \approx_N f_{h_N}(s)$. And we further know that $E(s_m, h, \mathbf{Ag}) = \{y \in S : \text{lub}(s_m, h) \approx y\}$, with $\text{lub}(s_m, h) = f_h(s_m)$.

Now suppose that $x \in \{s' : s' \approx \text{lub}(s, h_N)\}$, so $x \approx_N f_{h_N}(s)$. There is $s' \in h$ (namely s_m) that is \approx_N -related to s and $s'' \in E(s_m, h, \mathbf{Ag})$ (namely $f_h(s_m)$), such that $s'' \approx_N x$. That means $x \in E_N(s, h_N, \mathbf{Ag})$.

Conversely, suppose $x \in E_N(s, h_N, \mathbf{Ag})$. Then $s_m \approx_N s$ and there is $s'' \in E(s_m, h, \mathbf{Ag})$ with $s'' \approx_N x$. Since $s'' \in E(s_m, h, \mathbf{Ag})$, $s'' \approx \text{lub}(s_m, h) = f_h(s_m)$, and because $s \approx_N s_m$ and h is finite we know $f_h(s_m) \approx_N f_{h_N}(s) = \text{lub}(s, h_N)$. So by transitivity of \approx_N , we have $s'' \approx_N \text{lub}(s, h_N)$. Thus $x \approx_N \text{lub}(s, h_N)$ as we wanted.

If h is infinite, then $h = h_N$. If $x \approx_N \text{lub}(s, h_N) = f_h(s)$, then since $h = h_N$. Now there is $s' \in h'$ (where $x \in h'_N$) such that $s' \approx_N f_h(s)$ so by conservativity $s' \approx f_h(s)$. Since $\text{lub}(s, h_N) = f_h(s) = \text{lub}(s, h) \in E(s, h, \mathbf{Ag})$, $x \approx_N f_h(s)$. So $x \in E_N(s, h_N, \mathbf{Ag})$.

Conversely, If $x \in E_N(s, h_N, \mathbf{Ag})$, then there is $s' \in E(s, h, \mathbf{Ag})$ since $h = h_N$ s.t. $s' \approx_N x$. Now $s' \approx \text{lub}(s, h)$ since E obeys e and that means $s' \approx_N \text{lub}(s, h_N)$, and by transitivity $x \approx_N \text{lub}(s, h_N)$.

- (f) if $\mathbf{A} \subsetneq \mathbf{B}$, then $E_N(s, h_N, \mathbf{B}) \subseteq E_N(s, h_N, \mathbf{A})$. Now if we assume the condition, then suppose that $x \in E_N(s, h_N, \mathbf{B})$, we know by definition that the representative of s , call it s' , from h is such that there is $y \in E(s', h, \mathbf{B})$ with $y \approx_N x$. Since E obeys f, y will be in $E(s', h, \mathbf{A})$, but that means $x \in E_N(s, h_N, \mathbf{A})$.
- (g) For all \mathbf{A}, \mathbf{B} $(s, h), (s', h'), (s'', h'')$, if $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $s' \approx_N s \approx_N s''$, then there is (s''', h''') such that $s''' \approx_N s$ with $E_N(s''', h''', \mathbf{A})$ and $E_N(s''', h''', \mathbf{B})$ contained in $E_N(s', h'_N, \mathbf{A})$ and $E_N(s'', h''_N, \mathbf{B})$, respectively. Now suppose that $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $s' \approx_N s \approx_N s''$ where $s' \in h'_N, s'' \in h''_N$ and $s \in h_N$. That means that there is $s' \approx_N s_1 \in h', s'' \approx_N s_2 \in h''$ which are the representatives s' and s'' . By the transitivity of \approx_N , $s_1 \approx_N s_2$ so these relations hold in the original model by conservativity. But since E obeys g, we have a $s''' \in h'''$ such that $E(s''', h''', \mathbf{A}) \subseteq E(s_1, h', \mathbf{A})$ and $E_N(s''', h''', \mathbf{B}) \subseteq E(s_2, h'', \mathbf{B})$. Here we will do the case for \mathbf{A} since the other is symmetric. Suppose that $x \in E_N(s''', h''', \mathbf{A})$ then there is $y \in E(s''', h''', \mathbf{A})$ such that $y \approx_N x$. That means $y \in E(s_1, h', \mathbf{A})$ and $y \approx_N x$ and so by definition $x \in E_N(s', h'_N, \mathbf{A})$. We also note that since $s''' \approx s_1, s''' \approx_N s_1$ so $s''' \approx_N s'$, and so $s''' \approx_N s$.

Finally we can define v_N for the new model. The new valuation we define as

$$v_N(\mathbf{p}) = \{s^* \in S_N : \exists s' \in v(\mathbf{p}) \ \& \ s^* \approx_N s'\} \quad (7.11)$$

We can see pretty quickly that if $s \in v_N(\mathbf{p})$ and $s \approx_N s'$, then $s' \in v_N(\mathbf{p})$. Suppose that $s \in v_N(\mathbf{p})$, then there is $s^* \in v(\mathbf{p})$ such that $s^* \approx_N s$. If $s \approx_N s'$, then by transitivity, $s^* \approx_N s'$, thus $s' \in v_N(\mathbf{p})$. Therefore \mathfrak{M}_N is a disjoint neutral model for \mathcal{L} , but each history is an infinite IDDLF. Now what we will claim is that

Proposition 7.1.17. *If \mathfrak{M} is a disjoint neutral model, then the disjoint neutral model \mathfrak{M}_N just defined is such that for all (s, h_N) in \mathfrak{M}_N , and ψ from \mathcal{L} , if $s^* = rep(s, h_N)$, then*

$$\mathfrak{M}_N, (s^*, h_N) \models \psi \iff \mathfrak{M}_N, (s, h_N) \models \psi$$

Proof. We proceed by induction on the complexity of ψ . For the base case, suppose that $(s, h_N) \models \mathbf{p}$. Then $s \in v_N(\mathbf{p})$ by definition. Now notice that $s^* \approx_N s$ by our observation above. By definition of v_N , then there is $s' \in v(\mathbf{p})$ such that $s \approx_N s'$, but that means $s' \approx_N s^*$ by transitivity. Thus, $s^* \in v_N(\mathbf{p})$, i.e., $(s^*, h_N) \models \mathbf{p}$.

Now suppose that $s^* \in v_N(\mathbf{p})$. Since by definition $s^* \approx_N s$, we have immediately, $s \in v_N(\mathbf{p})$.

IH: for all (s, h_N) in \mathfrak{M}_N , where $\text{rep}(s, h_N) = s^*$ and φ of less complexity than ψ from \mathcal{L} , $\mathfrak{M}_N, (s, h_N) \models \varphi \iff \mathfrak{M}_N, (s^*, h_N) \models \varphi$

Now there are two situations that we have to consider, if h is infinite, and when it isn't. If it is infinite, then $\text{rep}(s, h_N) = s$ so all cases follow because they are identical. So we assume that h is finite, say of size n .

Suppose that $(s, h_N) \models X\varphi$. Now suppose that $s^* = s_m$ for $0 \leq m \leq n-1$. So $s = h_k$ and $k \equiv m \pmod{n}$. This means that $h_{k+1} = f_{h_N}(s)$ and $(h_{k+1}, h_N) \models \varphi$. Now there are two cases to consider: 1) $f_{h_N}(s_m) \in h$ in which case $f_{h_N}(s_m) = f_h(s_m) = s_{m+1}$. If not 2), then $f_{h_N}(s_m) = h_{m+1}$, and $f_h(s_m) = s_0$.

If 1), then $k+1 \equiv m+1 \pmod{n}$. So we have $(h_{k+1}, h_N) \models \varphi$, and by IH, since $\text{rep}(h_{k+1}, h_N) = s_{m+1}$, $(s_{m+1}, h_N) \models \varphi$. That means $(s_m, h_N) \models X\varphi$.

If 2), then $\text{rep}(h_{k+1}, h_N) = s_0$. But we also have that $\text{rep}(f_{h_N}(s_m), h_N) = s_0$, this means that $(s_0, h_N) \models \varphi$ by IH, so again by IH $(f_{h_N}(s_m), h_N) \models \varphi$. But that means $(s_m, h_N) \models X\varphi$.

Now conversely, $(s_m, h_N) \models X\varphi$. Again we have the same two cases. If $f_{h_N}(s_m) = s_{m+1}$, i.e., is in h , then $\text{rep}(h_{k+1}, h_N) = s_{m+1}$, so by assumption $(s_{m+1}, h_N) \models \varphi$. By IH, $(h_{k+1}, h_N) \models \varphi$, which means $(h_k, h_N) = (s, h_N) \models X\varphi$.

Now suppose that $f_{h_N}(s_m) \notin h$, then again $\text{rep}(f_{h_N}(s_m), h_N) = s_0$ so by IH $(s_0, h_N) \models \varphi$. This means that $(h_{k+1}, h_N) \models \varphi$. So again $(h_k, h_N) = (s, h_N) \models X\varphi$.

The P case is symmetric to the X one so we omit it. For the $[\mathbf{A} \text{ xstit}] \varphi$ case, it follows because $E_N(s, h_N, \mathbf{A}) = E_N(s^*, h_N, \mathbf{A})$.

Suppose that $(s, h_N) \models \Box\varphi$. Then let $s' \approx_N s^*$. Since $\text{rep}(s, h_N) = s^*$, $s \approx_N s^*$ so

$s' \approx_N s$, which means that $(s', h'_N) \models \varphi$, by supposition. Since s' was arbitrary, $(s^*, h_N) \models \Box\varphi$.

Now suppose that $(s^*, h_N) \models \Box\varphi$. If $s \approx_N s'$, then of course $s^* \approx_N s'$ so $(s', h_N) \models \varphi$. That means $(s, h_N) \models \Box\varphi$ since s' was arbitrary. \square

Now if we consider a disjoint neutral model \mathfrak{M} , and (s, h) from \mathfrak{M} , then (s, h_N) is in \mathfrak{M}_N . That means $rep(s, h_N) = s$. But what will follow is that

Proposition 7.1.18. *If \mathfrak{M} is a disjoint neutral model, then the disjoint neutral model \mathfrak{M}_N just defined is such that for all (s, h) in \mathfrak{M} , and φ from \mathcal{L}*

$$\mathfrak{M}, (s, h) \models \varphi \iff \mathfrak{M}_N, (s, h_N) \models \varphi$$

Proof. This follows from the previous proposition because each such (s, h) is its own representative. \square

Since we can see that each history in \mathfrak{M}_N is an infinite IDDLF, so it looks like a copy of \mathbb{Z} , call it an **NUZ** model, and refer to that class of models as **NUZ**. So we can conclude that

Theorem 7.1.19. *For all $\Gamma, \varphi \in \mathcal{L}$, $\Gamma \vdash_{xp} \varphi$ iff $\Gamma \models_{\mathbf{NUZ}} \varphi$, i.e., \vdash_{xp} is complete with respect to **NUZ**.*

Proof. Since $\mathbf{NUZ} \subseteq \mathbf{NU}$, we have soundness for \vdash_{xp} . By the completeness theorem (theorem 7.1.13) we know that \vdash_{xp} is complete with respect to the class of disjoint neutral models. Now if $\Gamma \not\models_{\mathbf{NU}} \varphi$, then there is a disjoint **NU** model \mathfrak{M} and (s, h) in \mathfrak{M} such that $\mathfrak{M}, (s, h) \models \Gamma$, but $\mathfrak{M}, (s, h) \not\models \varphi$. But then by proposition 7.1.18, we have $\mathfrak{M}_N \in \mathbf{NUZ}$, and \mathfrak{M}_N agrees with \mathfrak{M} for each (s, h) in $|\mathfrak{M}|$. Therefore, $\mathfrak{M}_N, (s, h_N) \models \Gamma$ and $\mathfrak{M}_N, (s, h_N) \not\models \varphi$. So, $\Gamma \models_{\mathbf{NU}} \varphi$ iff $\Gamma \models_{\mathbf{NUZ}} \varphi$. Thus, $\Gamma \vdash_{xp} \varphi$ iff $\Gamma \models_{\mathbf{NUZ}} \varphi$. \square

7.1.5 The Same Old Models

Where we started off was with a collection of models where all of the histories which have some point in common share a common past. However, what we have noticed is that the canonical

model for the logic \vdash_{xp} isn't like that. But we provided a more general class of models that \vdash_{xp} is complete with respect to, and in fact we have shown that it is complete with respect to a smaller class of models. Namely, those in which each history is an infinite IDDLF. What we would like to show now is that \vdash_{xp} is actually complete with respect to the class of models defined in definitions 5.1.1 and 5.1.2. The models defined in those definitions, as we have noted above, are among those in **NU**, but the relation \approx from definition 7.1.4 has to be identity, and each history must be infinite. Ensuring that all of the histories are infinite has been done, but in those models all of the histories are disjoint, a far cry from \approx being identity.

The big problem, and the one that faced us in the completeness proof using canonical models was that the natural way to make \approx^o into identity is to take the members of S^o to be equivalence classes of maximal sets rather than the maximal sets themselves. But there is the possibility of a line looping back on itself, or simply having a finite sequence of maximal sets that repeat the same collection of necessary truths about the past every n steps towards the past. So the equivalence classes of maximal sets would generate histories that loop. Models with those kinds of histories are certainly not regular. We will call these kinds of histories periodic.

Definition 7.1.11. Let h be a history from a model $\mathfrak{M} = \langle S, H, E, \approx, v \rangle$ from **NU**, then we say that

- 0 h has a period of $n \in \mathbb{N}$ at $s \in h$ iff $b_h^n(s) \approx s$.
- 1 h is periodic iff there is an $s \in h$ and $n \in \mathbb{N}$ such that $b_h^n(s) \approx s$, there is some $n \in \mathbb{N}$ such that h has a period of n .
- 2 h is partially periodic iff h is periodic, but there is $s \in h$ such that for all $n \in \mathbb{N}$, $f_h^n(s) \not\approx s$.
- 3 h is totally periodic iff for all $s \in h$, there is $n \in \mathbb{N}$ such that $f_h^n(s) \approx s$.

The kinds of histories defined by this definition have characteristic shapes. A periodic history is one that has a repeating past. A partially periodic history is one that has a repeating past,

but has, after some point, a future that doesn't repeat. A totally periodic history is one that repeats perpetually. If we consider totally periodic histories they are equivalent to finite histories, although they needn't be finite. Non-periodic histories are those that never have a repeating past at any point. Now what of the relationships between these kinds of histories. We catalogue these relationships in the following easily proved list of observations

Observation 7.1.20. *Let h and h' be histories from a model \mathfrak{M} from \mathbf{NU} , then:*

- 0 *If there is $s \in h$ and $s' \in h'$ with $s \approx s'$, and h has a period of n at s , then h' has a period of n at s' .*
- 1 *If there is $s \in h$ and h is partially periodic, then there is a first s for which h isn't periodic. I.e., there is an $s \in h$ where for all $n \in \mathbb{N}$ $f_h^n(s) \not\approx s$, and for all $m \in \mathbb{N}$, there is $k \in \mathbb{N}$ such that $f_h^k(b_h^m(s)) \approx b_h^m(s)$.*
- 2 *If h is periodic, then there is a least $n \in \mathbb{N}$ such that there is $s \in h$ where h has a period of n at s .*
- 3 *If h has a period of n at s and that is its least period, then for every $m \in \mathbb{N}$, h has a period of n at $b_h^m(s)$.*
- 4 *If h is partially periodic with a least period of k , and $s \in h$ is the least s such that it isn't periodic after that s , then h has a period of k at s .*
- 5 *If h is partially periodic, $s \in h$ is the least s such that h isn't periodic after that s , and the least period of h is n , then for all $k \geq n$ and $m > 0$, $b_h^m(f_h^k(s)) \not\approx f_h^k(s)$.*
- 6 *If h is partially periodic, s is the least s such that h isn't periodic after that s , and there are $s_0 \in h$ and $s_1 \in h'$ such that $s_0 \approx s_1$, then there is $s^* \in h'$ such that $s \approx s^*$.*

Proof. The first four observations follow from condition H2 from definition 7.1.4 and from observations 7.1.14 and 7.1.15.

The fifth observation follows because h has a period of k , and s is the least such state such that for all n , $f_h^n(s) \not\approx s$. So for all $0 < m \in \mathbb{N}$, there is $l \in \mathbb{N}$ such that $f_h^l(b_h^m(s)) \approx b_h^m(s)$.

But that means $f_h^k(b_h^k(s)) \approx b_h^k(s)$. However, $f_h^k(b_h^k(s)) = s$ so $b_h^k(s) \approx s$ by transitivity of \approx , so h has a period of k at s .

The sixth follows because s marks the beginning of the last period of h so after the n th successor of s , nothing from before the n th successor can relate to it or anything after it least another period start. The 7th follows because, if n is the least period for h , then there is $s' \in h$ such that $s_0 \approx s'$ and $f_h^k(s) = s'$ for some $k < n$. That is true because s_0 must relate to some state in the last period of h . But then $s = b_h^k(f_h^k(s)) \approx b_h^k(s_0) \approx b_{h'}^k(s_1)$. \square

Now we can draw some conclusions from these observations. If \mathfrak{M} is in **NUZ**, then it is clumpy, by which we mean that the set of histories can be divided into clumps, or partitioned. That is we could say that h and h' are in the same clump iff there is $s \in h$ and $s' \in h'$ such that $s \approx s'$. So in a sense \approx partitions H and not just S . Also, because of condition b on E the effectivity function, we have that if $s \in E(s', h', \mathbf{A})$, then s is in a history in the same clump as h' . So that means \vdash_{xp} is complete with respect to the class of **NUZ** models that has only one clump of histories. This is much like how S5 is complete with respect to the class of equivalence relation and the class of universal relations.

So now let's restrict our attention to the class of disjoint **NUZ** models that have only one clump. We can call these models *universal NUZ* models. We must also notice that if a history is periodic, then every history in that clump is periodic with the same minimal period. If there is a partially periodic history in the clump that doesn't mean that every history in the clump is partially periodic, there can be totally periodic histories in there too. But all of the periodic histories have the same minimal period. This follows from observation 7.1.20(0) and condition H2.

Now if we have a model \mathfrak{M} that is in our particular restricted class that we are considering, and it has the property that

$$\forall s, s' \in h \in H, s \approx s' \text{ only if } s = s' \quad (7.12)$$

we say it has the **Kamp Property** (Kamp for short)² What the Kamp property amounts to is that no histories in our universal model are periodic. Non-periodicity follows from the contrapositive of the Kamp property. So that is an equivalent formulation for a model to satisfy the Kamp property, for all histories to be non-periodic. Of course if even one history is non-periodic in a universal model, they all must be.

It is shown in Reynolds (2002) that for a different—but similar—language Models that satisfy the Kamp property are elementarily equivalent to those given in definition 5.1.1. We shall show, by the same sort of argument, that **NUZ** models that satisfy the Kamp property, are elementarily equivalent to the regular \mathcal{L} -models from definition 5.1.1.

Given a model that satisfies the Kamp property, we can then construct another model as follows:

1. $S^* = S / \approx$,
2. H^* will be given as the triples $\langle [h], f_{[h]}, b_{[h]} \rangle$ where
 - (a) $[h] = \{ [s] : s \in h \}$
 - (b) $f_{[h]}([s]) = [s']$ iff there is $s_0 \in h$ such that $s_0 \approx s$ and $f_h(s_0) \approx s'$
 - (c) $b_{[h]}([s]) = [s']$ iff there are $s_0 \in h$ such that $s_0 \approx s$ and $b_h(s_0) \approx s'$
3. $E^*([s], [h], \mathbf{A})$ is the set of $[s'] \in S^*$ such that there is $s_0 \in h$ with $s_0 \approx s$ and $s_1 \approx s'$ with $s_1 \in E(s_0, h, \mathbf{A})$.
4. $[s] \approx^* [s']$ iff $s \approx s'$ which means \approx^* is $=$, and
5. $v^*(\mathbf{p}) = \{ [s] \in S^* : s \in v(\mathbf{p}) \}$

From observation 7.1.2, we know that an infinite IDDLF is \mathbb{Z} -like, and that means that as long as each $\langle [h], f_{[h]}, b_{[h]} \rangle$ is an infinite IDDLF, and obeys H1 and H2, we will have that the structure satisfies conditions 2.a and 2.b from definition 5.1.1. Notice too that if $s \in h$ from

²This is because it defines what is elsewhere called a Kamp Model, see Thomason (1984), and Reynolds (2002).

\mathfrak{M} , then since $f_h(s) \approx f_h(s)$, $f_{[h]}([s]) = [f_h(s)]$. Next we check the conditions for an infinite injective DDLF in

Observation 7.1.21. *Here we show that each $\langle [h], f_{[h]}, b_{[h]} \rangle \in H^*$, is an infinite injective DDLF.*

Proof. The first thing to notice is that if $[s] \neq [s']$ are in $[h]$, then there are $s_0 \approx s$ and $s_1 \approx s'$ from h , and either $f_h^+(s_0) = s_1$ or $f_h^+(s_1) = s_0$ so either $f_{[h]}^+([s]) = [s']$ or $f_{[h]}^+([s']) = [s]$. So it obeys TRI. We can see that $f_{[h]}$ is a function since if $[s] = [s']$, and $f_{[h]}([s]) = [s^*]$, there is $s_0 \in h$ with $s_0 \approx s$ and $f_h(s_0) \approx s^*$. Since $s_0 \approx s$ and $[s'] = [s]$, $s_0 \approx s'$. So we have $f_{[h]}([s']) = f_{[h]}([s])$.

For injectivity, suppose that $f_{[h]}([s']) = f_{[h]}([s]) = [s^*]$. Then there are $s_0 \approx s$ and $s_1 \approx s'$ both in h such that $f_h(s_0) \approx s^*$ and $f_h(s_1) \approx s^*$. But that means $f_h(s_0) \approx f_h(s_1)$, and so by Kamp, $f_h(s_0) = f_h(s_1)$, and since f_h is injective, we have $s_0 = s_1$. Thus, $[s] = [s']$.

We can UFIX is satisfied since if $f_{[h]}([s]) = [s]$, and $f_{[h]}([s']) = [s']$, then there are $s_0 \in h$ and $s_1 \in h$ such that $s_0 \approx s$ and $s_1 \approx s'$ such that $f_h(s_0) \approx s$ and $f_h(s_1) \approx s'$. But that means $f_h(s_0) \approx s_0$ and $f_h(s_1) \approx s_1$. By Kamp we must have $f_h(s_0) = s_0$ and $f_h(s_1) = s_1$. Since f_h obeys UFIX, $s_1 = s_0$. So $[s] = [s']$.

All of the above will be basically the same for $b_{[h]}$.

Now suppose that $|[h]| \geq 2$, and $f_{[h]}([s]) = [s]$, so there is $s_0 \in h$ such that $f_h(s_0) = s_0$ as above. So by SIZE for f_h , since $|h|$ must be at least 2, $b_h(s_0) \neq s_0$. If $b_{[h]}([s]) = [s]$, then there is $s'_0 \in h$ with $s'_0 \approx s$ and $b_h(s'_0) \approx s$. But by Kamp that means $b_h(s'_0) = s'_0$. That also means that $s_0 \approx s \approx s'_0$ so again by Kamp $s_0 = s'_0$. So we would have $b_h(s_0) = s_0$, a contradiction. Thus, $b_{[h]}([s]) \neq [s]$.

Now for CONV1, the CONV2 case is similar. Let $[s] \neq [s']$, and $f_{[h]}([s]) = [s']$. So there is $s_0 \in h$ with $f_h(s_0) \approx s'$ and $s_0 \approx s$. $f_h(s_0) \neq s_0$ since if it did $s_0 \approx s'$ and that would contradict $[s] \neq [s']$. So $b_h(f_h(s_0)) = s_0$. That means $b_h(f_h(s_0)) \approx s_0$. So now there is an $s_0^* \in h$ such that $[s_0^*] = f_{[h]}([s])$, namely $f_h(s_0)$. So $b_{[h]}(f_{[h]}([s])) = [s]$.

The fact that $|[h]|$ must be infinite follows because if $[h]$ was finite, then there would have to be some $s \approx s'$ both in h , and not equal. But that is impossible. \square

Of course, that the functions are well defined is standard fare with such quotient structures. The 2.a condition is clearly satisfied, and so we can see the 2.b condition because, if $[s] \in [h] \cap [h']$, then there is $s_0 \in h$ and $s_1 \in h'$ such that $s_0 \approx s \approx s_1$. That means that for all n , $b_h^n(s_0) \approx b_{h'}^n(s_1)$ by H2. Now suppose that $b_{[h]}^n([s]) = [s']$. That means $b_{[h]}^n([s_0]) = [s']$, so by definition, $b_h^n(s_0) \approx s'$, but by transitivity of \approx , $b_{h'}^n(s_1) \approx s'$. So $b_{[h']}^n([s_1]) = [s']$, so $[s'] \in [h']$. The other direction is symmetric. The real work is of course is done in checking the conditions on effectivity functions.

Observation 7.1.22. E^* obeys conditions a–g.

Proof. (a) if $[s] \notin [h]$, then there is no $s_0 \approx s$ such that $s_0 \in h$. If $[s'] \in E^*([s], [h], \mathbf{A})$, then there would have to be $s_0 \approx s$ and $s_0 \in h$, contrary to assumption. So $E^*([s], [h], \mathbf{A}) = \emptyset$.

(b) Suppose that $[s'] \in E^*([s], [h], \mathbf{A})$. We want to show that $[s'] \in \text{lub}([s])$. So we need to find $[h']$ such that $[s] \in [h']$ and $f_{[h']}([s]) = [s']$. Thus we need $s_1 \in h'$ such that $s_1 \approx s$ and $f_{h'}(s_1) \approx s'$. From our assumption there is $s'_0 \in h$ such that $s'_0 \approx s$ and $s^* \approx s'$ with $s^* \in E(s'_0, h, \mathbf{A})$. That means $s^* \in \text{lub}(s'_0)$ since E obeys b, and that means there is h' and $s_1 \in h'$ with $s_1 \approx s'_0$ where $f_{h'}(s_1) = s^*$. So $f_{h'}(s_1) \approx s'$, and we are done.

(c) Suppose $[s] \in [h]$. Now by definition $\text{lub}([s], [h]) = f_{[h]}([s])$ is defined by our assumption, so suppose it is $[s']$. That means there is $s_0 \in h$ with $s_0 \approx s$ and $f_h(s_0) \approx s'$. Since E obeys c, $f_h(s_0) \in E(s_0, h, \mathbf{A})$. Thus $\text{lub}([s], [h]) = f_{[h]}([s]) \in E^*([s], [h], \mathbf{A})$.

(d) Suppose that $[s] \in [h]$. From b we know that $E^*([s], [h], \emptyset) \subseteq \text{lub}([s])$, so now assume that $[s'] \in \text{lub}([s])$. We want to show that $[s'] \in E^*([s], [h], \emptyset)$, so we need $s_0 \approx s$ such that $s_0 \in h$ and an $s_1 \approx s'$ in $E(s_0, h, \emptyset)$. From our first supposition there is $s_0 \approx s$ and $s_0 \in h$. By our assumption there is h' such that $f_{[h']}([s]) = [s']$. Thus there is $s'_0 \in h' \approx$ -related to s such that $f_{h'}(s'_0) \approx s'$. Since $s \approx s'_0$, we have that $s_0 \approx s'_0$, so $f_{h'}(s'_0) \in \text{lub}(s_0)$ by definition of lub , and that means, since E obeys d, that $f_{h'}(s'_0) \in E(s_0, h', \emptyset)$, which is the s_1 that we needed.

- (e) Suppose that $[s] \in [h]$. From c we know that $\text{lub}([s], [h]) \in E^*([s], [h], \mathbf{Ag})$. Now we want to show that that is the only element in that set. Recall that $\text{lub}([s], [h]) = f_{[h]}([s])$. Now assume that $[s'] \in E^*([s], [h], \mathbf{Ag})$. So there is $s_0 \in h$ with $s_0 \approx s$ and $s_1 \approx s'$ with $s_1 \in E(s_0, h, \mathbf{Ag})$. That means that $s_1 \approx \text{lub}(s_0, h) = f_h(s_0)$. What we want to show is that $f_{[h]}([s]) = [s']$, so we need $s^* \in h$ such that $s^* \approx s$ and $f_h(s^*) \approx s'$. Since $s_1 \approx s'$, let $s^* = s_0$. Thus, $\{\text{lub}([s], [h])\} = E^*([s], [h], \mathbf{Ag})$.
- (f) If $\mathbf{A} \subsetneq \mathbf{B}$, then $E^*([s], [h], \mathbf{B}) \subseteq E^*([s], [h], \mathbf{A})$, because E obeys f. Finally, we have,
- (g) Suppose that $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $[s] \in [h] \cap [h']$. That means that there is $s_0 \approx s \approx s_1$ such that $s_0 \in h$ and $s_1 \in h'$. Since E obeys g, we have $s'' \in h''$ such that $s'' \approx s$ and $E(s'', h'', \mathbf{A}) \subseteq E(s_0, h, \mathbf{A})$ and $E(s'', h'', \mathbf{B}) \subseteq E(s_1, h', \mathbf{B})$. We will do the \mathbf{B} case since the \mathbf{A} is the same. So we have $[s''] \in [h'']$ and $[s''] = [s]$, so $[s] \in [h'']$. Now suppose that $[s^*] \in E^*([s''], [h''], \mathbf{B})$. So there is $s_2 \in h''$ with $s_2 \approx s''$ and $s_3 \in E(s_2, h'', \mathbf{B})$ with $s_3 \approx s^*$. Since both of s_2 and s'' are in h'' , by Kamp we have that $s_2 = s''$. That means that $s_3 \in E(s'', h'', \mathbf{B})$, so $s_3 \in E(s_1, h', \mathbf{B})$. But recall that $s_1 \approx s$, so that is the condition for $[s^*] \in E^*([s], [h'], \mathbf{B})$.

□

From these observations we can see that \mathfrak{M}^* is a (perhaps non-universal) regular model, i.e., a (non-universal) neutral model where \approx is identity. We can then prove the following:

Theorem 7.1.23. *Let $\mathfrak{M} = \langle S, H, E, \approx, v \rangle$ be a disjoint, **NUZ** model that satisfies the Kamp property. If $\mathfrak{M}^* = \mathfrak{M} / \approx$ (the quotient model just defined), then for all $(s, h) \in |\mathfrak{M}|$ and $\varphi \in \mathcal{L}$,*

$$\mathfrak{M}, (s, h) \models \varphi \iff \mathfrak{M}^*, ([s], [h]) \models \varphi$$

Proof. We do this by induction on the complexity of φ . Clearly the base case holds. We will do the X , \square and $[\mathbf{A} \text{ xstit}]$ cases as examples since the rest are standard.

Suppose that $\mathfrak{M}, (s, h) \models X\psi$. So we have $\mathfrak{M}, (f_h(s), h) \models \psi$, so by IH, $\mathfrak{M}^*, ([f_h(s)], [h]) \models \psi$. But $[f_h(s)] = f_{[h]}([s])$, so we have $\mathfrak{M}^*, (f_{[h]}([s]), [h]) \models \psi$, and so by the truth condition

for X , \mathfrak{M}^* , $([s], [h]) \models X\psi$. Each of those steps was an equivalence.

Suppose that $\mathfrak{M}, (s, h) \models \Box\psi$. That happens iff for each s' such that $s' \approx s$, $\mathfrak{M}, (s', h') \models \psi$ where $s' \in h'$ (there is always only one h'). Any $[h']$ that contains $[s]$ will be such that there is $s^* \in h'$ and $s^* \approx s$, so $\mathfrak{M}^*, ([s'], [h']) \models \psi$ by IH. So this holds for any $[h']$; thus, $\mathfrak{M}^*, ([s], [h]) \models \Box\psi$. Each step here was actually an equivalence.

Now suppose that $\mathfrak{M}, (s, h) \models [\mathbf{A} \text{ xstit}] \psi$. So for each $s' \in E(s, h, \mathbf{A})$, where $s' \in h'$, $\mathfrak{M}, (s', h') \models \psi$. Suppose that $[s^*] \in E^*([s], [h], \mathbf{A})$. That means there is $s_0 \in h$ with $s_0 \approx s$ and $s_1 \in E(s_0, h, \mathbf{A})$ such that $s_1 \approx s^*$. Since both s_0 and s are in h , $s_0 = s$ by Kamp so $s_1 \in E(s, h, \mathbf{A})$. That means $\mathfrak{M}, (s_1, h_1) \models \psi$ so by IH, $\mathfrak{M}^*, ([s_1], [h_1]) \models \psi$. But $[s_1] = [s^*]$, so $[s^*] \in [h_1]$ and $\mathfrak{M}^*, ([s^*], [h_1]) \models \psi$. Therefore, $\mathfrak{M}^*, ([s], [h]) \models [\mathbf{A} \text{ xstit}] \psi$ since $[s^*]$ was arbitrary.

Conversely, suppose that $\mathfrak{M}^*, ([s], [h]) \models [\mathbf{A} \text{ xstit}] \psi$. So for each $[s^*] \in E^*([s], [h], \mathbf{A})$, $\mathfrak{M}^*, ([s^*], [h^*]) \models \psi$. Now let $s' \in E(s, h, \mathbf{A})$. That means $[s'] \in E^*([s], [h], \mathbf{A})$, and so by IH, $\mathfrak{M}, (s', h') \models \psi$ where $s' \in h'$. Since s' was arbitrary, $\mathfrak{M}, (s, h) \models [\mathbf{A} \text{ xstit}] \psi$. \square

This means that if a sentence fails in a Kamp model, it will fail in a regular model, i.e., \mathfrak{M}^* . We introduce a special name for models derived from these quotient structures: \mathfrak{M}/\approx . Now the problem facing us is that not all **NUZ** models satisfy Kamp; there are models that are periodic. What we will now argue is that given a model that is periodic where some argument fails, we can generate a Kamp model from it where that argument fails. For this we construct another model $\mathfrak{M}_K = \langle S_K, H_K, E_K, \approx_K, v_K \rangle$ as follows.

Let $\mathfrak{M} = \langle S, H, E, \approx, v \rangle$ be a universal and disjoint **NUZ** \mathcal{L} -model. So it is also disjoint and suppose that it doesn't satisfy Kamp; thus, it is periodic. Every history must be periodic, and those histories all have the same shortest period $n \in \mathbb{N}$ since there is only one clump (since the model is universal). Let $h \in H$ and $s \in h$. Now h is either partially or totally periodic. We will define a new *set* of histories $\{h_s : s \in h\}$ for each $h \in H$, where each h_s is defined as

follows:

$$\left\{ \begin{array}{c} \vdots \\ h_{s+2} \\ h_{s+1} \\ h_{s+0} = s \\ h_{s-1} \\ h_{s-2} \\ \vdots \end{array} \right.$$

where each $h_{s\pm i}$, $i \neq 0$, is new to S and doesn't overlap with any other $h'_{s'}$. Thus they are all new elements, and every $h'_{s'}$ is disjoint from every other. The idea of these new histories is that each history shifts by one. So, $s \in h_s$, but $s \notin h_{s'}$ when $s \neq s'$ but $s' \in h$. We will impose the following conventions that $h_{s-0} = h_{s+0} = s$. Often we will have to deal with separate cases where a static state is, whether it is above or below the element from the original model. If $x = h_{s+k}$ we will call it a positive case, if $x = h_{s-k}$ we will call it a negative case.

Now we define the triples $\langle h_s, f_{h_s}, b_{h_s} \rangle$. We define the functions $f_{h_s}(x)$ and $b_{h_s}(x)$ as

$$f_{h_s}(x) = \begin{cases} h_{s+(i+1)} & \text{if } i \in \mathbb{N} \text{ \& } x = h_{s+i} \\ h_{s-(i-1)} & \text{if } i \in \mathbb{N} \text{ \& } x = h_{s-i} \end{cases}$$

and

$$b_{h_s}(x) = \begin{cases} h_{s+(i-1)} & \text{if } i \in \mathbb{N} \text{ \& } x = h_{s+i} \\ h_{s-(i+1)} & \text{if } i \in \mathbb{N} \text{ \& } x = h_{s-i} \end{cases}$$

Each of these histories is isomorphic to $\langle \mathbb{Z}, z+1, z-1 \rangle$. Since each triple is an infinite IDDLF, b and f are inverses. So we will write $f_h^{-k}(s) = b_h^k(s)$ where $k \in \mathbb{N}$. This way $f_h^k(f_h^{-l}(s)) = f_h^{k-l}(s)$. We can picture the *lub* of an h_{s+k} as in figure 7.3.

Now we have to define the new parts S_K, H_K, \approx_K . What we do for H_K is simply take the set $\bigcup \{ \{ \langle h_s, f_{h_s}, b_{h_s} \rangle \}_{s \in h} : h \in H \}$. Then to get S_K we take $\bigcup_{s \in S} h_s$. Notice that $S \subseteq S_K$.

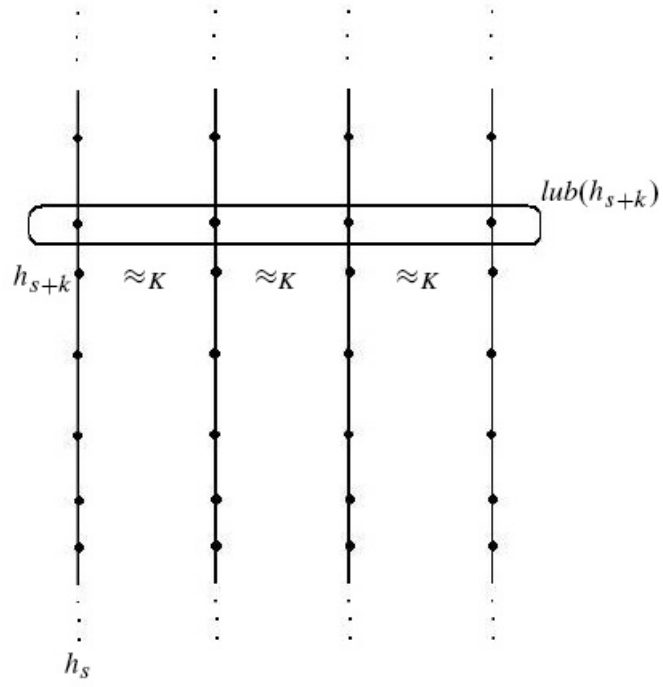
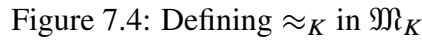


Figure 7.3: $lub(h_{s+k})$ in \mathfrak{M}_K


$$\left\{ \langle h_{s-j}, h'_{s'-k} \rangle : b_h^j(s) \approx b_{h'}^k(s') \text{ \& } j = k \right\} \cup \left\{ \langle h_{s+j}, h'_{s'+k} \rangle : f_h^j(s) \approx f_{h'}^k(s') \text{ \& } j = k \right\} \quad (7.13)$$

This relation will relate everything to itself since \approx is reflexive, and it relates all of the members of h_s and $h'_{s'}$ that come at the same intervals before s and s' in h_s and $h'_{s'}$, when $s \approx s'$. It also relates things that continue to be related after s and s' . The idea behind the new model

is to generate new histories, one for each $s \in h$, and $h \in H$. The new history for $s \in h$ is supposed to mimic the ordering and relations to other states and histories in \mathfrak{M} , but all from the perspective of s . Each new history only contains one static state from the original model. This will eliminate the possibility of a static state “looping” back on itself in the sense that $s \approx s'$ for $s, s' \in h$. However, we will keep the same relations between the static states from the original model.

Clearly, \approx_K is an equivalence relation since \approx is. And if $h_{s-j} \approx_K h'_{s'-k}$, then if they are different points, it must be that $j = k$ and $s \approx s'$.

Observation 7.1.24. H_K obeys condition H2.

Proof. Suppose that $h_{s-j} \approx_K h'_{s'-j}$, without loss of generality. Then let $m \in \mathbb{N}$, so we will have $b_{h_s}^m(h_{s-j}) = h_{s-(j+m)}$, and $h'_{s'-(j+m)} = b_{h'_{s'}}^m(h'_{s'-j})$. Also from our supposition we have $b_h^j(s) \approx b_{h'}^j(s')$. So by condition H2 for \approx , we have $b_h^{j+m}(s) \approx b_{h'}^{j+m}(s')$, and since $j + m = j + m$, $h_{s-(j+m)} \approx_K h'_{s'-(j+m)}$.

Suppose that $h_{s+j} \approx_K h'_{s'+j}$, without loss of generality. Then let $m \in \mathbb{N}$, so we will have $b_{h_s}^m(h_{s+j}) = h_{s+(j-m)}$, and $h'_{s'+(j-m)} = b_{h'_{s'}}^m(h'_{s'+j})$. Also from our supposition we have $f_h^j(s) \approx f_{h'}^j(s')$. So by condition H2 for \approx , we have $f_h^{j-m}(s) \approx f_{h'}^{j-m}(s')$, and since $j - m = j - m$, $h_{s+(j-m)} \approx_K h'_{s'+(j-m)}$. So condition H2 is satisfied. \square

Now we define the effectivity function E_K .

$$E_K(x, h_s, \mathbf{A}) = \begin{cases} \emptyset & \text{if } x \notin h_s \\ \left\{ h'_{s'+j} : f_{h'}^j(s') \in E(f_h^k(s), h, \mathbf{A}) \right\} \cap \text{lub}(h_{s+k}) & \text{if } x = h_{s+k} \\ \left\{ h'_{s'-j} : b_{h'}^j(s') \in E(b_h^k(s), h, \mathbf{A}) \right\} \cap \text{lub}(h_{s-k}) & \text{if } x = h_{s-k} \end{cases} \quad (7.14)$$

With this function we define it in cases because of where its input sits, either above or below s , determines what we are going to put into the result of the function. What it does is relate the E_K function to the E function from the original model. To explain how this works let's define

the representative of an element of the new model. Define the function rep as follows:

$$rep(x) = \begin{cases} (f_h^k(s), h) & \text{if } x = (h_{s+k}, h_s) \\ (b_h^k(s), h) & \text{if } x = (h_{s-k}, h_s) \end{cases} \quad (7.15)$$

The idea is that each element of the new model has a representative in the original model. We will abuse this notation by saying that sometimes $rep(s)$ is simply the static state coordinate. Notice that different x s will be assigned the same representative.

So the rationale behind E_K is that a static state $h'_{s'+j}$ is in $E_K(h_{s+k}, h_s, \mathbf{A})$, when the representative of $(h'_{s'+j}, h'_{s'})$, i.e., $f_{h'}^j(s')$ is in $E(f_h^k(s), h, \mathbf{A})$. However, that could lead to E_K making some wild jumps into the future, so we restrict it by requiring that only elements from the immediate future be admitted, i.e., we intersect E_K with $lub(h_{s+k})$. The same rationale goes for the negative case.

So now we must check that E_K has the conditions *a–g* of definition 7.1.4. We should notice a few things first.

Observation 7.1.25. *1. In general, if $s \approx s'$, then $lub(s) = lub(s')$. This will hold for \approx_K as well.*

2. If $h''_{s''+k} \in lub(h_{s+j})$, then $k = j + 1$. And if $h'_{s'-k} \in lub(h_{s-j})$, then $k = j - 1$. This also means that the only time h_{s+j} (a positive case) will be a possible successor to $h'_{s'-k}$ (a negative case) is when $j = 0$, and $k = 1$.

3. Finally, there is no way that h_{s-k} would ever be a successor of $h'_{s'+j}$.

Proof. For 1, it follows from the definitions immediately. For 2, consider the definition of $lub(h_{s+j}) =_{Df}$

$$\left\{ h'_{s'+\pm k} : \exists h'_{s'}, h_{s'+\pm l} \approx h_{s+j} \ \& \ f_{h'_{s'}}(h_{s'+\pm l}) = h_{s'+\pm k} \right\}$$

Now for $h_{s'+\pm l} \approx_K h_{s+j}$ to hold, $h_{s'+\pm l} = h'_{s'+l}$, and $l = j$. But then also $f_{h'_{s'}}(h_{s'+l}) = h_{s'+\pm k} = h_{s'+k} = h_{s'+(l+1)}$. So given that $l = j$, $k = j + 1$.

For the negative case consider the definition of $\text{lub}(h_{s+j}) =_{Df}$

$$\left\{ h'_{s' \pm k} : \exists h'_{s'}, h_{s' \pm l} \approx h_{s-j} \ \& \ f_{h'_{s'}}(h_{s' \pm l}) = h_{s' \pm k} \right\}$$

Now for $h_{s' \pm l} \approx_K h_{s-j}$ to hold, $h_{s' \pm l} = h'_{s' - l}$, and $l = j$. But then also $f_{h'_{s'}}(h_{s' - l}) = h_{s' \pm k} = h_{s' - k} = h_{s' - (l-1)}$. So given that $l = j$, $k = j - 1$.

For 3, it is just inconsistent with the construction. \square

So there are really two cases to consider in the next proofs: when both x states are positive cases or negative cases.

Lemma 7.1.26. E_K has the properties a–g of definition 7.1.4.

Proof. [a] If $x \notin h_s$, then $E_K(x, h_s, \mathbf{A}) = \emptyset$. This follows by definition of E_K

[b] If $x' \in E_K(x, h_s, \mathbf{A})$, then $x' \in \text{lub}(x)$. This also follows by definition of E_K .

[c] Suppose that $x = h_{s+k} \in h_s$. We want to show that $\text{lub}(h_{s+k}, h_s) \in E_K(h_{s+k}, h_s, \mathbf{A})$. By definition $\text{lub}(h_{s+k}, h_s) = f_{h_s}(h_{s+k}) = h_{s+(k+1)}$. Clearly, $h_{s+(k+1)} \in \text{lub}(h_{s+k})$, and since E obeys c we have that $f_h^{k+1}(s)$, the representative of $h_{s+(k+1)}$, is in $E(f_h^k(s), h, \mathbf{A})$, and that means $h_{s+(k+1)} \in E_K(h_{s+k}, h_s, \mathbf{A})$. In the negative case, i.e., $x = h_{s-k}$ we have the same situation since $f_h(b_h^k(s)) \in E(b_h^k(s), h, \mathbf{A})$, and $f_h(b_h^k(s)) = b_h^{k-1}(s)$ which is the representative of $h_{s-(k-1)}$.

[d] Suppose that $x = h_{s+k} \in h_s$, we want to show that $E_K(x, h_s, \emptyset) = \text{lub}(x)$. From b we know that $E_K(h_{s+k}, h_s, \emptyset) \subseteq \text{lub}(h_{s+k})$. So suppose that $h'_{s'+j} \in \text{lub}(h_{s+k})$. That means $j = k + 1$, as we noted in observation 7.1.25, and so $h'_{s'+(j-1)} = h'_{s'+(k)} \approx_K h_{s+k}$. That means by definition of \approx_K , that $f_{h'}^k(s') \approx f_h^k(s)$. And that means $f_{h'}^{k+1}(s') \in E(f_h^k(s), h, \emptyset)$ since E obeys d, and $f_{h'}^{k+1}(s') \in \text{lub}(f_h^k(s))$. That means $h'_{s'+j} \in E_K(h_{s+k}, h_s, \emptyset)$ since $f_{h'}^{k+1}(s')$ is the representative of $h'_{s'+j}$ ($k + 1 = j$). If $x = h_{s-k}$, then suppose that $h'_{s'-j} \in \text{lub}(h_{s-k})$. Here $j = k - 1$, and so $h'_{s'-k} \approx_K h_{s-k}$, thus $b_{h'}^k(s') \approx b_h^k(s)$. Then $f_{h'}(b_{h'}^k(s')) = b_{h'}^j(s') \in \text{lub}(b_h^k(s)) = E(b_h^k(s), h, \emptyset)$, and $b_{h'}^j(s')$ is the representative of $h'_{s'-j}$. So $h'_{s'-j} \in E_K(x, h_s, \emptyset)$. Therefore, $E_K(x, h_s, \emptyset) = \text{lub}(x)$, if $x \in h_s$.

[e] Suppose $x \in h_s$, and $x = h_{s+k}$. We want to show that

$E_K(h_{s+k}, h_s, \mathbf{Ag}) = \{h'_{s'+j} : h'_{s'+j} \approx_K \text{lub}(h_{s+k}, h_s)\}$. We know from c that $\text{lub}(h_{s+k}, h_s) = f_{h_s}(h_{s+k})$ and so $f_{h_s}(h_{s+k}) = h_{s+(k+1)} \in E_K(h_{s+k}, h_s, \mathbf{Ag})$. Now suppose that $h'_{s'+j} \in E_K(h_{s+k}, h_s, \mathbf{Ag})$. By b $h'_{s'+j} \in \text{lub}(h_{s+k}, h_s)$, so $j = k+1$ by observation 7.1.25(2). That means the representative of $h'_{s'+j} = h'_{s'+(k+1)}$, $f_{h'}^{k+1}(s')$ is a member of $E(f_h^k(s), h, \mathbf{Ag})$. Thus, $f_{h'}^{k+1}(s') \approx f_h^{k+1}(s) = \text{lub}(f_h^k(s))$ since $E(f_h^k(s), h, \mathbf{Ag}) = \{s' : s' \approx \text{lub}(f_h^k(s), h)\}$ because E obeys e. But that means $h'_{s'+j} = h'_{s'+(k+1)} \approx_K h_{s+(k+1)}$ by definition of \approx_K . Now suppose that $h'_{s'+j} \approx_K h_{s+(k+1)}$, and so $h'_{s'+j} = h'_{s'+(k+1)}$, and we have by definition of \approx_K , $f_{h'}^{k+1}(s') \approx f_h^{k+1}(s)$. That means $f_{h'}^k(s') \approx f_h^k(s)$ by H2 for \approx , and so $h'_{s'+(k+1)} \in \text{lub}(h_{s+k}, h_s)$. But it also means that the representative of $h'_{s'+j}$, $f_{h'}^{k+1}(s')$ is a member of $E(f_h^k(s), h, \mathbf{A})$. So $h'_{s'+j} = h'_{s'+(k+1)} \in E_K(h_{s+k}, h_s, \mathbf{Ag})$.

For the negative case, suppose that $x = h_{s-k}$. We know from c that $\text{lub}(h_{s-k}, h_s) = f_{h_s}(h_{s-k}) = h_{s-(k-1)} \in E_K(h_{s-k}, h_s, \mathbf{Ag})$. Now suppose that $h'_{s'-j} \in E_K(h_{s-k}, h_s, \mathbf{Ag})$. By observation 7.1.25(2), $j = k-1$ and $b_{h'}^{k-1}(s') \in E(b_h^k(s), h, \mathbf{Ag})$. Thus, $b_{h'}^{k-1}(s') = f_{h'}(b_h^k(s')) \approx f_h^{k+1}(s) = \text{lub}(f_h^k(s))$ since $E(b_h^k(s), h, \mathbf{Ag}) = \{s' : s' \approx \text{lub}(b_h^k(s), h)\}$ because E obeys e. But that means $h'_{s'-j} = h'_{s'-(k-1)} \approx_K h_{s-(k-1)}$, which is what we want. Now suppose that $h'_{s'-j} \approx_K h_{s-(k-1)}$, and so $h'_{s'-j} = h'_{s'-(k-1)}$, and we have by definition of \approx_K , $b_{h'}^{k-1}(s') \approx b_h^{k-1}(s)$. That means $f_{h'}(b_h^k(s')) \approx f_h(b_h^k(s))$ and so $h'_{s'-(k-1)} \in \text{lub}(h_{s-k}, h_s)$. But it also means that $b_{h'}^{k-1}(s') \in E(b_h^k(s), h, \mathbf{A})$ which is the representative of $h'_{s'-(k-1)}$. So $h'_{s'-j} = h'_{s'-(k-1)} \in E_K(h_{s-k}, h_s, \mathbf{Ag})$.

[f] Suppose that $\mathbf{A} \subsetneq \mathbf{B}$. Now suppose that $h'_{s'+j} \in E_K(h_{s+k}, h_s, \mathbf{B})$. So $j = k+1$ and $f_{h'}^{k+1}(s') \in E(f_h^k(s), h, \mathbf{B})$ and $h'_{s'+j} \in \text{lub}(h_{s+k}, h_s)$. But we know that $E(f_h^k(s), h, \mathbf{B}) \subseteq E(f_h^k(s), h, \mathbf{A})$ because E obeys f, thus $f_{h'}^{k+1}(s') \in E(f_h^k(s), h, \mathbf{A})$ and so $h'_{s'+j} \in E_K(h_{s+k}, h_s, \mathbf{A})$.

Now suppose that $h'_{s'-j} \in E_K(h_{s-k}, h_s, \mathbf{B})$. So $j = k-1$ and $b_{h'}^{k-1}(s') \in E(b_h^k(s), h, \mathbf{B})$, and $h'_{s'-j} \in \text{lub}(h_{s-k}, h_s)$. But we know that $E(b_h^k(s), h, \mathbf{B}) \subseteq E(b_h^k(s), h, \mathbf{A})$ because E obeys f, thus $b_{h'}^{k-1}(s') \in E(b_h^k(s), h, \mathbf{A})$, and so $h'_{s'-j} \in E_K(h_{s-k}, h_s, \mathbf{A})$.

[g] Suppose that $x' = h'_{s'+k'}$, $x = h_{s+k}$ and $x'' = h''_{s''+k''}$ and $\mathbf{A} \cap \mathbf{B} = \emptyset$ with $x' \approx_K x \approx_K x''$, i.e., $h'_{s'+k'} \approx_K h_{s+k} \approx_K h''_{s''+k''}$. That means $k' = k = k''$, and then that $f_{h'}^k(s') \approx f_h^k(s) \approx f_{h''}^k(s'')$. Since E obeys g, we have that there is (s''', h''') such that $s''' \approx f_h^k(s)$ with $E(s''', h''', \mathbf{A})$ and $E(s''', h''', \mathbf{B})$ contained in $E(f_{h'}^k(s'), h', \mathbf{A})$ and $E(f_{h''}^k(s''), h'', \mathbf{B})$, respectively.

Now since $s''' \approx f_h^k(s)$, $b_{h'''}^k(s''') \approx b_h^k(f_h^k(s)) = s$ by H2 for \approx . Let's call $b_{h'''}^k(s''')$, ' s_3 '. So $f_{h'''}^k(s_3) = f_{h'''}^k(b_{h'''}^k(s''')) \approx f_h^k(s)$. That means $h''_{s_3+k} \approx_K h_{s+k}$ by definition of \approx_K . But we will also, by the same argument, have that $h''_{s_3+k} \approx_K h'_{s'+k}$ and $h''_{s_3+k} \approx_K h''_{s''+k}$. By observation 7.1.25(1), we have $\text{lub}(h_{s_3+k}) = \text{lub}(h'_{s'+k})$ and $\text{lub}(h_{s_3+k}) = \text{lub}(h''_{s''+k})$.

Now suppose that $h_{s^*+m}^* \in E_K(h''_{s_3+k}, h''_{s_3}, \mathbf{A})$. That means $m = k + 1$ by observation 7.1.25(2), and $f_{h^*}^{k+1}(s^*) \in E(s''', h''', \mathbf{A})$ since $s''' = f_{h'''}^k(s_3)$ and $f_{h^*}^{k+1}(s^*)$ is the representative of $h_{s^*+m}^*$. So $f_{h^*}^{k+1}(s^*) \in E(f_{h'}^k(s'), h', \mathbf{A})$. Since $\text{lub}(h_{s_3+k}) = \text{lub}(h'_{s'+k})$, $h_{s^*+m}^* \in E_K(h'_{s'+k}, h'_{s'}, \mathbf{A})$. The same will go for the \mathbf{B} case using $h'_{s'+k}$.

Now suppose that $x' = h'_{s'-k'}$, $x = h_{s-k}$ and $x'' = h''_{s''-k''}$ and $\mathbf{A} \cap \mathbf{B} = \emptyset$ with $x' \approx_K x \approx_K x''$, i.e., $h'_{s'-k'} \approx_K h_{s-k} \approx_K h''_{s''-k''}$. That means $k' = k = k''$, and then that $b_{h'}^k(s') \approx b_h^k(s) \approx b_{h''}^k(s'')$. Since E obeys g, there must be (s''', h''') such that $s''' \approx b_h^k(s)$ with $E(s''', h''', \mathbf{A})$ and $E(s''', h''', \mathbf{B})$ contained in $E(b_{h'}^k(s'), h', \mathbf{A})$ and $E(b_{h''}^k(s''), h'', \mathbf{B})$, respectively. Now let $f_{h'''}^k(s''') = s_3$, then $b_{h'''}^k(s_3) = b_{h'''}^k(f_{h'''}^k(s''')) = s'''$. So $b_{h'''}^k(s_3) \approx b_h^k(s)$, and thus $h''_{s_3-k} \approx_K h_{s-k}$.

Now suppose that $h_{s^*-m}^* \in E_K(h''_{s_3-k}, h''_{s_3}, \mathbf{A})$. That means $m = k - 1$, and $b_{h^*}^{k-1}(s^*) \in E(s''', h''', \mathbf{A})$ since $s''' = b_{h'''}^k(s_3)$. So $f_{h^*}^{k-1}(s^*) \in E(b_{h'}^k(s'), h', \mathbf{A})$. Since $h_{s_3-k} \approx_K h'_{s'-k}$, $\text{lub}(h_{s_3-k}) = \text{lub}(h'_{s'-k})$ by observation 7.1.25, so $h_{s^*-m}^* \in E_K(h'_{s'-k}, h'_{s'}, \mathbf{A})$. The same will go for the \mathbf{B} case.

□

So we have that \mathfrak{M}_K is a neutral frame, it may not be universal even when \mathfrak{M} is, but that won't bother us. Now we must show that it acts like \mathfrak{M} in certain ways. For that we need the

new valuation v_K which we define as

$$v_K(\mathbf{p}) = \{x \in S_K : \text{rep}(x) \in v(\mathbf{p})\} \quad (7.16)$$

here we abuse the rep notation as we mentioned above. From the way that we have defined \approx_K , if $s \approx_K s'$, then $\text{rep}(s) \approx \text{rep}(s')$. So if $s \in v_K(\mathbf{p})$ and $s \approx_K s'$, then $\text{rep}(s) \in v(\mathbf{p})$ and so $\text{rep}(s') \in v(\mathbf{p})$, i.e., $s' \in v_K(\mathbf{p})$. Thus, \mathfrak{M}_K is a disjoint **NUZ** \mathcal{L} -model that obeys Kamp. Now we will show that,

Theorem 7.1.27. *For all (s, h) from \mathfrak{M}_K , and all $\varphi \in \mathcal{L}$,*

$$\mathfrak{M}_K, (s, h) \models \varphi \iff \mathfrak{M}, \text{rep}(s, h) \models \varphi$$

Proof. One thing to remember is that each static state in \mathfrak{M} , and in \mathfrak{M}_K is in exactly one history.

[**p**] Suppose that $\mathfrak{M}_K, (h_{s \pm k}, h_s) \models \mathbf{p}$, so $h_{s \pm k} \in v_K(\mathbf{p})$, that means $\text{rep}(h_{s \pm k}) \in v(\mathbf{p})$ so $\mathfrak{M}, (\text{rep}(h_{s \pm k}), h) \models \mathbf{p}$. Of course it works conversely as well.

[**IH:**] suppose that for all ψ of less complexity than φ , $\mathfrak{M}_K, (s, h) \models \psi \iff \mathfrak{M}, \text{rep}(s, h) \models \psi$.

[**X, P**] Suppose that $\mathfrak{M}_K, (h_{s-k}, h_s) \models P\psi$. Then by definition $\mathfrak{M}_K, (h_{s-(k+1)}, h_s) \models \psi$, so then by IH, $\mathfrak{M}, (\text{rep}(h_{s-(k+1)}), h) \models \psi$. Of course, $\text{rep}(h_{s-(k+1)}) = b_h^{k+1}(s)$, so $\mathfrak{M}, (b_h^{k+1}(s), h) \models \psi$, of course $\text{glb}(b_h^k(s), h) = b_h^{k+1}(s)$, so $\mathfrak{M}, \text{rep}(h_{s-k}, h_s) \models P\psi$. The same holds for the positive case. Now suppose that $\mathfrak{M}, \text{rep}(h_{s-k}, h_s) \models P\psi$. Then we have that $\text{rep}(h_{s-k}) = b_h^k(s)$, so $\text{glb}(b_h^k(s), h) = b_h^{k+1}(s)$. Thus by the truth condition of P , $\mathfrak{M}, (b_h^{k+1}(s), h) \models \psi$. So by the inductive hypothesis and that $\text{rep}(h_{s-(k+1)}) = b_h^{k+1}(s)$ we have $\mathfrak{M}_K, (h_{s-(k+1)}, h_s) \models \psi$. Therefore, $\mathfrak{M}_K, (h_{s-k}, h_s) \models P\psi$. The X case works in analogy with the P case.

[**□**] Suppose that $\mathfrak{M}_K, (h_{s-k}, h_s) \models \square\psi$. Then by definition for all $h'_{s'-k} \approx_K h_{s-k}$, $\mathfrak{M}_K, (h'_{s'-k}, h_{s'}) \models \psi$. We also have that $\text{rep}(h_{s-k}, h_s) = (b_h^k(s), h)$, so suppose that $s^* \approx b_h^k(s)$. Then say $s^* \in h^*$ since it has to be in some history, so let $f_{h^*}^k(s^*) = s_*$, then $b_{h^*}^k(s_*) \approx b_h^k(s)$. Thus $h_{s_*-k}^* \approx_K h_{s-k}$. That means $\mathfrak{M}_K, (h_{s_*-k}^*, h_{s_*}^*) \models \psi$, so by IH,

$\mathfrak{M}, (b_{h^*}^k(s_*), h^*) \models \psi$, i.e., $\mathfrak{M}, (s^*, h^*) \models \psi$. Thus, $\mathfrak{M}, (b_h^k(s), h) \models \Box\psi$. The positive case is similar, but one has $s^* \approx f_h^k(s)$ so you use $b_{h^*}^k(s^*) = s_*$.

Now suppose that $\mathfrak{M}, (b_h^k(s), h) \models \Box\psi$, so for all s^* such that $s^* \approx b_h^k(s)$, $\mathfrak{M}, (s^*, h^*) \models \psi$. Assume that $h'_{s'-k} \approx_K h_{s-k}$. That means $b_{h'}^k(s') \approx b_h^k(s)$, so $\mathfrak{M}, (b_{h'}^k(s'), h') \models \psi$. By IH, $\mathfrak{M}_K, (h'_{s'-k}, h'_{s'}) \models \psi$, since $h'_{s'-k}$ was arbitrary, $\mathfrak{M}_K, (h_{s-k}, h_s) \models \Box\psi$. The positive case is the same.

[[A xstit]] Now suppose that $\mathfrak{M}_K, (h_{s-k}, h_s) \models [\mathbf{A} \text{ xstit}] \psi$. By definition for all $h'_{s'-m} \in E_K(h_{s-k}, h_s, \mathbf{A})$, $m = k-1$ and $\mathfrak{M}_K, (h'_{s'-m}, h'_{s'}) \models \psi$. Now suppose that $s^* \in E(b_h^k(s), h, \mathbf{A})$. Again, we will have $b_{h^*}(s^*) \approx b_h^k(s)$ since $E(b_h^k(s), h, \mathbf{A}) \subseteq \text{lub}(b_h^k(s))$, so then $s^* = b_{h^*}^k(f_{h^*}^{k-1}(s^*)) \approx b_h^k(s)$, and we can call $f_{h^*}^{k-1}(s^*)$, s_* . Thus $b_{h^*}^k(s_*) \approx b_h^k(s)$, thus $h_{s^*-k}^* \approx_K h_{s-k}$. That means $\mathfrak{M}_K, (h_{s^*-k}^*, h_{s^*}^*) \models \psi$ and so by IH, $\mathfrak{M}, (b_{h^*}^k(s_*), h^*) \models \psi$, i.e., $\mathfrak{M}, (s^*, h^*) \models \psi$. And since s^* was arbitrary, $\mathfrak{M}, (b_h^k(s), h) \models [\mathbf{A} \text{ xstit}] \psi$. The positive case where $\mathfrak{M}_K, (h_{s+k}, h_s) \models [\mathbf{A} \text{ xstit}] \psi$ we have by definition, for all $h'_{s'+m} \in E_K(h_{s+k}, h_s, \mathbf{A})$, $m = k+1$ and $\mathfrak{M}_K, (h'_{s'+m}, h'_{s'}) \models \psi$. Now taking $s^* \in E(f_h^k(s), h, \mathbf{A})$, it must be that $b_{h^*}(s^*) \approx f_h^k(s)$, and then $b_{h^*}^{k+1}(s^*) \approx s$ by H2, so let $s_* = b_{h^*}^{k+1}(s^*)$, then $f_{h^*}^k(s_*) \approx f_h^k(s)$, and then we proceed as above.

Conversely, for the positive case, if $\mathfrak{M}, (f_h^k(s), h) \models [\mathbf{A} \text{ xstit}] \psi$. Then for all $s' \in E(f_h^k(s), h, \mathbf{A})$, $\mathfrak{M}, (s', h') \models \psi$. Now suppose that $h'_{s'+m} \in E_K(h_{s+k}, h_s, \mathbf{A})$. Then $\text{rep}(h_{s+k}) = f_h^k(s)$ and $\text{rep}(h'_{s'+m}) = f_{h'}^m(s')$, and $m = k+1$. So by definition of E_K we have $f_{h'}^m(s') = f_{h'}^{k+1}(s') \in E(f_h^k(s), h, \mathbf{A})$; thus, $\mathfrak{M}, (f_{h'}^{k+1}(s'), h) \models \psi$. So by IH $\mathfrak{M}_K, (h'_{s'+k+1}, h'_{s'}) \models \psi$. Since $h'_{s'+m}$ was arbitrary, $\mathfrak{M}_K, (h_{s+k}, h_s) \models [\mathbf{A} \text{ xstit}] \psi$. For the negative case, we have $\mathfrak{M}, (b_h^k(s), h) \models [\mathbf{A} \text{ xstit}] \psi$. Then for all $s' \in E(b_h^k(s), h, \mathbf{A})$, $\mathfrak{M}, (s', h') \models \psi$. Then $\text{rep}(h_{s-k}) = b_h^k(s)$ and $\text{rep}(h'_{s'-m}) = b_{h'}^m(s')$, with $m = k-1$. So by definition of E_K we have $b_{h'}^m(s') = b_{h'}^{k-1}(s') \in E(b_h^k(s), h, \mathbf{A})$; thus, $\mathfrak{M}, (f_{h'}^{k-1}(s'), h) \models \psi$. So by IH $\mathfrak{M}_K, (h'_{s'-k-1}, h'_{s'}) \models \psi$. Since $h'_{s'-m}$ was arbitrary, $\mathfrak{M}_K, (h_{s-k}, h_s) \models [\mathbf{A} \text{ xstit}] \psi$.

□

If we look at the proofs involved in proving theorem 7.1.23, we will see that we didn't use

the universal property at all. We did of course use the Kamp property extensively. This means that every disjoint neutral model \mathfrak{M} has an elementarily equivalent regular model, i.e., the Kamp quotient structure \mathfrak{M}^* . Now we should observe that \mathfrak{M}_K will have the Kamp property. This is quite easy to see: if $h_{s-k} \approx_K h_{s-k'}$, i.e., they are in the same history, h_s , we must have $k = k'$, so $h_{s-k} = h_{s-k'}$. The same will go for h_{s+k} . That means that \mathfrak{M}_K has a regular model \mathfrak{M}_K^* such that for all $(h_{s\pm k}, h_s) \in |\mathfrak{M}_K|$, $\mathfrak{M}_K, (h_{s\pm k}, h_s) \models \varphi$ iff $\mathfrak{M}^*, ([h_{s\pm k}], [h_s]) \models \varphi$.

Now, we might notice that if $(s, h) \in |\mathfrak{M}|$, then $rep(h_{s-0}, h_s) = (s, h)$. So for each (s, h) , from \mathfrak{M} ,

$$\mathfrak{M}_K, (h_{s-0}, h_s) \models \varphi \iff \mathfrak{M}, (s, h) \models \varphi$$

and that means that if $\mathfrak{M}, (s, h) \not\models \varphi$, but $\mathfrak{M}, (s, h) \models \Gamma$, then $\mathfrak{M}_K, (h_{s-0}, h_s) \not\models \varphi$ and $\mathfrak{M}_K, (h_{s-0}, h_s) \models \Gamma$. But then $\mathfrak{M}_K^*, ([h_{s-0}], [h_s]) \not\models \varphi$ while, $\mathfrak{M}_K^*, ([h_{s-0}], [h_s]) \models \Gamma$. So if Γ holds, while φ fails in a universal **NUZ** \mathcal{L} -model, then Γ will hold and φ will fail in a regular \mathcal{L} -model. It is easy to see that it will then hold in a universal regular \mathcal{L} -model by a standard generated submodel argument:

Observation 7.1.28. *If $\mathfrak{M} = \langle S, H, E, v \rangle$ is a regular \mathcal{L} -model, with $(s, h) \in |\mathfrak{M}|$, then $\mathfrak{M}^{(s,h)}$ defined as*

1. $H^{(s,h)} = \{h' \in H : \exists s' \in S \text{ s.t. } s' \in h \cap h'\}$
2. $S^{(s,h)} = \cup H^{(s,h)}$
3. $E^{(s,h)}(s', h', \mathbf{A}) = E(s', h', \mathbf{A}) \cap S^{(s,h)}$
4. $v^{(s,h)}(\mathbf{p}) = v(\mathbf{p}) \cap S^{(s,h)}$

is such that for all $\varphi \in \mathcal{L}$,

$$\mathfrak{M}, (s, h) \models \varphi \iff \mathfrak{M}^{(s,h)}, (s, h) \models \varphi$$

Thus we have proved.

Theorem 7.1.29. \vdash_{xp} is complete with respect to the class of universal regular models:

$$\Gamma \vdash_{\text{xp}} \varphi \text{ iff } \Gamma \models_{\text{xp}} \varphi.$$

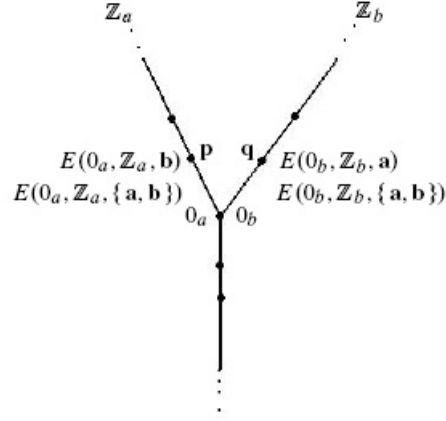


Figure 7.5: Counter Example to Indep-G

7.2 Completeness of \vdash_{xp}^I

For this we just have to make a model that doesn't satisfy condition g and falsifies Indep-G. Let the language be $\mathbf{Ag} = \{\mathbf{a}, \mathbf{b}\}$ $H = \{\mathbb{Z}_a, \mathbb{Z}_b\}$ such that if $z_b, z_a \leq 0$, $z_a = z_b$, and after 0 not. And $\mathbf{At} = \{\mathbf{p}, \mathbf{q}\}$, then let

1. $S = \mathbb{Z}_a \cup \mathbb{Z}_b$,
2. $E(z, \mathbb{Z}_a, \mathbf{a}) = (z + 1_a, \mathbb{Z}_a)$
3. $E(z, \mathbb{Z}_b, \mathbf{b}) = (z + 1_b, \mathbb{Z}_b)$ for non 0 z .
4. $E(0_a, \mathbb{Z}_a, \mathbf{a}) = (1_a, \mathbb{Z}_a) = E(0_a, \mathbb{Z}_a, \mathbf{b})$
5. $E(0_b, \mathbb{Z}_b, \mathbf{b}) = (1_b, \mathbb{Z}_b) = E(0_b, \mathbb{Z}_b, \mathbf{a})$
6. $v(\mathbf{p}) = \{1_a\}$
7. $v(\mathbf{q}) = \{1_b\}$

We provide an image of this counter model in figure 7.5. This is clearly a regular model, but at 0 we would have that $(0_a, \mathbb{Z}_a) \models \Diamond([\mathbf{a} \text{ xstit}] \mathbf{p}) \wedge \Diamond([\mathbf{b} \text{ xstit}] \mathbf{q})$ (the right conjunct is true because

of $(0_b, \mathbb{Z}_b)$). But relative to $(0_a, \mathbb{Z}_a)$, $[\mathbf{b} \text{ xstit}] \mathbf{q}$ fails, and relative to $(0_b, \mathbb{Z}_b)$, $[\mathbf{a} \text{ xstit}] \mathbf{p}$ fails, and $0_a = 0_b$, and those are the only choices, so $(0_a, \mathbb{Z}_a) \not\models \Diamond([\mathbf{a} \text{ xstit}] \mathbf{p} \wedge [\mathbf{b} \text{ xstit}] \mathbf{q})$. So the class of regular frames, with effectivity functions that may not satisfy g doesn't validate Indep-G. Of course this is in a restricted language, but if we add more agent terms we can still get the same result, similarly with more atomic sentences. This was also done with agent terms not with roles, but those are simply notational variants. Let's call the class of models like this for \mathcal{L}^I in which E may not satisfy condition g, \mathcal{L}^I -models. We can let the entailment relation defined by: For all \mathcal{L}^I -models \mathfrak{M} , and for all $(s, h) \in |\mathfrak{M}|$, $\mathfrak{M}, (s, h) \models \Gamma$ only if $\mathfrak{M}, (s, h) \models \varphi$, be denoted by $\Gamma \models_{\text{xp}}^I \varphi$. If we replace the agent terms in \mathcal{L} with those of role terms, we would get \mathcal{L}^I . So using the axioms of \vdash_{xp} with agent terms replaced with role terms we get an institutional version of \vdash_{xp} . We call that logic, i.e., its consequence relation, \vdash_{xp}^* . We can then show:

Theorem 7.2.1. *If $\Gamma; \varphi \subseteq \mathcal{L}^I$, then $\Gamma \vdash_{\text{xp}}^I \varphi$ iff $\Gamma \models_{\text{xp}}^I \varphi$.*

Proof. For completeness just repeat the completeness proof of section 7.1 without condition g on effectivity functions. □

7.3 Completeness of the Rest: \vdash_{Ixp}

We will now stop ignoring \Subset . As a brief aside, notice that since the logic of \vdash_{xp}^I is a sub-logic of \vdash_{Ixp} , if a set Δ of \mathcal{L}^I -sentences is \vdash_{xp}^I -inconsistent, i.e., $\Delta \vdash_{\text{xp}}^I \perp$, then $\Delta \vdash_{\text{Ixp}} \perp$. This means that \vdash_{Ixp} -consistency implies \vdash_{xp}^I -consistency. The situation is better than that though. To see how much better, we need a lemma.

Lemma 7.3.1. *If $\mathfrak{M} = \langle \mathcal{D}, \mathfrak{F}_x, v \rangle$ is an $\mathcal{L}_{\subseteq}^I$ -model, and $\mathfrak{M}^* = \langle \mathfrak{F}^*, v^* \rangle$ is a universal regular \mathcal{L}^I -model, then $\mathfrak{M}^c = \langle \mathcal{D}, \mathfrak{F}^*, v^c \rangle$ where v^c is defined as*

1. $v^c(\mathbf{p}) = \langle \llbracket \mathbf{p} \rrbracket_1^{\mathfrak{M}}, v^*(\mathbf{p}) \rangle$
2. $v^c(\mathbf{r}) = v(\mathbf{r})$

is such that for all $\psi \Subset \psi'$ and $\varphi \in \mathcal{L}^I$,

1. for all $(s, h) \in |\mathfrak{M}^c| = |\mathfrak{M}^*|$, $\mathfrak{M}^c, (s, h) \models \varphi$ iff $\mathfrak{M}^*, (s, h) \models \varphi$, and
2. $\mathfrak{M}^c \models \psi \in \psi'$ iff $\mathfrak{M} \models \psi \in \psi'$.

Proof. The proof is by induction on the complexity of ψ , ψ' and φ . It is straight forward, so we will omit the details. The overall reason is that v^c agrees with v^* with respect to the formulas of \mathcal{L}^I , and we keep the content relations from \mathcal{D} . That means it maintains the content sentences from \mathfrak{M} . \square

Now we will argue that \vdash_{Ixp} is sound with respect to \models_{Ixp} .

Proposition 7.3.2. *For all $\Gamma; \varphi \subseteq \mathcal{L}_{\subseteq}^I$, $\Gamma \vdash_{\text{Ixp}} \varphi$ only if $\Gamma \models_{\text{Ixp}} \varphi$.*

Proof. In proposition 7.1.3, we argued that the XP-stit group axioms from definition 6.3.5 were sound for the neutral models, so they are sound for the regular models. Therefore they are sound for $\mathcal{L}_{\subseteq}^I$ -models. The classical axioms are clearly sound. In proposition 5.3.1 we argued that the \subseteq axioms were sound for SI -frames, and \mathcal{D} in an $\mathcal{L}_{\subseteq}^I$ -model is an SI -frame, so they are sound here too. We have a few new cases: PCX1–10, but those are clearly sound as well. For example PCX10 holds since each $\mathbf{r} \in \mathbf{Rol}$ and $\mathbf{p} \in \mathbf{At}$ is assigned an atom, but in the definition of an $\mathcal{L}_{\subseteq}^I$ -model, $\llbracket \mathbf{p} \rrbracket_1 \neq \llbracket \mathbf{r} \rrbracket_1$, so if either $\llbracket \mathbf{p} \rrbracket_1 \lesssim \llbracket \mathbf{r} \rrbracket_1$ or $\llbracket \mathbf{r} \rrbracket_1 \lesssim \llbracket \mathbf{p} \rrbracket_1$, then $\llbracket \mathbf{p} \rrbracket_1 = \llbracket \mathbf{r} \rrbracket_1$ since $\llbracket \mathbf{p} \rrbracket_1$ and $\llbracket \mathbf{r} \rrbracket_1$ are atoms in \mathcal{D} . We leave the other cases for the reader. \square

With this, then we can show that,

Proposition 7.3.3. *If $\Delta \vdash_{\text{Ixp}} \varphi$, and Δ and φ are both in \mathcal{L}^I , then $\Delta \vdash_{\text{xp}}^I \varphi$. As a corollary we have that \mathcal{L}^I -consistency implies $\mathcal{L}_{\subseteq}^I$ -consistency.*

Proof. Suppose that $\Delta \not\vdash_{\text{xp}}^I \varphi$. Then there is an \mathcal{L}^I -model $\mathfrak{M} = \langle S, H, E, v \rangle$ and (s, h) in $|\mathfrak{M}|$ that satisfies Δ , but doesn't satisfy φ by theorem 7.2.1. Let \mathcal{D} be any SI -frame for $\mathcal{L}_{\subseteq}^I$. Such things exists since \vdash_{Ixp} is sound, so \emptyset has a model, so \mathcal{D} could be a model for \emptyset . By lemma 7.3.1(1), $\mathfrak{M}^c = \langle \mathcal{D}, S, H, E, v^c \rangle$ is an $\mathcal{L}_{\subseteq}^I$ -model such that for all $(s', h') \in |\mathfrak{M}^c| = |\mathfrak{M}|$, $\mathfrak{M}^c, (s', h') \models \varphi$ iff $\mathfrak{M}, (s', h') \models \varphi$. So a fortiori, $\mathfrak{M}^c, (s, h) \models \varphi$ iff $\mathfrak{M}, (s, h) \models \varphi$. So

$\mathfrak{M}^c, (s, h) \models \Delta$, and $\mathfrak{M}^c, (s, h) \not\models \varphi$. That means by proposition 7.3.2, $\Delta \not\vdash_{\text{Ixp}} \varphi$. So we have the first result. For the corollary, if $\Delta \not\vdash_{\text{xp}}^I \perp$, i.e., it is \vdash_{xp}^I -consistent, then $\Delta \not\vdash_{\text{Ixp}} \perp$ by the previous result. That means \vdash_{Ixp} is a *conservative extension* of \vdash_{xp}^I . \square

To show completeness of \vdash_{Ixp} with respect to \models_{Ixp} , we construct what is known as a parameterization³ of the logic \vdash_{SI} by the logic \vdash_{xp}^I . In Caleiro et al. (1999) the authors show that completeness of logics can be preserved by a combination by parameterization. However, their demonstration is with respect to a general class of structures that would take us too far away from the current models to explain. Instead we will show that the logic is complete in a more direct manner.

First we note that a model, like that in definition 5.3.1, can be constructed from a \vdash_{Ixp} -maximal consistent set. Of course such maximal sets exist and any \vdash_{Ixp} -consistent set can be extended to a maximal one by Lindenbaum's lemma. We construct the special PO-set as follows:

Definition 7.3.1. Let Γ be a \vdash_{Ixp} -maximal consistent set. Define the canonical SI -frame for Γ , as \mathcal{D}_Γ as follows:

1. $\mathcal{D}_\Gamma = \langle D_\Gamma, \lesssim_\Gamma, \gamma_\Gamma, \odot_\Gamma \rangle$ where $D_\Gamma = \mathcal{L}^I / \sim_{SI}$.
2. \sim_{SI} is defined between $A, B \in \mathcal{L}^I \cup \mathcal{P}(\mathbf{Rol})$ ($A \sim_{SI} B$) when $(A \in B) \wedge (B \in A) \in \Gamma$.
 $[A]$ refers to the equivalence class $\{ B \in \mathcal{L}^I \cup \mathcal{P}(\mathbf{Rol}) : A \sim_{SI} B \}$, of A under \sim_{SI} .
3. For all $A, B \in \mathcal{L}^I \cup \mathcal{P}(\mathbf{Rol})$ $[A] \lesssim_\Gamma [B]$ iff $A \in B \in \Gamma$ or $[A] = [B]$, and $[A] \gamma_\Gamma [B] = [A \wedge B]$.
4. $D_{\Gamma_A} = \{ [\mathbf{p}] : \mathbf{p} \in \mathbf{At} \} \cup \{ [\mathbf{r}] : \mathbf{r} \in \mathbf{Rol} \}$, and $\odot_\Gamma = [\perp]$.

By inspecting the PC and PCX axioms from definition 6.3.5, we can see that \mathcal{D}_Γ is the right kind of PO-set. That means for any \mathcal{L}^I -model $\mathfrak{M} = \langle S, H, E, v \rangle$, $\langle \mathcal{D}_\Gamma, S, H, E, v^c \rangle$ where v^c is as in observation 7.3.1, is an \mathcal{L}_\subseteq^I -model. That is we define $v^c(\mathbf{p}) = \langle [\mathbf{p}], v(\mathbf{p}) \rangle$ and

³See Caleiro et al. (1999).

$v^c(\mathbf{r}) = [\mathbf{r}]$. In this model, clearly $\mathfrak{M}, (s, h) \models A \subseteq B$ iff $A \subseteq B \in \Gamma$. Now we show the following lemma.

Lemma 7.3.4. *Suppose that Γ is a \vdash_{Ixp} -maximally consistent set. Then for any $\mathcal{L}_{\subseteq}^I$ -model \mathfrak{M} and $(s, h) \in |\mathfrak{M}|$, if*

1. $\mathfrak{M}, (s, h) \models A \subseteq B$ iff $A \subseteq B \in \Gamma$, and
2. $\mathfrak{M}, (s, h) \models \theta$ for all $\theta \in \Gamma \cap \mathcal{L}^I$, then

For all $\varphi \in \mathcal{L}_{\subseteq}^I$, $\mathfrak{M}, (s, h) \models \varphi$ iff $\varphi \in \Gamma$.

Proof. Suppose that

1. $\mathfrak{M}, (s, h) \models A \subseteq B$ for all $A \subseteq B \in \Gamma$, and
2. $\mathfrak{M}, (s, h) \models \theta$ for all $\theta \in \Gamma \cap \mathcal{L}^I$.

The proof is by induction on the complexity of φ . Suppose $\mathbf{p} \in \mathbf{At}$, then suppose that $\mathbf{p} \in \Gamma$, then $\mathbf{p} \in \mathcal{L}^I$, so $\mathfrak{M}, (s, h) \models \mathbf{p}$ by 2. Conversely, suppose that $\mathfrak{M}, (s, h) \models \mathbf{p}$, then $\mathbf{p} \in \mathcal{L}^I$. If $\mathbf{p} \notin \Gamma$, then $\neg \mathbf{p} \in \Gamma$ by maximality, and $\neg \mathbf{p} \in \mathcal{L}^I$, so $\mathfrak{M}, (s, h) \models \neg \mathbf{p}$ by 2. Therefore $\mathfrak{M}, (s, h) \not\models \mathbf{p}$, a contradiction.

IH: Suppose for all $\theta' \in \mathcal{L}_{\subseteq}^I$ of less complexity than φ , $\mathfrak{M}, (s, h) \models \theta'$ iff $\theta' \in \Gamma$.

The Boolean cases are standard, and follow by the use of the induction hypothesis. Note that if $\varphi = [*]\theta'$ for $[*] \in \{[\mathbf{R} \text{ xstit}], X, P, \Box : \mathbf{R} \subseteq \mathbf{Rol}\}$, then $\varphi \in \mathcal{L}^I$ by the construction of the language $\mathcal{L}_{\subseteq}^I$. So each of those cases follow from 2. If $\varphi = A \subseteq B$, then $\mathfrak{M}, (s, h) \models A \subseteq B$ iff $A \subseteq B \in \Gamma$, by 1. \square

What this lemma says, in essence, is that the truth of formulas in $\mathcal{L}_{\subseteq}^I$ is separable into the truth of the \subseteq part, and the truth of the \mathcal{L}^I part. But we can use this to show the following fact.

Observation 7.3.5. *If*

1. $\Gamma \subseteq \mathcal{L}_{\subseteq}^I$,
2. $\Gamma \not\vdash_{\text{Ixp}} \perp$,

3. $\varphi \in \mathcal{L}^I$, and

4. $\Gamma; \varphi \vdash_{\text{Ixp}} \perp$,

then there is $\Gamma' \subseteq \Gamma \cap \mathcal{L}^I$ such that $\Gamma'; \varphi \vdash_{\text{xp}}^I \perp$. I.e., if a set of \mathcal{L}_{\in}^I sentences is \vdash_{Ixp} -consistent, but it is \vdash_{Ixp} -inconsistent with a \mathcal{L}^I -formula φ , then there is a subset of \mathcal{L}^I sentences in Γ that is \vdash_{xp}^I -inconsistent with φ .

Proof. Suppose 1,2,3, and 4. Suppose for reductio that all $\Gamma' \subseteq \Gamma \cap \mathcal{L}^I$, are such that $\Gamma'; \varphi \not\vdash_{\text{xp}}^I \perp$. Since Γ is \vdash_{Ixp} -consistent by 2, there is a maximally \vdash_{Ixp} -consistent extension of Γ , call it Γ^+ . With Γ^+ we can form \mathcal{D}_{Γ^+} as in definition 7.3.1. Let $\Gamma'' = \Gamma \cap \mathcal{L}^I$. Since $\Gamma'' \subseteq \Gamma$, it is \vdash_{xp}^I -consistent by proposition 7.3.3. Also, by assumption $\Gamma''; \varphi$ is \vdash_{xp}^I -consistent. That means there is a \mathcal{L}^I -model $\mathfrak{M} = \langle S, H, E, v \rangle$, and $(s, h) \in |\mathfrak{M}|$ such that $\mathfrak{M}, (s, h) \models \Gamma''; \varphi$ by theorem 7.2.1. Then we can form $\mathfrak{M}^c = \langle \mathcal{D}_{\Gamma^+}, S, H, E, v^c \rangle$ as before, and we will have that for all $\theta \in \mathcal{L}^I$,

1. for all $(s, h) \in |\mathfrak{M}^c| = |\mathfrak{M}|$, $\mathfrak{M}^c, (s, h) \models \theta$ iff $\mathfrak{M}, (s, h) \models \theta$, and
2. $\mathfrak{M}^c \models \psi \in \psi'$ iff $\psi \in \psi' \in \Gamma^+$.

We will then have that $\mathfrak{M}^c, (s, h) \models \theta$ for all $\theta \in \Gamma; \varphi$. But that means $\Gamma; \varphi$ is not \vdash_{Ixp} -inconsistent (by soundness) contrary to 4.

So there is some $\Gamma' \subseteq \Gamma \cap \mathcal{L}^I$ such that $\Gamma'; \varphi \vdash_{\text{xp}}^I \perp$. □

Finally we can show completeness.

Theorem 7.3.6. *If $\Gamma; \varphi \subseteq \mathcal{L}_{\in}^I$, then $\Gamma \vdash_{\text{Ixp}} \varphi$ iff $\Gamma \models_{\text{Ixp}} \varphi$.*

Proof. The only if direction follows by soundness. So suppose that $\Gamma \not\vdash_{\text{Ixp}} \varphi$ for the if direction. That means $\Gamma; \neg\varphi$ is \vdash_{Ixp} -consistent and so there is a maximally \vdash_{Ixp} -consistent extension of $\Gamma; \neg\varphi$, call it Γ^+ . Now we can form \mathcal{D}_{Γ^+} as before, but we can also look at $\Gamma^+ \cap \mathcal{L}^I$. Since it is contained in Γ^+ it is \vdash_{Ixp} -consistent, and so it is \vdash_{xp}^I -consistent. Therefore it has a \mathcal{L}^I -model $\mathfrak{M} = \langle S, H, E, v \rangle$, i.e., there is $(s, h) \in |\mathfrak{M}|$, and $\mathfrak{M}, (s, h) \models \Gamma^+ \cap \mathcal{L}^I$. So we can form $\mathfrak{M}^c = \langle \mathcal{D}_{\Gamma^+}, S, H, E, v^c \rangle$, as before and we get that for all $\theta \in \Gamma^+ \cap \mathcal{L}^I$, $\mathfrak{M}^c, (s, h) \models \theta$,

and $A \subseteq B \in \Gamma^+$ iff $\mathfrak{M}^c, (s, h) \models A \subseteq B$. So by lemma 7.3.4, we have that for all $\theta \in \Gamma^+$, $\mathfrak{M}^c, (s, h) \models \theta$. Therefore $\mathfrak{M}^c, (s, h) \models \Gamma$, but $\mathfrak{M}^c, (s, h) \not\models \varphi$. Thus $\Gamma \not\models_{\text{xp}} \varphi$. \square

7.4 Proof of Proposition 6.6.1

We recall proposition 6.6.1 and the definition of an implementation for clarity.

Proposition 6.6.1. For any implementation \mathfrak{I} , if $\Omega \vdash_N \varphi$, and $\Omega \not\models_{\text{Ixp}} \perp$ and $\varphi \in \mathcal{L}^I$, then $\mathfrak{I}(\Omega) \vdash_{\text{xp}}^\Omega \delta$ for all $\delta \in \mathfrak{I}(\varphi)$. \square

Definition of Implementation. Let Ω be a code. An *implementation* of Ω is a triple $\mathfrak{I} = \langle \text{holds}, \pi, \mathfrak{I}(\Omega) \rangle$ such that

1. $\text{holds} \subseteq \mathcal{P}(\mathbf{Ag}) \times \mathbf{Rol}$, such that for each $\mathbf{a} \in \mathbf{Ag}$ there is $\mathbf{A} \subseteq \mathbf{Ag}$ and $\mathbf{r} \in \mathbf{Rol}$ such that $\mathbf{a} \in \mathbf{A}$ and $\langle \mathbf{A}, \mathbf{r} \rangle \in \text{holds}$.
2. π is a partition of \mathbf{Rol} , where $\pi_{\mathbf{r}}$ is the cell of the partition containing \mathbf{r} , such that
3. if $\langle \mathbf{A}, \mathbf{r} \rangle \in \text{holds}$, then for all $\mathbf{r}' \in \pi_{\mathbf{r}}$, $\langle \mathbf{A}, \mathbf{r}' \rangle \in \text{holds}$, and
4. $\mathfrak{I}(\Omega) \subseteq \mathcal{L}_\Omega^B$ such that each $\delta \in \mathfrak{I}(\Omega)$ is a substitution instance of some $\varphi \in \Omega$, where each role term \mathbf{r} mentioned in φ is replaced uniformly in φ by an agent term \mathbf{A} such that $\langle \mathbf{A}, \mathbf{r} \rangle \in \text{holds}$.

\square

If we have a set $Q \subseteq \text{holds}$ such that for each $\mathbf{r} \in \mathbf{Rol}$, there is a unique $\mathbf{A} \subseteq \mathbf{Ag}$ with $\langle \mathbf{A}, \mathbf{r} \rangle \in Q$, then we can define $\text{holds}_Q^{-1}(\mathbf{r}) = \mathbf{A}$ such that $\langle \mathbf{A}, \mathbf{r} \rangle \in Q$. We also define $\mathfrak{I}_Q(\mathbf{R}) = \cup \{ \mathbf{A} : \langle \mathbf{A}, \mathbf{r} \rangle \in Q \ \& \ \mathbf{r} \in \mathbf{R} \}$. Now for each $\varphi' \in \Omega$, form the set of substitution instances $\mathfrak{I}_Q(\varphi')$ using the pairs in Q . That is for each $\varphi' \in \Omega$ and \mathbf{r} mentioned in φ' we uniformly substitute \mathbf{A} for \mathbf{r} when $\langle \mathbf{A}, \mathbf{r} \rangle \in Q$. We will call this set of formulas $\mathfrak{I}_Q(\Omega)$. Notice that for each $\varphi' \in \Omega$, $\mathfrak{I}_Q(\varphi')$ will consist of just one formula since there is only one possible substitution for each \mathbf{r} mentioned in φ' .

To prove proposition 6.6.1, we show the following

Lemma 7.4.1. *If \mathfrak{F} is an implementation of Ω , $\mathfrak{M} = \langle D, \mathfrak{F}_x, v \rangle$ an $\mathcal{L}_{\subseteq}^I$ -model,*

$\mathfrak{M}^ = \langle S^*, H^*, E^*, v^* \rangle$ is a \mathcal{L}_{Ω}^B -model, and $Q \subseteq \text{holds}$ such that for each \mathbf{r} , with $\langle \mathbf{A}, \mathbf{r} \rangle \in Q$, then \mathbf{A} is the unique $\mathbf{A} \subseteq \mathbf{Ag}$ such that $\langle \mathbf{A}, \mathbf{r} \rangle \in Q$, then $\mathfrak{M}^+ = \langle D, \langle S^*, H^*, E^+ \rangle, v^+ \rangle$ where v^+ is defined as*

1. $v^+(\mathbf{p}) = \langle \llbracket \mathbf{p} \rrbracket_1^{\mathfrak{M}}, v^*(\mathbf{p}) \rangle$
2. $v^+(\mathbf{r}) = v(\mathbf{r})$

and E^+ is defined as

1. $E^+(s, h, \emptyset) = E^*(s, h, \emptyset)$,
2. $E^+(s, h, \{ \mathbf{r} \}) = E^*(s, h, \text{holds}_Q^{-1}(\mathbf{r}))$ for $\mathbf{r} \in \mathbf{Rol}$, and more generally,
3. $E^+(s, h, \mathbf{R}) = E^*(s, h, \mathfrak{F}_Q(\mathbf{R}))$.

Then for all $\varphi \in \mathcal{L}^I$, and (s, h) from $|\mathfrak{M}^+|$, $\mathfrak{M}^+, (s, h) \models \varphi$ iff $\mathfrak{M}^, (s, h) \models \mathfrak{F}_Q(\varphi)$.*

Proof. Suppose all that we need. Then we first have to show that \mathfrak{M}^+ is an $\mathcal{L}_{\subseteq}^I$ -model. The thing to check is that E^+ obeys conditions a–f

- (a) If $s \notin h$, then for all \mathbf{A} , $E^*(s, h, \mathbf{A}) = \emptyset$, so the same will hold for any \mathbf{R} .
- (b) If $s' \in E^+(s, h, \mathbf{R})$, then $s' \in E^*(s, h, \mathfrak{F}(\mathbf{R})) \subseteq E^*(s, h, \emptyset) = E^+(s, h, \emptyset)$.
- (c) If $s \in h$, then $\text{lub}(s, h) \in E^*(s, h, \mathbf{A})$, for any \mathbf{A} , so it will hold for $E^*(s, h, \mathfrak{F}(\mathbf{R})) = E^+(s, h, \mathbf{R})$.
- (d) $\text{lub}(s) = E^*(s, h, \emptyset) = E^+(s, h, \emptyset)$.
- (e) If $s \in h$, then $E^+(s, h, \mathbf{Rol}) = E^*(s, h, \mathfrak{F}(\mathbf{Rol}))$, and
 $\mathfrak{F}(\mathbf{Rol}) = \cup \{ \mathbf{A} : \langle \mathbf{A}, \mathbf{r} \rangle \in \text{holds} \ \& \ \mathbf{r} \in \mathbf{Rol} \}$. But each $\mathbf{a} \in \mathbf{Ag}$ is assigned to some role, so $\mathfrak{F}(\mathbf{Rol}) = \mathbf{Ag}$. So $E^+(s, h, \mathbf{Rol}) = \{ \text{lub}(s, h) \}$ since $E^*(s, h, \mathbf{Ag}) = \{ \text{lub}(s, h) \}$.
- (f) Suppose that $\mathbf{R} \subseteq \mathbf{R}'$. Then $\mathfrak{F}(\mathbf{R}) \subseteq \mathfrak{F}(\mathbf{R}')$, so $E^+(s, h, \mathbf{R}') = E^*(s, h, \mathfrak{F}(\mathbf{R}')) \subseteq E^*(s, h, \mathfrak{F}(\mathbf{R})) = E^+(s, h, \mathbf{R})$.

So $\langle S^*, H^*, E^+ \rangle$ is a model of the right kind, now we just have to apply lemma 7.3.1(1), and we have our result. \square

For proposition 6.6.1, we suppose that $\Omega \vdash_N \varphi$. Now suppose that $\mathfrak{F}(\Omega) \not\vdash_{xp}^\Omega \delta$ and $\delta \in \mathfrak{F}(\varphi)$. By completeness we will have a \mathcal{L}_Ω^B -model $\mathfrak{M} = \langle S, H, E, v \rangle$ with a dynamic state (s, h) such that $\mathfrak{M}, (s, h) \models \mathfrak{F}(\Omega)$, but $\mathfrak{M}, (s, h) \not\models \delta$. What we need to do is specify a set $Q \subseteq$ holds that interprets each \mathbf{r} with a unique agent term \mathbf{A} . Since $\delta \in \mathfrak{F}(\varphi)$, δ is the result of substituting an agent term \mathbf{A} for each \mathbf{r} mentioned in φ . Recall that if there is a formula like $[\mathbf{r}, \mathbf{r}' \text{ xstit}] \mathbf{p}$ in Ω , and $\langle \mathbf{A}, \mathbf{r} \rangle, \langle \mathbf{B}, \mathbf{r}' \rangle \in$ holds in \mathfrak{F} , then $[\mathbf{A} \cup \mathbf{B} \text{ xstit}] \mathbf{p} \in \mathfrak{F}(\Omega)$. So, for each \mathbf{r} mentioned in φ , there is $\langle \mathbf{A}, \mathbf{r} \rangle \in$ holds where \mathbf{A} is the agent term used to replace \mathbf{r} in φ to make δ . So define Q^- as the set of all $\langle \mathbf{A}, \mathbf{r} \rangle$ for each \mathbf{r} mentioned in φ where \mathbf{A} is used to replace \mathbf{r} to make δ . For each $\mathbf{r}' \in \mathbf{Rol} \setminus Q^-$, we choose one pair $\langle \mathbf{B}, \mathbf{r}' \rangle \in$ holds. Call that set Q^+ , then $Q = Q^- \cup Q^+$. We are assuming that all of the roles are used in some member of Ω . So for each $\mathbf{r} \in \mathbf{Rol}$, there is $\langle \mathbf{A}, \mathbf{r} \rangle \in Q$ and \mathbf{A} is unique.

Clearly $\mathfrak{F}_Q(\Omega) \subseteq \mathfrak{F}(\Omega)$ since $Q \subseteq$ holds. Also, since $\mathfrak{F}(\Omega) \not\vdash_{xp}^\Omega \delta$, it follows that $\mathfrak{F}_Q(\Omega) \not\vdash_{xp}^\Omega \delta$. And we will further recognize that $\mathfrak{M}, (s, h) \models \mathfrak{F}_Q(\Omega)$.

Since Ω is \vdash_{Ixp} -consistent, there is a \mathcal{L}_\in^I -model $\mathfrak{M}' = \langle D', \mathfrak{F}'_x, v' \rangle$ and $(s', h') \in \mathfrak{M}'$ such that $\mathfrak{M}', (s', h') \models \Omega$. Now we use \mathfrak{M} , \mathfrak{M}' , and Q to form \mathfrak{M}^+ as in lemma 7.4.1. I.e., let $\mathfrak{M}^+ = \langle D', \langle S, H, E \rangle, v^+ \rangle$ where v^+ is defined as in the previous lemma. But since Ω is a code, i.e., $\Omega \subseteq \mathcal{L}^I$, and $\varphi \in \mathcal{L}^I$, $\mathfrak{M}^+, (s, h) \models \Omega$, and $\mathfrak{M}^+, (s, h) \not\models \varphi$. That means $\Omega \not\vdash_{\text{Ixp}} \varphi$, by completeness. Thus, $\Omega \not\vdash_N \varphi$, a contradiction.

Chapter 8

Normative Consistency

If two duties, equally sacred, conflict, an exercise of the will can settle the conflict, but not a calculation of values.

van Fraassen (1973, p. 9)

In sections 6.2–6.6 we developed a language in which to formulate institutional norms, i.e., classification sentences. That language is the language of institutions, particularly the language that is under institutional control, i.e., the sentences φ for which $Ic(\varphi) = 1$. Within that language we represent the various legal relations discussed by Holfeld (see 6.4). But now we want to take up another topic in the following chapter: we will investigate the notion of *Normative Consistency*. We will first discuss it via examples, then at theoretical level. Finally, we will make those theoretical notions clear within the new formal framework.

8.1 Actual Inconsistencies

A first question to raise is whether there are really inconsistent codes. On the face of it, such codes do exist. For example, the articles 97 and 112 of the Civil Code of Louisiana, at one time, were at odds. “According to the former ‘The minor of either sex, who has attained the competent age to marry, must have received the consent of his father and mother’ But article 112 prescribes that ‘The marriage of minors, contracted without the consent of the father and mother, can not for that cause be annulled’” (Alchourrón, 1996, p. 338). What exactly is the meaning these two statutes? Article 97, says that for a marriage to be valid, it is a necessary condition that the minors’ parents consent to the marriage. But article 112, says that, if for some reason there is a marriage that is valid *without* that necessary condition, the lack of that condition isn’t a reason to annul the marriage.

Article 112 is rather curious since it is really a rule about what can be cited as justification for

the annulment of a marriage, or rather a justification that can't be cited. What seems to follow from the existence of such a rule is that it might be possible for a minor to be married without the consent of his/her parents. The two statutes are not at odds unless there was a marriage, a genuine marriage, between minors. But article 97 seems to preclude that possibility.

What might have happened is that article 97 was brought into existence, and then people noticed that it would be possible to annul many marriages, marriages that took place between two minors, but by the time article 97 was in force those minors were now adults. But article 97 would make those marriages invalid. Since they didn't want article 97 to be applied retroactively, they made article 112. So through interpretation we see that there is no inconsistency because we interpret the putatively inconsistent situations away.

Consider that in the Drafting manual of the state of Maine it says

When law proposed by a bill conflicts with existing law, the existing law should be expressly changed or repealed since the courts are generally unwilling to find that a law is repealed by implication. In addition, a drafter should not rely on general language such as "This Act applies notwithstanding any other law to the contrary" to take care of inconsistent law. Use of such general language is confusing and does not make clear which of several inconsistent laws is to prevail. of Statutes (2009, pp. 29–30)

This directive seems to indicate a preference toward new law over old law, but it recognizes the problem and ambiguity that arises in legal applications of laws that are inconsistent but left in the master book. This is really an imperative to be careful in editing legal texts. Nonetheless, it seems possible that incompatibilities get missed, and so the master book could contain inconsistencies.

But the kind of incompatibility above isn't so terrible. If there were an explicit exception made for preexisting marriages in article 97, then there would be no incompatibility at all. The kinds of incompatibility that are really vicious are those that make a code unusable. Are such incompatibilities possible? Consider the following actual law

§8059. Inconsistent rules

When 2 rules are inconsistent or in conflict with one another, so that compliance with both is impossible, then compliance with either rule shall be deemed to be compliance with the other. [1985, c. 680, 7 (RPR).]

Title 5: ADMINISTRATIVE PROCEDURES AND SERVICES Part 18: ADMINISTRATIVE PROCEDURES Chapter 375: MAINE ADMINISTRATIVE PROCEDURE ACT Subchapter 2: RULEMAKING

This law, also from Maine, provides a way to correct for inconsistencies. Rules that would end someone up in a spot where there is no way to obey one law without being subject to sanctions from another remove the agent's liability to sanction. This doesn't mean that the law as a whole is made consistent, it is just that the agents bound by it, those for which it is in force, are not going to face situations where they have no choice but to be sanctioned. In a sense any action in relation to those incompatible laws would be a violation, but not a prosecutable violation. But not all legal systems have rules like §8059.

What have not, and will not deal with, how to handle inconsistencies. Many actual systems of law use various techniques. For example, *lex posterior derogat lex priori* is a principle used in applying the law that gives priority to older law over new law, unlike the principle from Maine above. But we also have other ways. Sometimes norms are put into hierarchies so that superior laws will override inferior laws. These are ways of actually handling inconsistent law, but it doesn't rule out the possibility of inconsistency. We will not make any detailed recommendations on how one might reason within systems that are genuinely inconsistent. Now we will look at some previous accounts of inconsistency.

8.2 (In)Consistency of Codes

In this section we will discuss normative inconsistency. Ultimately, we will come to rest on an account of normative consistency in the sense of a code being followable. To come to this

conclusion we consider some theories put forward by von Wright (1991), and Hamblin (1972).

8.2.1 Von Wright

To start our formal discussion of normative inconsistency we will refer to von Wright (1991). In that paper von Wright discusses a way of making sense of deontic logic as a logic of rational norm-making. When von Wright says ‘norm’ he means either a proposition ‘ $O\varphi$ ’ or a proposition ‘ $P\varphi$ ’ which are interpreted as ‘ φ is obligatory’ and ‘ φ is permitted’, respectively. The norms $O\varphi$ are called O-norms and $P\varphi$ are called P-norms. Also, von Wright says that *genuine* norms are those where φ isn’t necessarily true or necessarily false.

A necessary condition for rational norm-making, on von Wright’s view, is that norm-makers intend their norms to be followable. This claim leads to a definition of normative inconsistency. Intuitively, a set of norms is normatively consistent if and only if it is followable.

The concept of followability is explained in terms of the contents of sets of norms. The content of $O\varphi$ is φ , and the content of $P\varphi$ is φ . But there are different ways that a set of norms can fail to be followable. A set of obligations can be unfollowable when they conflict. That is, if Γ is a set of O-norms, then Γ is followable if and only if $\{\varphi : O\varphi \in \Gamma\}$ is consistent.

But von Wright notes that a set of P-norms is always followable. Even when $P\neg\varphi$ and $P\varphi$ are both in the set, any agent can do one or the other, they needn’t do both. However, a set of mixed norms, i.e., both P- and O-norms will be unfollowable if something is both obligatory and omissible, i.e., $O\varphi$ and $P\neg\varphi$ are both in the set. This leads von Wright to a definition of normative consistency as (CON_N): for a set of mixed norms Γ , $CON_N(\Gamma)$ iff for each $P\varphi \in \Gamma$, $\{\psi : O\psi \in \Gamma\} \cup \{\varphi\}$ is consistent.

This provides an interesting formalization of normative inconsistency, but the language lacks expressiveness. We can’t really say much with this language. There are two kinds of normative inconsistency. Requiring that something be done, while at the same time permitting that it not be done, that is the kind of inconsistency in the case of mixed norms. The other is having conflicting obligations. Of course the unfollowability of a set of O-norms implies that

the set is not consistent in the CON_N sense. Thus, normative inconsistency can be reduced to the CON_N sense alone. But we might do better if we thought about normative inconsistency in another framework, i.e., one similar to ours.

8.2.2 Hamblin

Hamblin (1972) offers some ideas how norms can conflict with each other, what he calls *quandaries*. A quandary is where there are no ways of acting that are in line with the norms. Everything transgresses the norms somehow.

Mathematically speaking, we can model quandaries in our stit framework in a way similar to what Hamblin does. Recall that each history from an xstit model is a way the world might unfold, and a single transition, i.e., pair of static states $\langle s, s' \rangle$ such that $s' \in \text{lub}(s)$, represents a possible “next step” from s in the model. Hamblin thinks of norms as sets of transitions between the static states in an xstit model. The set of transitions corresponding to each norm represents all of the *illegal transitions* according to a set of norms. So if a transition is in the norm N , then that transition transgresses against the code.

This model for thinking about action and norms allows Hamblin to identify a sense of normative inconsistency that departs from the technical definition von Wright gives in terms of logical consistency, but remains true to the spirit of von Wright’s idea. For Hamblin, inconsistency in a set of norms, i.e., a set of transitions in an xstit model, is for an agent to be placed in a situation where all of the transitions available to her are illegal. This is what, formally, Hamblin calls a *quandary*. We will not discuss the various kinds of quandaries Hamblin distinguishes since we are just using Hamblin’s work to motivate our own. That will be a topic for future work.¹

However, there is an issue to do with expressiveness in Hamblin’s work as well. Norms are simply sets of states of transitions, and so are actions on his account. Thus what is needed to formulate norms syntactically rather than semantically is a formalism that is expressive enough

¹Hamblin’s notions of norms are also similar to the work of Braybrooke et al. (1995), Ågotnes et al. (2007), and Segerberg (2009).

to represent that a set of rules being broken. Of course we have developed such a formalism in \mathcal{L}_{\in}^I .

8.3 Formal Account of Normative Consistency

We look at normative inconsistency via the kinds of quandaries that are possible. But we will set one kind of inconsistency to the side before we venture further. An explicit inconsistency φ and $\neg\varphi$ in a code would make the code problematic, and unusable or unfollowable, but in a rather uninteresting way. Indeed, we are more interested in ways that codes can be problematic although they are *logically consistent*. As it will turn out, if a code is normatively consistent, then it will also be logically consistent, so we will not be worried about our restricted view.

In relation to our discussion of Hamblin we will interpret the notion of normative inconsistency in terms of norms being transgressed. In Hamblin's case, norms are sets of transitions, i.e., pairs of static states (s, s') such that $s' \in \text{lub}(s)$. In the xstit framework norms are formulas, and we can represent when a code has been transgressed by the truth of the violation constant V . To this end, as an abuse of notation let's shorten 'for any $s' \in E(s, h, \mathbf{A})$, and $s' \in h', s', h' \models V$ ', to ' $E(s, h, \mathbf{A}) \subseteq \llbracket V \rrbracket$ '.² Our goal is to represent normative consistency, i.e., followability, by some condition(s) on \mathcal{L} -models of codes.

Now that we have looked at how to represent transgressions of a normative code, we simply define a code in this formal framework to be a set of formulas. Since we have only introduced one violation constant, we are only dealing with one code in this paper, but we could add other violation constants, both to represent different codes and different sorts of violations within a code. A code needn't be a theory since it is supposed to represent the collection of norms explicitly promulgated; what the norm consequences of a code are is another discussion. Now that we have these bits of terminology we can ask: When would it be that a code would really be unfollowable?

²The fact that this formalism is a complete logic with respect to the semantics is also significant. It means that we can discuss things in the semantic setting and translate them back into the syntax. Also, adding the violation constant does nothing to the logic since V is treated simply as an atomic sentence.

To evaluate whether a code is followable we have to look at it in all situations. But, for the moment, we will look at an implementation of a code in all situations. Since we are looking at an implementation of a code, rather than a code, i.e., a set in \mathcal{L} , rather than a set in \mathcal{L}^I , we will refer to codes by Δ s.

We say that a code is ‘in force’ when its norms are in effect, i.e., declared by the institutional authority. In these models the way that we model a code Δ as being in force is for its sentences to be true. To judge whether a code is really followable we have to look at whether it can come into force and stay in force thereafter without causing too many problems. Our goal is to formally describe what it means for a code to cause *too many problems*.

Before we describe the problems formally we have to formalize the notion of a code being in force and staying in force thereafter. We can make sense of the target situation formally as follows.

Definition 8.3.1. We say that a code Δ is *sustained in force after a point* (s, h) in an xstit model \mathfrak{M} , in symbols $\mathfrak{M}, (s, h) \triangleleft \Delta$, iff $\mathfrak{M}, (s, h) \models \Delta$, and for each $(s', h') \in |\mathfrak{M}|$ such that $s \in h'$, and $s \leq_{h'} s'$, $\mathfrak{M}, (s', h') \models \Delta$.

We say that \mathfrak{M} sustains Δ in force, $\mathfrak{M} \triangleleft \Delta$, iff there is $(s, h) \in |\mathfrak{M}|$ such that $\mathfrak{M}, (s, h) \triangleleft \Delta$.

We can also refer to the set of states (s', h') after (s, h) , as

$$\text{AFT}(s, h) = \{ (s', h') \in |\mathfrak{M}| : s \in h' \ \& \ s \leq_{h'} s' \}$$

We will say that $s' \in \text{AFT}(s, h)$ iff there is h' with $s' \in h'$ and $(s', h') \in \text{AFT}(s, h)$.

Recall that $|\mathfrak{M}| = \{ (s, h) \in S \times H : s \in h \}$. The idea is that there is some state at which Δ is satisfied, and it is satisfied at every “state” after that one. Every state that is potentially “after” (s, h) , is a way the world could turn out after (s, h) . This definition requires any code that can be sustained in force to be xstit-consistent (by completeness).

Following Hamblin’s ideas, a code is bad when it puts its subjects into a quandary, i.e., leaves them with no legal continuations. In this framework a legal continuation is a path from (s, h) to (s', h') where $s \in h'$, $\text{lub}(s, h') = s'$ and $s' \notin \llbracket V \rrbracket$. So \mathbf{A} is in a quandary at s when

for each $h \ni s$, $E(s, h, \mathbf{A}) \subseteq \llbracket V \rrbracket$. But the mere existence of a situation where an agent/group is in a quandary isn't by itself damning for a code: it may be possible to make a series of bad decisions and end up in a quandary. But that series of decisions isn't the fault of the code, it is the fault of the agent. If quandaries are too easy to come by, then we can say that the code is unusable. Thus we have to provide a formal way of characterizing what it means to say that a quandary is "too easy to come by".

Part of being in a quandary is to be in a position to xstit a violation, i.e., $E(s, h, \mathbf{A}) \subseteq \llbracket V \rrbracket$. A code that could never lead to any quandary would be a kind of utopian code. So clearly isolated incidents of xstit-ing V is a kind of unproblematic consequence of the application of a code: agents should be capable of breaking the rules. We will say that \mathbf{A} is in a *Bad Situation* at (s, h) when $E(s, h, \mathbf{A}) \subseteq \llbracket V \rrbracket$. What we should be worried about are codes that give rise to quandaries in too many cases; those are the genuine quandaries. Now we have to explain what makes up a 'case'.

The 'situations' in the logic of xstit can be represented by the dynamic states in a model since those are the points of evaluation. However, at any dynamic state in a model, there is also the set of agents. So situations are composed of: a model, a dynamic state in that model, and a set of agents. One benefit of introducing formalism is the ability to precisely characterize a totality of possible situations: it makes the precise expression of *all* possibilities *possible*. In our case, there are three parameters to consider in how frequently bad situations come about after a code is sustained in force. Frequency, logically speaking, can only be represented by the two quantifiers *all* and *some*.

These three parameters allow us to formulate properties of classes of xstit models that sustain a code Δ in force that represent putative ways that a code can be unfollowable. These properties are what we will call *quandaries*. Each of the three parameters corresponds to a kind of restricted quantifier (either \forall or \exists), and alternating the quantifiers in relation to a code Δ gives all of the possible, relevant conditions on models. The three restricted quantifiers are: (A) $\forall/\exists \mathfrak{M}$, $(s, h) \leq \Delta$, (B) $\forall/\exists (s', h') \in \text{AFT}(s, h)$, and (C) $\forall/\exists \mathbf{A} \subseteq \mathbf{Ag}$. This means that we can

get a well defined range of quandaries by looking at all of the ways to arrange these quantifiers.

There are certain restrictions on the permutations of the quantifiers as follows: The (s, h) -values in the kind A quantifier depend on the model \mathfrak{M} , so we cannot alternate the A and B quantifiers like $\forall(s, h)\forall\mathfrak{M}$. However, the agent term variables \mathbf{A} are not model dependent; they are part of the language. So alternating the type A and C quantifiers as in $\forall\mathbf{A}\forall\mathfrak{M}$ is intelligible. But since a permutation like $\forall\mathbf{A}\forall\mathfrak{M}$ will be equivalent to one like $\forall\mathfrak{M}\forall\mathbf{A}$, we will only represent the latter. Given the intelligibility of these arrangements of quantifiers, we have the full range of possibilities for quandaries. Each potential quandary will take the form $Q_1, Q_2, Q_3, E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$ where Q_1 is either $\forall/\exists\mathfrak{M}$ or $\forall/\exists\mathbf{A} \subset \mathbf{Ag}$, Q_2 is either $\forall/\exists\mathfrak{M}$ or $\forall/\exists\mathbf{A} \subset \mathbf{Ag}$ or $\forall/\exists(s', h') \in \text{AFT}(s, h)$, and Q_3 is either $\forall/\exists\mathbf{A} \subset \mathbf{Ag}$ or $\forall/\exists(s', h') \in \text{AFT}(s, h)$. To make this a bit clearer we list all of the combinations below.

1. $\forall\mathfrak{M}, (s, h) \triangleleft \Delta, \forall\mathbf{A} \subseteq \mathbf{Ag}, \forall(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
2. $\forall\mathfrak{M}, (s, h) \triangleleft \Delta, \forall\mathbf{A} \subseteq \mathbf{Ag}, \exists(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
3. $\forall\mathfrak{M}, (s, h) \triangleleft \Delta, \exists\mathbf{A} \subseteq \mathbf{Ag}, \forall(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
4. $\forall\mathfrak{M}, (s, h) \triangleleft \Delta, \exists\mathbf{A} \subseteq \mathbf{Ag}, \exists(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
5. $\exists\mathfrak{M}, (s, h) \triangleleft \Delta, \forall\mathbf{A} \subseteq \mathbf{Ag}, \forall(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
6. $\exists\mathfrak{M}, (s, h) \triangleleft \Delta, \forall\mathbf{A} \subseteq \mathbf{Ag}, \exists(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
7. $\exists\mathfrak{M}, (s, h) \triangleleft \Delta, \exists\mathbf{A} \subseteq \mathbf{Ag}, \forall(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
8. $\exists\mathfrak{M}, (s, h) \triangleleft \Delta, \exists\mathbf{A} \subseteq \mathbf{Ag}, \exists(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
9. $\forall\mathfrak{M}, (s, h) \triangleleft \Delta, \forall(s', h') \in \text{AFT}(s, h), \exists\mathbf{A} \subseteq \mathbf{Ag}, E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
10. $\exists\mathfrak{M}, (s, h) \triangleleft \Delta, \forall(s', h') \in \text{AFT}(s, h), \exists\mathbf{A} \subseteq \mathbf{Ag}, E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
11. $\exists\mathbf{A} \subseteq \mathbf{Ag}, \forall\mathfrak{M}, (s, h) \triangleleft \Delta, \exists(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$
12. $\exists\mathbf{A} \subseteq \mathbf{Ag}, \forall\mathfrak{M}, (s, h) \triangleleft \Delta, \forall(s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$

$$13. \forall \mathbf{A} \subseteq \mathbf{Ag}, \exists \mathfrak{M}, (s, h) \triangleleft \Delta, \exists (s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$$

$$14. \forall \mathbf{A} \subseteq \mathbf{Ag}, \exists \mathfrak{M}, (s, h) \triangleleft \Delta, \forall (s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$$

There are four possibilities that are not listed here since they just involve $\mathfrak{M}, (s, h) \triangleleft \Delta$, and $(s', h') \in \text{AFT}(s, h)$. Quandaries not involving reference agents we call total quandaries or T-quandaries for short. This is to say that whether there is a bad situation “after” the code has come into force is independent of the groups. The possible T-quandaries are

$$\text{T1 } \forall \mathfrak{M}, (s, h) \triangleleft \Delta, \forall (s', h') \in \text{AFT}(s, h), \text{lub}(s') \subseteq \llbracket V \rrbracket$$

$$\text{T2 } \forall \mathfrak{M}, (s, h) \triangleleft \Delta, \exists (s', h') \in \text{AFT}(s, h), \text{lub}(s') \subseteq \llbracket V \rrbracket$$

$$\text{T3 } \exists \mathfrak{M}, (s, h) \triangleleft \Delta, \forall (s', h') \in \text{AFT}(s, h), \text{lub}(s') \subseteq \llbracket V \rrbracket$$

$$\text{T4 } \exists \mathfrak{M}, (s, h) \triangleleft \Delta, \exists (s', h') \in \text{AFT}(s, h), \text{lub}(s') \subseteq \llbracket V \rrbracket$$

So each of these conditions 1–12, T1–4 may represent a problem for a code Δ . Each condition says that the class of models that sustains the code Δ in force has a certain property, i.e., leads to a bad situation “frequently”. But now we can do some pruning to cut away unproblematic conditions, conditions that are too strict on what models are problematic. Any condition that has ‘ $\exists \mathfrak{M}$ ’ in it, simply says that there is a model where there is some—even many—bad situations. These conditions aren’t really problematic, they are expected. Even a case where there is a model where every situation is bad for every group does not spell disaster for a code. The intuition is that it should be possible for everyone to be bad all of the time. It shouldn’t be an act of logic or of legislation that at least someone is good at each moment. So we can ignore cases 5–8 and 10, 13 and 14, as well as T3 and T4.

We focus on the first list of conditions and consider, in particular, the first four. Condition 4 says that in each model there is a group that is capable of being in a bad situation at some point. And what we say is that condition 4 can be true of the class of models that make the code sustained in force without the code being problematic, so we won’t require its falsity. Condition 2 is problematic, i.e., it does represent a kind of inconsistency, but we will deal with that later.

Condition 11 says that there is a particular group that in any model *can* be in trouble. This condition is a bit more worrisome since it is kind of discriminatory. That one particular group must be extra careful and not to step out of line. Of course we have groups like that, e.g., police. So, intuitively, a particular group being capable of being in trouble shouldn't worry us.

The really worrisome conditions are those where someone is always in trouble in every model after the code is sustained in force. The conditions that instantiate that worry are 1, 3, 9, and 12. In the case of 1, in every model, everyone is always in a bad situation. That's bad when it is forced by legislation. Since in every model everyone is in a bad situation always, we can say that that condition is due to the formulation of the code and not the actions of the code's subjects. Similarly with condition 3. In each model, there is a group that is always in trouble. That is discrimination, it is only a misanthropic code that would require that somebody must be persecuted. Condition 9 is also a problem since it is selective discrimination; some group is preselected to always in trouble in every model. Finally, in condition 12 someone/group is always in trouble, again it is a kind of misanthropic code that would force that condition. So we suggest that codes that force conditions like 1, 3, 9 or 12 are bad. We are indifferent to the rest of the conditions from the first list.

In the second list the remaining possibilities are T1 and T2. T1 says that in every model every situation after the code comes into force is a bad situation. That is clearly undesirable. Condition T2 is also, in a sense undesirable. It says that in every model, after the code is in force we are guaranteed some point at which things go bad for everyone. That means that it is just a matter of time before things go bad, and that is a problem for a code.

Now we can deal with condition 2. Notice that condition 2 is equivalent to T2. If condition 2 holds, then for any model that sustains Δ in force at (s, h) , every group can, at some point in $\text{AFT}(s, h)$, xstit a violation. So that is true of \emptyset , and $\text{lub}(s') = E(s', h', \emptyset) \subseteq \llbracket V \rrbracket$ so T2 is true. And if T2 is true, then $E(s', h', \emptyset) = \text{lub}(s') \subseteq \llbracket V \rrbracket$, so for any \mathbf{A} , \mathbf{A} will xstit a violation at (s', h') . So condition 2 is true as well. So although condition 2 is problematic, it can be folded into T2.

Now let's give these special conditions some names to make them easy to refer to. We have already introduced the T-quandary terminology for quandaries that result from V being true everywhere. What we will do now is refine these a bit. There is another kind of total quandary that might arise where a particular state s is such that $\text{lub}(s) \subseteq \llbracket V \rrbracket$ after the code comes into force whenever it comes into force. This means that there is no way to proceed without everyone ending up in a quandary, and if that follows by legislation that is a problem. We call that a TE-quandary (total existential quandary). Condition T1 is a T-quandary and condition T2 is a TE-quandary. Note that if there is a T-quandary, then there is a TE-quandary.

We introduce another kind of quandary called a global quandary or G-quandary for short. We also introduce a second kind of G-quandary that we call a GE-quandary to parallel the T-types. The G-types are made up of conditions 1 and 9. In condition 1, the quandary is global since everyone is in trouble. Condition 9 is global in the sense that in every situation someone is in trouble; who is in trouble may depend on the dynamic state. Finally, we will call conditions 3 and 12 Discriminatory Quandaries (DE- and D-quandaries, respectively). This is because they are similar in that they discriminate against at least one particular group independent of some aspect of the situation. In 3 that group may depend on the model, but not the dynamic state. In 12 it doesn't depend on either the dynamic state or the model. Again we note that D-quandaries imply DE-quandaries.

One final point before we give the definition of the quandaries. We will require that $\mathfrak{M} \triangleleft \Delta$ is true for Δ to have a quandary. To define quandaries we needed to extend the notion of when a code was in force. When a code is in force it is used to evaluate any situation, not just the current situation; the code is in force everywhere/when. If it isn't possible to sustain a code in force, then it isn't really possible to use that code to evaluate every situation. But we needed a notion stronger than consistency to guarantee that. To see this reason better, consider a code Δ that has a model, i.e., there is (s, h) in \mathfrak{M} that satisfies Δ , then it may not be the case that it is satisfied everywhere in that model. In fact it may not be *possible* to satisfy it everywhere after a certain state. Consider the set $\{ X \neg \mathbf{p}, \mathbf{p} \}$. If that is to be true everywhere after the current state

as well, then it must be true at the successor state. So at the current state $X \neg \mathbf{p}$ would have to be true, but then both \mathbf{p} and $\neg \mathbf{p}$ would have to be true at the successor state. So we want to restrict the notion of a code having a quandary to non-trivially having a quandary in the sense that the code could be used, but gave rise to problems. This means that there is \mathfrak{M} , such that $\mathfrak{M} \triangleleft \Delta$ as in definition 8.3.1 below.

So we can finally arrive at a definition for the types of quandaries.

Definition 8.3.2. We say that a code Δ , such that there is \mathfrak{M} and $\mathfrak{M} \triangleleft \Delta$, has a

1. T-quandary iff every model \mathfrak{M} with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$ is such that

$$\forall s' \in \text{AFT}(s, h), \text{lub}(s') \subseteq \llbracket V \rrbracket$$
2. TE-quandary iff every model \mathfrak{M} with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$ is such that

$$\exists s' \in \text{AFT}(s, h), \text{lub}(s') \subseteq \llbracket V \rrbracket$$
3. G-quandary iff every model \mathfrak{M} with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$ is such that

$$\forall \mathbf{A} \subseteq \mathbf{Ag}, \forall (s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$$
4. GE-quandary iff every model \mathfrak{M} with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$ is such that

$$\forall (s', h') \in \text{AFT}(s, h), \exists \mathbf{A} \subseteq \mathbf{Ag}, E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$$
5. D-quandary iff $\exists \mathbf{A} \subseteq \mathbf{Ag}$ such that for every model \mathfrak{M} with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$,

$$\forall (s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$$
6. DE-quandary iff every model \mathfrak{M} with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$ is such that

$$\exists \mathbf{A} \subseteq \mathbf{Ag}, \forall (s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$$

There is another philosophical question worth discussing briefly: what about responsibility? If a code has a T or TE-quandary (a T-type quandary), then no one is properly responsible for the violations. Indeed, no one need be responsible in the sense of *xdstit* for any quandary. If a code has a T-type quandary, there is definitely a problem, but we can't hold anyone responsible for the badness. But, one might argue, we recognize that these quandaries aren't anyone's fault, so we won't try to punish anyone for something they couldn't help. The thought that

responsibility is important for justly administering punishment is a conceptual one to do with the law or morality. Even if no one is responsible for the violations, that doesn't make the conditions the code imposes useful. Indeed, when there is a T-quandary the code doesn't allow us to distinguish which actions should be "really counted" as wrongdoing. If a code doesn't allow anyone to do something right, ever, that isn't a very useful code.

Now we can raise two questions about what kinds of codes lead to G, T, or D-type quandaries. The definitions of the quandaries were given in terms of conditions on models, so we may want to attempt to characterize those conditions by what those codes can prove, the consequences of those codes. So we can first ask: what conditions on the consequences of codes correspond, if any, to the quandaries? The second question to raise is whether there are connections between the kinds of quandaries.

As it turns out the conditions on quandaries are very similar given the formalism that we are working with. We can connect them according to the following theorem where the arrow is implication: $X \implies Y$ means If Δ has an X -type quandary, then Δ has a Y -type quandary.

Theorem 8.3.1.

$$\begin{array}{ccccc}
 G & \iff & D & \iff & T \\
 \Downarrow & & \Uparrow & & \Downarrow \\
 GE & \iff & DE & & TE
 \end{array}$$

This is proved in section 8.4. The upshot of this theorem is that the only difference between the quandaries, really, is 5/6 of the quandaries are equivalent and TE on the other. But the other five all imply TE. This result is definitely an artifact of the formalism. The big jumps, like from D to T, that cause the collapse are forced because of the anti-monotonicity of the effectivity function, i.e., condition f, and the fact that $E(s, h, \mathbf{Ag}) = \{lub(s, h)\}$ for any $s \in h$, i.e., condition e. Giving up either of condition e or f would help separate these conditions. However, conditions e and f can be justified by referencing certain intuitions about action, condition f in particular. That investigation will be passed to future work. Does this mean that the formalism is useless in distinguishing the kinds of quandaries? No, there is still the distinction between the

TE quandaries and the rest. But this result permits a simple definition of normative consistency that we give below.

What this means from the semantic side of things is that we can define the consistency of a normative system as quandary freeness, and all that we need to ensure quandary freeness is that there is no TE-quandary. Thus we are in a position much like that of von Wright's condition. However, it doesn't reduce directly to regular inconsistency since to have a TE-quandary the code must be consistent.

Notice that the notions of D and DE quandary are not usually considered *logical* notions of inconsistency. They aren't conditions that have to do with whether the code is followable. Even if the code discriminates against some group, it is still followable by others. That kind of condition is usually described as a code being *unfair* or *unjust*. But in the xstit framework unfairness is on par with other notions of followability since it is extensionally the same as those other notions.

We can move on to our second question about characterizing the quandaries in terms of provability. We don't characterize them in terms of normative consequence, because we have to look at the code in a different light. We will see how this works soon. We first have to build up some technical results to support our investigation.

Given a set Δ define a set Δ_{if} as follows:

$$\Delta_{if} = \{ \Box X^n \delta \mid \delta \in \Delta \text{ \& } n \in \mathbb{N} \} \quad (8.1)$$

So for any $\delta \in \Delta$, we have $\Box \delta \in \Delta_{if}$, $\Box XXXX \delta \in \Delta_{if}$, but also $XXXX \delta \in \mathbb{C}_{\vdash_{xp}}(\Delta_{if})$ and $\delta \in \mathbb{C}_{\vdash_{xp}}(\Delta_{if})$. The latter two facts follow since \Box is an S5 modality. Now we can make an observation:

Observation 8.3.2. $\mathfrak{M}, (s, h) \triangleleft \Delta$ if and only if $\mathfrak{M}, (s, h) \models \Delta_{if}$.

We require one more lemma.

Lemma 8.3.3. Suppose that there is a $(s, h) \in |\mathfrak{M}|$, such that $\mathfrak{M}, (s, h) \models \Delta_{if}$, and $(s', h') \in \text{AFT}(s, h)$. Then $\mathfrak{M}, (s', h') \models \Delta_{if}$

Proof. Suppose that there is a $(s, h) \in |\mathfrak{M}|$, such that $\mathfrak{M}, (s, h) \models \Delta_{if}$, and $(s', h') \in \text{AFT}(s, h)$. By observation 8.3.2, $\mathfrak{M}, (s, h) \prec \Delta$. But that means that at each $(s'', h'') \in \text{AFT}(s, h)$, $\mathfrak{M}, (s'', h'') \models \Delta$. So if $n \in \mathbb{N}$, $\delta \in \Delta$, $\Box X^n \delta$ is always true at (s', h') since δ is true everywhere after that point a fortiori. Thus $\mathfrak{M}, (s', h') \models \Delta_{if}$. \square

Now that we have these results we can show how to define conditions on provability that will translate into quandaries and back. Because we have theorem 8.3.1, we really only need two results, one for T and one for TE. However, we can provide more specific results for most of the types of quandaries. We collect the conditions together in theorem 8.3.4. Recall that the set \mathbf{Ag} is finite, and so $\mathcal{P}(\mathbf{Ag})$ is finite. As another bit of notational convenience let $\Delta_V = \{ \neg \Diamond X^n \Box X V \mid n \in \mathbb{N} \}$.

Theorem 8.3.4. *If Δ is a code, and $\Delta_{if} \not\vdash_{\text{xp}} \perp$ then*

1. Δ has a T quandary iff $\Delta_{if} \vdash_{\text{xp}} \Box X V$
2. Δ has a TE quandary iff $\Delta_{if} \cup \Delta_V \vdash_{\text{xp}} \perp$
3. Δ has a G quandary iff $\Delta_{if} \vdash_{\text{xp}} \bigwedge_{\mathbf{A} \subseteq \mathbf{Ag}} [\mathbf{A} \text{ xstit}] V$
4. Δ has a GE quandary iff $\Delta_{if} \vdash_{\text{xp}} \bigvee_{\mathbf{A} \subseteq \mathbf{Ag}} [\mathbf{A} \text{ xstit}] V$
5. Δ has a D quandary iff $\Delta_{if} \vdash_{\text{xp}} [\mathbf{A} \text{ xstit}] V$ for some $\mathbf{A} \subseteq \mathbf{Ag}$

The proof appears below. There isn't a clear condition that matches up with the DE case. But because of the equivalence of DE to D, we don't have to worry about representing the condition here. We want to point out an interesting fact as an aside. The TE condition is unlike the others since it put in terms of Δ_{if} 's inconsistency with Δ_V . As is used in the proof in the appendix, that conditions means, by compactness of \vdash_{xp} , that there is some finite set $X \subseteq \mathbb{N}$ such that $\Delta_{if} \cup \{ \Diamond X^n \Box X V : n \in X \} \vdash_{\text{xp}} \perp$. And so there must be a largest n in that set. That means if $\mathfrak{M}, (s, h) \models \Delta_{if}$, there is an upper bound on how far into the future one has to go before $\text{lub}(s') \subseteq \llbracket V \rrbracket$.

But now we are finally in a position to take up the notion of normative consistency for our system. Standardly, one has a notion of consequence, \vdash , and then one defines inconsistency as proving some formula, or some set of formulas, C . Then consistency can be defined as $\Delta \not\vdash C$. Here we take a slightly different view since consequence is normative consequence \vdash_N , and that isn't what is used to define quandaries.

Normative consistency CON_N will have something to do with the existence of quandaries. What is nice in the current situation is that if there is a quandary of any type (T,G,D), then there is a TE-quandary. But that means if we say that a code doesn't have a TE-quandary, then it won't have any of the problematic quandaries either. But there is also the situation of standard inconsistency, i.e., $\Delta \vdash_{\text{xp}} \perp$. A code that is simply inconsistent is rather problematic, at least from a logical standpoint. As we discussed in section 8.2, in practice we find ways to get along without the code when things are broken. There may be ways to model that, but in this logical setting we are setting those possibilities aside.

Now a code Δ that is inconsistent can't be made true at all, so it can't do its job in our setting. But it is also a problem if a code isn't able to be used to evaluate all situations, i.e., $\Delta_{if} \vdash_{\text{xp}} \perp$. As we noticed in observation 8.3.2, the consistency of Δ_{if} is necessary and sufficient for Δ to be usable as a code. By the transitivity of \vdash_{xp} , if $\Delta \vdash_{\text{xp}} \perp$, then $\Delta_{if} \vdash_{\text{xp}} \perp$ since $\Delta \subseteq \mathbb{C}_{\vdash_{\text{xp}}}(\Delta_{if})$. That means the consistency of Δ_{if} implies the consistency of Δ . Thus part of being a normatively consistent code is for Δ_{if} to be consistent. The other part, as you might have guessed, is for there to be no TE-quandary.

But for there to be a TE quandary, it must be that $\Delta_{if} \cup \Delta_V \vdash_{\text{xp}} \perp$. If $\Delta_{if} \vdash_{\text{xp}} \perp$, then by monotonicity of \vdash_{xp} , $\Delta_{if} \cup \Delta_V \vdash_{\text{xp}} \perp$. So if there is no TE-quandary, then $\Delta_{if} \not\vdash_{\text{xp}} \perp$. Thus if no TE-quandary, Δ is normatively consistent. So far we have been discussing quandaries and normative consistency in the setting of \mathcal{L} , but codes are formulated in \mathcal{L}^I . For this reason we have the following definition:³

Definition 8.3.3. A code $\Omega \subseteq \mathcal{L}^I$ is normatively inconsistent, relative to an implementation \mathfrak{Z}

³This notion is a much like a formal version the notion of deontic consistency from Marcus (1980).

iff $\mathfrak{Z}(\Omega)$ has a TE-quandary. So we can define normative inconsistency $\overline{\text{CON}}_{\mathfrak{N}}^{\mathfrak{Z}}(\Omega)$ formally as follows

$$\overline{\text{CON}}_{\mathfrak{N}}^{\mathfrak{Z}}(\Omega) \iff [\mathfrak{Z}(\Omega)_{if} \cup \Delta_V \vdash_{\text{xp}}^{\Omega} \perp]$$

This is an implementation relative notion of consistency. So it remains to discuss what role the implementation plays in the normative consistency of Ω tout court. There are only two possibilities: either it is normatively consistent relative to all implementations of a code, or just some. But as in the previous discussion about which possible conditions for quandaries are problematic, there simply being a way for things to go wrong doesn't seem too bad. Of course, if there exists an implementation that implies a T-quandary that is problematic. But all that means is that we have to be mindful about the implementation. So again, we go with the universal quantifier:

Definition 8.3.4. A code Ω is normatively inconsistent iff every implementation $\mathfrak{Z}(\Omega)$ has a TE-quandary. So we can define normative inconsistency $\overline{\text{CON}}_N(\Omega)$ formally as follows,

$$\overline{\text{CON}}_N(\Omega) \iff \forall \mathfrak{Z}, [\mathfrak{Z}(\Omega)_{if} \cup \Delta_V \vdash_{\text{xp}}^{\Omega} \perp]$$

8.4 Proofs from Section 8.3

Proof of theorem 8.3.1. We will proceed by first noticing that in each type of quandary X, X implies XE for X=T,G,D. Also notice the rather trivial result that if Δ has a T-quandary, then Δ has both G and D-quandaries too. Let \mathfrak{M} be any model of Δ such that $(s, h) \in |\mathfrak{M}|$, and $\mathfrak{M}, (s, h) \prec \Delta$. If Δ has T quandary, all states after the code is in force only have successor states that are V states, i.e., for all $(s', h') \in \text{AFT}(s, h)$, $\text{lub}(s') \subseteq \llbracket V \rrbracket$. So whatever any group is effective for is going to be a set of V states, i.e., for all $A \subseteq \mathbf{Ag}$, and $(s', h') \in \text{AFT}(s, h)$, $E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$ since $E(s', h', \mathbf{A}) \subseteq \text{lub}(s')$. Since \mathfrak{M} was arbitrarily chosen, Δ has a G-quandary. Under the same condition of Δ having a T-quandary, for any $\mathfrak{M}, (s, h) \prec \Delta$, and $(s', h') \in \text{AFT}(s, h)$, $E(s', h', \emptyset) = \text{lub}(s') \subseteq \llbracket V \rrbracket$. But that will be true for $\mathbf{A} = \emptyset \subseteq \mathbf{Ag}$ for any model, so Δ has a D-quandary.

Now we show that G implies T. Suppose Δ has a G-quandary. Then in any model with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$, every group is such that it is always effective for violations, i.e., for all $\mathbf{A} \subseteq \mathbf{Ag}$, and $(s', h') \in \text{AFT}(s, h)$, $E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$. But that means \emptyset is effective for violations, i.e., $E(s', h', \emptyset) \subseteq \llbracket V \rrbracket$ for all $(s', h') \in \text{AFT}(s, h)$. But that is just to say that $\text{lub}(s') \subseteq \llbracket V \rrbracket$ by condition d on effectivity functions. Since \mathfrak{M} was arbitrarily chosen Δ has a T-quandary.

To show D implies T suppose Δ has a D-quandary, i.e., there is \mathbf{A} such that for any model with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$, for all $(s', h') \in \text{AFT}(s, h)$, $E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$. But then by the anti monotonicity of E and because $\mathbf{A} \subseteq \mathbf{Ag}$, $E(s', h', \mathbf{Ag}) \subseteq \llbracket V \rrbracket$. But that holds for every $s' \in h'$. Let $s'' \in \text{lub}(s')$, arbitrary s' from $\text{AFT}(s, h)$. Then there is h'' such that $s'' \in h''$ and (s'', h'') is in $\text{AFT}(s, h)$ since s' is. But also $\text{lub}(s', h'') = (s'', h'')$, and $E(s', h'', \mathbf{Ag}) \subseteq \llbracket V \rrbracket$ (because it holds for all elements of $\text{AFT}(s, h)$). But then $(s'', h'') \models V$, so $s'' \in \llbracket V \rrbracket$. Since s'' was arbitrary, $\text{lub}(s') \subseteq \llbracket V \rrbracket$, i.e., Δ has a T-quandary.

From DE to GE is simple since if for each model with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$, there is an \mathbf{A} , call it \mathbf{B} , such that for any $(s', h') \in \text{AFT}(s, h)$, $E(s', h', \mathbf{B}) \subseteq \llbracket V \rrbracket$, then for any $(s', h') \in \text{AFT}(s, h)$ there is an \mathbf{A} , viz. \mathbf{B} , such that $E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$, i.e., Δ has a GE-quandary

So we are left with showing that GE implies G. Suppose that for each model with (s, h) that $\mathfrak{M}, (s, h) \triangleleft \Delta$, and any $(s', h') \in \text{AFT}(s, h)$, there is $\mathbf{A} \subseteq \mathbf{Ag}$, such that $E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$, i.e., Δ has a GE-quandary. We want to show that Δ has a G-quandary, that is for any $\mathfrak{M}, (s, h) \triangleleft \Delta$, and $\mathbf{A} \subseteq \mathbf{Ag}$, and $(s', h') \in \text{AFT}(s, h)$, $E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$. So let $\mathfrak{M}, (s, h) \triangleleft \Delta$, and $\mathbf{B} \subseteq \mathbf{Ag}$ and $(s', h') \in \text{AFT}(s, h)$. We will show that $\text{lub}(s') \subseteq \llbracket V \rrbracket$. That will imply $E(s', h', \mathbf{B}) \subseteq \llbracket V \rrbracket$, and since \mathbf{B} and (s', h') were chosen arbitrarily, Δ will have a G-quandary.

Suppose that $h'' \ni s'$. Then $(s', h'') \in \text{AFT}(s, h)$, so there is \mathbf{A}' such that $E(s', h'', \mathbf{A}') \subseteq \llbracket V \rrbracket$. By the anti-monotonicity of E , $E(s', h'', \mathbf{Ag}) \subseteq \llbracket V \rrbracket$ so $\text{lub}(s', h'') \in \llbracket V \rrbracket$. Since h'' was arbitrarily chosen, $\text{lub}(s') \subseteq \llbracket V \rrbracket$ as we wanted. So Δ has a G-quandary.

Just for fun we can show that GE implies DE. If GE is true, then for each model with (s, h)

that $\mathfrak{M}, (s, h) \triangleleft \Delta$, and any $(s', h') \in \text{AFT}(s, h)$, there is $\mathbf{A} \subseteq \mathbf{Ag}$, such that $E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$. As we noticed, that means $E(s', h', \mathbf{Ag}) \subseteq \llbracket V \rrbracket$, for each $(s', h') \in \text{AFT}(s, h)$. But then T is true, and so D is true and D implies DE. \square

Proof of observation 8.3.2. Suppose $\mathfrak{M}, (s, h) \models \Delta_{if}$. Note that $\mathfrak{M}, (s, h) \models \Delta$ since $\Box[\Delta] \subseteq \Delta_{if}$ and $\Box[\Delta] \vdash_{\text{xp}} \delta$ for each $\delta \in \Delta$, and Ixstit is sound. Then let (s', h') be “after” (s, h) , so $s, s' \in h'$, and let $\delta \in \Delta$. Suppose, without loss of generality, that s' is the n -successor of h' from s . Then $\Box X^n \delta \in \Delta_{if}$ by definition. But that means $(s, h) \models \Box X^n \delta$, $(s, h') \models X^n \delta$ and $(s', h') \models \delta$ also by definition. Since δ was arbitrarily chosen, $(s', h') \models \Delta$. (s', h') was also arbitrary, so $\mathfrak{M}, (s, h) \triangleleft \Delta$.

For the other direction suppose $\mathfrak{M}, (s, h) \triangleleft \Delta$. Then let $\delta \in \Delta$ (i.e., $\Box X^n \delta \in \Delta_{if}$). Clearly, $(s, h) \models \Delta$ since $\Box[\Delta] \subseteq \Delta_{if}$. Let h' be such that $s \in h'$. Take the n -th h' -successor of s , call it s' , then by assumption $(s', h') \models \delta$. But that means that $(s, h') \models X^n \delta$, and since h' was arbitrarily chosen, $(s, h) \models \Box X^n \delta$, and since n was also arbitrary this holds for all $n \in \mathbb{N}$. That means Δ_{if} is satisfied at (s, h) . \square

Proof of theorem 8.3.4. Here we will prove the T case, the TE case and the GE case, all of the other cases proceed in a similar manner.

(T case) $[\Rightarrow]$ Suppose that Δ has a T quandary. So every model \mathfrak{M} with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$ is such that $\forall s' \in \text{AFT}(s, h), \text{lub}(s') \subseteq \llbracket V \rrbracket$, and there is $\mathfrak{M}' \triangleleft \Delta$. That means $\mathfrak{M}', (s'', h'') \models \Delta_{if}$, for some (s'', h'') and so by completeness $\Delta_{if} \not\vdash_{\text{xp}} \perp$. Let $\mathfrak{M}, (s, h) \models \Delta_{if}$. By observation 8.3.2 $\mathfrak{M}, (s, h) \triangleleft \Delta$. But that means $\text{lub}(s') \subseteq \llbracket V \rrbracket$ for any s' in $\text{AFT}(s, h)$. Suppose $s \in h'$. Then $(s, h') \in \text{AFT}(s, h)$, so $\text{lub}(s) \subseteq \llbracket V \rrbracket$. But that means $\mathfrak{M}, (s, h') \models XV$, and since h' was arbitrary $\mathfrak{M}, (s, h) \models \Box XV$. Since $\mathfrak{M}, (s, h)$ was arbitrary, $\Delta_{if} \models \Box XV$, and by completeness $\Delta_{if} \vdash_{\text{xp}} \Box XV$.

(T case) $[\Leftarrow]$ Suppose that $\Delta_{if} \vdash_{\text{xp}} \Box XV$ and $\Delta_{if} \not\vdash_{\text{xp}} \perp$, then by soundness $\Delta_{if} \models \Box XV$, and by completeness and observation 8.3.2 there is \mathfrak{M}' such that $\mathfrak{M}' \triangleleft \Delta$. Suppose $\mathfrak{M}, (s, h) \triangleleft \Delta$, and let $s' \in \text{AFT}(s, h)$. Then $(s', h') \models \Delta_{if}$ by lemma 8.3.3 for any h' with $s \in h'$, so

$(s', h') \models \Box XV$. But that happens only when $\text{lub}(s') \subseteq \llbracket V \rrbracket$. Since s' was arbitrary it holds for any s' , and since $\mathfrak{M}, (s, h)$ was arbitrary, T holds.

(TE case) $[\Rightarrow]$ Suppose that Δ has a TE quandary. So every model \mathfrak{M} with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$ is such that $\exists s' \in \text{AFT}(s, h), \text{lub}(s') \subseteq \llbracket V \rrbracket$, and there is $\mathfrak{M}' \triangleleft \Delta$. The latter assumption means $\mathfrak{M}', (s'', h'') \models \Delta_{if}$ for some (s'', h'') , and so by completeness $\Delta_{if} \not\vdash_{\text{xp}} \perp$. Let $\mathfrak{M}, (s, h) \models \Delta_{if}$. Then, $\mathfrak{M}, (s, h) \triangleleft \Delta$ by observation 8.3.2, and that means there is $s' \in \text{AFT}(s, h)$ such that $\text{lub}(s') \subseteq \llbracket V \rrbracket$. And there is a history h' with $s' \in h'$. s' must be the n th h' -successor from s for some $n \in \mathbb{N}$, and $(s', h') \models \Box XV$. But that means, $(s, h') \models X^n \Box XV$. And so $(s, h) \models \Diamond X^n \Box XV$. Since \mathfrak{M} and (s, h) were arbitrary, there are no models of $\Delta_{if} \cup \{ \neg \Diamond X^n \Box XV \mid n \in \mathbb{N} \}$. That means, by completeness of \vdash_{xp} , $\Delta_{if} \cup \{ \neg \Diamond X^n \Box XV \mid n \in \mathbb{N} \} \vdash_{\text{xp}} \perp$.

(TE case) $[\Leftarrow]$ Suppose that $\Delta_{if} \cup \Delta_V \vdash_{\text{xp}} \perp$, and $\Delta_{if} \not\vdash_{\text{xp}} \perp$, then by completeness and observation 8.3.2 there is \mathfrak{M}' such that $\mathfrak{M}' \triangleleft \Delta$ from the latter assumption. Suppose $\mathfrak{M}, (s, h) \triangleleft \Delta$. Then $(s, h) \models \Delta_{if}$ by observation 8.3.2. By compactness of \vdash_{xp} there must be $n_1, \dots, n_k \in \mathbb{N}$ such that $\Delta_{if} \cup \{ \neg \Diamond X^{n_i} \Box XV \mid 1 \leq i \leq k \} \vdash_{\text{xp}} \perp$. But that means $\Delta_{if} \vdash_{\text{xp}} \bigvee_{1 \leq i \leq k} \Diamond X^{n_i} \Box XV$ by classical logic, and by soundness of \vdash_{xp} , $\Delta_{if} \models \bigvee_{1 \leq i \leq k} \Diamond X^{n_i} \Box XV$. What this means is that $\Diamond X^{n_i} \Box XV$ is true at (s, h) for some $1 \leq i \leq k$. Let it be n_i , i.e., $(s, h) \models \Diamond X^{n_i} \Box XV$. Then there is h' with $s \in h'$ and $(s, h') \models X^{n_i} \Box XV$. Let s' be the n_i th h' -successor from s , then $(s', h') \models \Box XV$. But that means $\text{lub}(s') \subseteq \llbracket V \rrbracket$. Since $\mathfrak{M}, (s, h)$ were arbitrary, there is a TE-quandary.

(GE case) $[\Rightarrow]$ Suppose that Δ has a GE quandary. So every model \mathfrak{M} with (s, h) such that $\mathfrak{M}, (s, h) \triangleleft \Delta$ is such that $\forall (s', h') \in \text{AFT}(s, h), \exists \mathbf{A} \subseteq \mathbf{Ag}$, such that $E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$, and there is $\mathfrak{M}' \triangleleft \Delta$. That means $\mathfrak{M}', (s'', h'') \models \Delta_{if}$ for some (s'', h'') . Suppose that $\mathfrak{M}, (s, h) \models \Delta_{if}$. Then there is $\mathbf{A} \subseteq \mathbf{Ag}$ such that $E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$ by assumption for any $(s', h') \in \text{AFT}(s, h)$. That means, since $(s, h) \in \text{AFT}(s, h)$, $(s, h) \models [\mathbf{A} \text{ xstit}] V$, and so $(s, h) \models \bigvee_{\mathbf{A} \in \mathcal{P}(\mathbf{Ag})} [\mathbf{A} \text{ xstit}] V$. Since $\mathfrak{M}, (s, h)$ was arbitrary, $\Delta_{if} \models \bigvee_{\mathbf{A} \in \mathcal{P}(\mathbf{Ag})} [\mathbf{A} \text{ xstit}] V$. By completeness, $\Delta_{if} \vdash_{\text{xp}} \bigvee_{\mathbf{A} \in \mathcal{P}(\mathbf{Ag})} [\mathbf{A} \text{ xstit}] V$.

(GE case) [\Leftarrow] Suppose that $\Delta_{if} \vdash_{xp} \bigvee_{A \in \mathcal{P}(\mathbf{Ag})} [A \text{ xstit}] V$, and $\Delta_{if} \not\vdash_{xp} \perp$, then by soundness $\Delta_{if} \models \bigvee_{A \in \mathcal{P}(\mathbf{Ag})} [A \text{ xstit}] V$, and by completeness and observation 8.3.2 there is \mathfrak{M}' such that $\mathfrak{M}' \triangleleft \Delta$. Suppose $\mathfrak{M}, (s, h) \triangleleft \Delta$, and let $(s', h') \in \text{AFT}(s, h)$. Then $(s', h') \models \Delta_{if}$ by lemma 8.3.3, and so $(s', h') \models \bigvee_{A \in \mathcal{P}(\mathbf{Ag})} [A \text{ xstit}] V$ from our assumption. Thus for some $A \subseteq \mathbf{Ag}$, $E(s', h', A) \subseteq \llbracket V \rrbracket$. So there must be A such that $E(s', h', A) \subseteq \llbracket V \rrbracket$ for any (s', h') since it was arbitrary. And because \mathfrak{M} and (s, h) were arbitrary, GE holds.

The other cases follow similar patterns.

□

It might worry the reader that TE-quandaries are not independent of the conditions 4 and 11 above. Indeed, if Δ has a TE-quandary, then both 4 and 11 will hold of Δ . But we have said that a code can be normatively consistent while 4 and 11 are true of the code. That means what we have to ensure is that neither 4 nor 11 imply that a code has a TE-quandary.

Proposition 8.4.1. *There is a code Δ such that although Δ is such that, $\exists A \subseteq \mathbf{Ag}, \forall \mathfrak{M}, (s, h) \triangleleft \Delta, \exists (s', h') \in \text{AFT}(s, h), E(s', h', A) \subseteq \llbracket V \rrbracket$, Δ does not have a TE-quandary*

Proof. Consider the set $\Delta = \{ \Diamond XV \}$. Now consider the set \mathbf{Ag} . Claim: this set is such that for any model \mathfrak{M} , and $(s, h) \in |\mathfrak{M}|$, if $\mathfrak{M}, (s, h) \triangleleft \Delta$, then there is $(s', h') \in \text{AFT}(s, h)$ such that $E(s', h', \mathbf{Ag})$. Suppose $\mathfrak{M}, (s, h) \triangleleft \Delta$. Then $\mathfrak{M}, (s, h) \models \Diamond XV$, so there is $h' \ni s$ such that $\mathfrak{M}, (s, h') \models XV$. But then $\mathfrak{M}, (\text{lub}(s, h'), h') \models V$, and $E(s, h', \mathbf{Ag}) = \{ \text{lub}(s, h') \}$, so $E(s, h', \mathbf{Ag}) \subseteq \llbracket V \rrbracket$. So 11 holds for Δ .

Now consider the model \mathfrak{M}^* defined as follows: S is the negative integers, and after 0 it is a tree such that each node has 2 successor nodes. We will call them the left and right nodes. H is all of the paths through that tree. Now suppose that $v(V)$ is the set of all right nodes. This model is such that $\mathfrak{M}^*, (0, h) \triangleleft \Delta$. Since at each $\mathfrak{M}^*, (s', h') \in \text{AFT}(0, h)$ its right successor node is a V -state, $\mathfrak{M}^*, (s', h') \models \Diamond XV$. But also, since each left successor state is a non- V -state, $\text{lub}(s') \not\subseteq \llbracket V \rrbracket$. Thus, $\mathfrak{M}^*, (0, h) \triangleleft \Delta$, but there is no $s' \in \text{AFT}(0, h)$ such that $\text{lub}(s') \subseteq \llbracket V \rrbracket$. Therefore Δ does not have a TE-quandary.

□

Corollary 8.4.2. *It is not necessary that if Δ satisfies 4, i.e., $\forall \mathfrak{M}, (s, h) \prec \Delta, \exists \mathbf{A} \subseteq \mathbf{Ag}, \exists (s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$, then Δ has a TE-quandary.*

Proof. Since $\exists \mathbf{A} \subseteq \mathbf{Ag}, \forall \mathfrak{M}, (s, h) \prec \Delta, \exists (s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$ implies $\forall \mathfrak{M}, (s, h) \prec \Delta, \exists \mathbf{A} \subseteq \mathbf{Ag}, \exists (s', h') \in \text{AFT}(s, h), E(s', h', \mathbf{A}) \subseteq \llbracket V \rrbracket$, 11 implies 4. Now, if Δ satisfying 4 implied that it also had a TE-quandary, that would mean that, Δ satisfying 11 implied it having a TE-quandary. But as we have just seen Δ satisfying 11 doesn't imply that Δ has a TE-quandary. Therefore Δ satisfying 4 doesn't imply that Δ has a TE-quandary. \square

Chapter 9

Reflections on the Logic

“Then my sunset?” the little prince reminded him: for he never forgot a question once he had asked it.

“You shall have your sunset. I shall command it. But, according to my science of government, I shall wait until conditions are favourable.”

“When will that be?” inquired the little prince.

“Hum! Hum!” replied the king; and before saying anything else he consulted a bulky almanac. “Hum! Hum! That will be about—about—that will be this evening about twenty minutes to eight. And you will see how well I am obeyed!”

The Little Prince, Antoine de Saint Exupéry

In this chapter we want to reflect on what we have done over the course of this essay. We start by summarizing what has happened, then we look at three things: work by others in this area, how our system stands up in philosophical applications, and whether it is aphiosophical, i.e., philosophically neutral. We end by considering some future directions of research.

In chapter 2 we discussed the logic of institutions in general and found that a logic of institutions is determined by two things: a conception of norms and a conception of in forceness for institutional norms. We argued that it isn’t fruitful to construct an aphiosophical logic of norms because it would be trivial, so we chose one prong of this fork.

In chapter 3 we elaborated on the details of the prong that we chose, which was roughly Searle’s conception of institutions. Searle’s conception of institutions gave us both a conception of norms: all institutional norms are representations of status functions, and a conception of in forceness: declarations by collectively recognized institutional authorities. That provided us with a way to conceive of a consequence relation. The notion of consequence is given by the classical consequences that are preserved by the illocutionary act of status function declaration, e.g., promulgation. In chapter 4 we argued that that consequence relation is captured by Vanderveken’s notion of strong implication.

Part I of the essay provides an answer to question α from the introduction: is a logic of

norms possible? The answer is yes, a logic of norms is possible in so far as our interpretation of Searle is plausible, and assuming Searle's theory is plausible. By a logic of norms we mean *institutional norms*, of course. Part two set about answering question β : What does a logic of norms look like?

We constructed a formal language for representing institutional roles, institutional facts, and a certain account of action via xstit logic. In chapter 6 we formalized the informal notion of norm consequence introduced in chapter 4. We also showed how to interpret the formal language, and how the language of institutions is to be applied to the "real world".

In chapter 7 we proved the formal results that are standard with the introduction of a logic, i.e., soundness and completeness. In chapter 8 we used those formal results to look at an issue in relation to a logic of institutions: normative consistency. We showed how to represent various notions of inconsistency based on the existence of quandaries which reduced to the existence of violations states. Then we demonstrated that these notions in the formal framework can be reduced to just one condition which we named CON_N .

We are now interested in providing some analysis of the formal and philosophical positions defended in this paper. First we compare the formal theory to other work in a similar vein, we then look at what the philosophical range of this theory is, and finally we turn to the range of the formal theory.

9.1 Formal Accounts of Norms

9.1.1 Other Work on Action

In this section we want to briefly mention some work on logics of norms that come from a different conception than ours. That makes a complete comparison rather difficult. From a strictly formal point of view, i.e., the sets of formulas of the consequence relations, there may be some comparisons to make, but that comparison would be almost meaningless since the interpretations of what those formal sentences mean would be so radically different.

Braybrooke, Brown, and Schotch (1995) develop a semantics for norms which we will call TRACK.¹ The kinds of norms that interest the authors are social norms. These norms may be those of etiquette, law or morality. The semantics for norms is based on their definition of norms developed in their chapter 2. Simply put, a norm or rule is a system of imperatives. Also, the authors believe that all norms are of, or are reducible to, one fundamental kind: prohibitions. This is, of course, open to refutation, if there are permissions that cannot be reduced to prohibitions or explained as exceptions to prohibitions. But if we grant their account of norms, all norms can be so reduced without the loss of any logically relevant character. On the TRACK view, norms have three components of logical interest. There is the group of agents for whom the norms are in force (this is the ‘who’ of the norm), the set of circumstances that determine when the norm applies (this is the ‘when’ of the norm), and the actions (or action types) the norm prohibits—what they call the ‘nono’.² Norms have a certain amount of specificity on the TRACK view. The who and when are crucial, but not problematic. How to characterize the actions prohibited is the central project in TRACK.

But as we said, a norm is something radically different from our view, their norms are triples of people, situations and actions. Also, their conception of a norm being in force is rather opaque and a long discussion would be impractical at the moment. Suffice it to say that it is different, but we will make one criticism. The TRACK view takes a stance on what an action is, it is a transition from one state to another. But their characterization of the ‘nono’, i.e., the set of prohibited actions requires a commitment to action individuation which is a notoriously difficult and contentious subject in the philosophy of action. On our view the agentive sentences don’t commit us to a particular view on actions nor act individuation. So our project has the virtue of parsimony over the TRACK theory. However, it is a project of its own to fully compare our two accounts.

Next we consider a system developed by Ågotnes, van der Hoek, Rodríguez-Aguilar, Sierra, and Wooldridge (2007) called normative temporal logic (NTL). This logic allows reasoning

¹That account shares much in common with an account of imperatives in Hamblin (1972).

²This analysis is much like von Wright’s of the norm-kernel of a regulation cf. von Wright (1963, p. 70–1).

about normative systems and adds an explicit temporal dimension. NTL allows reasoning about obligations, and changes in obligations over time.

The formal semantics of NTL is defined on a frame $\langle W, W_0, R \rangle$ where W is a non-empty set called the domain, W_0 a set of start states and R a relation on W . A normative system η is interpreted as a set of forbidden transitions from R such that $R - \eta$ is a serial relation on W . The transitions are supposed to be transitions of one social state to another. Paths are defined as being sequences in W of length ω each step of which is in R . Paths in NTL frames resemble histories in the models of xstit logics. A path conforms to a normative system η when it doesn't share any elements in common with η , that is, a path obeys the rules when it doesn't contain any of the transitions prohibited by η . This system is very closely related to that of Hamblin (1972), but the syntax is very different.

Syntactically speaking, each deontic operator is relativized to a normative system η and must be followed by a temporal operator. Thus obligations and permissions are always relative to a time and a normative system. NTL can be extended to consider unions and intersections of normative systems which the authors claim will give a kind of calculus of normative systems. NTL is also a comparative model of how liberal a system is, which is to say we can compare whether one system has more prohibitions or permissions than another. However, union and intersection are operations on the semantic representations of normative systems. There is also no mention of agents in this system, although these could be added fairly easily. But more importantly, there are no representations of the norms themselves in the object language, only the obligations that arise from the imposition of the norms.³ But our system represents the norms in the object language rather than just what obligations or permissions issue from the norm system being in force. Also, our system makes sense of norm systems being those that follow legal transitions, i.e., an illegal transition is one that ends up in a violation state. However, their conception of what in forceness is or how it is preserved by logical consequence isn't clear. Also, their system doesn't incorporate a syntactic account of codes, the η are simple sets of

³A similar system is found in Segerberg (2009), but that system suffers from similar deficiencies for the intended applications.

transitions. This puts our formulation on a better path. But to be fair we have to mention that their semantics of obligation is in terms of what holds in all permitted continuations. This is at least conceptually distinct from our account of section 6.4, so it makes comparison difficult.

9.1.2 The Input/Output Camp

One of the guiding principles behind input/output logic is that it treats logic as an operation that takes as inputs premises and outputs conclusions. One of the attractive aspects of i/o logic is that it doesn't make any assumptions about the underlying language. The language indifference of i/o logic makes it a continuation of the general theory of consequence operators started—perhaps inadvertently—by Tarski in the 1930s.

The theory uses the consequence operator Cn as a basis for defining the operation, which is simply \mathbb{C}_{CL} . The idea of an i/o logic is that it takes a consequence operator and uses it to extend a collection of conditionals. I/o logic was developed in Makinson and van der Torre (2000), but it was applied most extensively to the theory of norms in the work of Stolpe, e.g., Stolpe (2008b). The attractive thing about the i/o paradigm is that the conditional used to compose conditional norms isn't assumed to have any special logical properties. A norm on this view is made up of a pair of sentences (φ, ψ) where φ describes the condition in which a state of affairs described by ψ is obligatory. The sense in which the norms are arbitrary is that the connection between the condition and the obligation isn't assumed to have any of the logical properties of material or even strict conditionals. Thus a normative code G is a set of pairs (φ, ψ) as just described. But (φ, ψ) isn't really a conditional since it doesn't have any introduction conditions; the connections between conditions and obligations are arbitrary. Given a set of sentences Γ the application of G to Γ , $G(\Gamma)$ is defined as $\{\varphi : \psi \in \Gamma \ \& \ (\psi, \varphi) \in G\}$.

Stolpe's concern is how to extend the explicit norms PN (proper norms) to the set of implicit norms IN by logical consequence. He sees one way of extending PN , called *chaining*, as unproblematic: if (φ, ψ) and (ψ, φ') are both in PN , then (φ, φ') is an implicit norm, and $(\varphi, \varphi') \in PN$. But we have to ask if (φ, ψ) and (φ', ψ') are in PN and $\psi \vdash_{CL} \varphi'$, then should

(φ, ψ') be included in PN ? Stolpe thinks that this latter closure condition, called mediated transitivity, is an acceptable way to extend the PN . But a rule that is unacceptable is weakening the output with arbitrary logical consequences: if (φ, ψ) is in PN , and $\psi \vdash_{CL} \psi'$, then (φ, ψ') is in PN . The reason stems from an argument from Carmo and Jones (2002), and is essentially why Ross's paradox is problematic. If someone is obligated to mail a letter, then they are obligated to mail or burn the letter. If the obligation to mail or burn the letter is on par with the obligation to mail the letter, then failing to do both mail and burn the letter would be to violate the obligation to mail or burn the letter. If the letter isn't mailed, then why not burn it? At least then they would violate fewer of their obligations. So Ross's paradox isn't harmless because it seems to mix up the priority of duties when arbitrary logical consequences of obligations are taken to have the same status as obligations. Obligations not only have fulfillment conditions, they also have violation conditions, and each of those conditions must be accounted for when looking at derived obligations.

Now one might express doubt about counting such derived obligations as on par with explicit obligations at all. But that is a form of scepticism about a logic of norms. But what Ross's paradox can partially show us is that extending PN by arbitrary logical consequence includes too much. So Stolpe's goal is to find some middle ground.

But, Stolpe thinks, that when norms are obeyed, i.e., when (φ, ψ) is a norm, φ is true and whomever this norm is applied to does ψ , the situation that arises from doing ψ also has an impact on what other norms are triggered. What he wants is that

When a norm is used to produce an output, then its consequent—i.e. what the norm decrees to be ideal or obligatory—is dissociated from logically weaker items so that its normative force, so to speak, does not extend to items that are merely true upon fulfillment. Hence all obligations generated are *genuine* in the sense that they correspond to accumulations of *explicitly* given duties pertaining to the circumstances.

Stolpe (2008a, p. 181)

This idea is captured by a particular formal system, given as follows:

Definition 9.1.1. $Der(G)$ is the set of all (φ, ψ) such that either $(\varphi, \psi) \in G; (\top, \top)$ or it is derivable from $G; (\top, \top)^4$ by the rules: such

$$\begin{array}{c} SI \frac{(\varphi', \psi')}{(\varphi'', \psi')} \text{ and } \varphi'' \vdash_{CL} \varphi' \\ Eq \frac{(\varphi', \psi')}{(\varphi', \psi'')} \text{ and } \psi' \equiv \psi'' \\ AND \frac{(\varphi', \psi'), (\varphi', \psi'')}{(\varphi', \psi' \wedge \psi'')} \end{array}$$

and

$$MCT \frac{(\varphi', \psi'), (\varphi' \wedge \psi'', \psi^*)}{(\varphi', \psi^*)} \text{ and } \psi' \vdash_{CL} \psi''$$

The first rule SI is clearly that a normative code is closed when the conditions for the applications of norms are weakened. Eq says that the code is closed by substituting equivalent obligations in norms. AND says that obligations can be combined under the *same* conditions. Finally, MCT , or Mediated Cumulative Transitivity, says that logical consequences of contexts of fulfillment of obligations can be recycled to generate other norms. But it is restricted since there already has to be a norm (θ, ψ^*) in the norms with, at least, $\varphi' \wedge \psi'' \vdash_{CL} \theta$.

Now Stolpe's ideas connect to ours, but his concerns with how norms qua pairs of conditions and obligations can be chained together differ. In our system the situation is more complex since it includes institutional facts, and some of those facts define what the duties are. Now in our system a conditional duty is expressed as $1) \Box(\psi \supset \Box(\neg [\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V))$, although we have not discussed this (we only did it for non-conditional duties). If $\psi \vdash_{xp}^I \theta$, then as long as $\vdash_{Ixp} \theta \subseteq \psi_1 \wedge \dots \wedge \psi_n$ for some subset $\{\psi_1, \dots \psi_n\}$ of the code Ω that 1 is a part of, then $\Box(\psi \supset \Box(\neg [\mathbf{R} \text{ xstit}] \theta \supset [\mathbf{R} \text{ xstit}] V))$ will also be an institutional duty. So this will be weaker than adding arbitrary logical consequences, but it doesn't capture the spirit of what Stolpe is after.

His concern is with the use that gets made of the conditions of norms in triggering the obligations of other norms in the system, i.e., the use made of φ from (φ, ψ) in determining what other norms to include in PN .

⁴ \top is a symbol for the logical constant that is always true.

But is the worry about Ross’s paradox from Carmo and Jones (2002) a worry for our system? The answer is yes, but in our case it isn’t a problem. For Stolpe the problem is that $O(\varphi \vee \psi)$ will be violated when both φ and ψ are false, even when that obligation issues from $O\varphi$, and ψ undermines fulfillment of φ . Ross’s paradox turns on the ability to add anything to an obligation. Not just anything can be added to a duty in our system, it must be something whose content is in the system already. But that doesn’t mean problematic things can’t be added to duties. So once $\Box(\psi \supset \Box(\neg [\mathbf{R} \text{ xstit}] \varphi \supset [\mathbf{R} \text{ xstit}] V))$ is in Ω , $\Box(\psi \supset \Box(\neg [\mathbf{R} \text{ xstit}](\varphi \vee \theta) \supset [\mathbf{R} \text{ xstit}] V))$ will be in Ω as well, and the fulfillment of θ could undermine the fulfillment of φ .

But in our system the reasoning that led to ‘I should burn the letter since I haven’t mailed it, that way I will disobey fewer duties’ doesn’t hold. The approach that we take to practical reasoning *with respect to institutional norms* is that what one should do, *with respect to the institution*, is avoid violations. Now if \mathbf{A} is in a situation s where $\neg \Diamond [\mathbf{A} \text{ xstit}] \varphi$, (i.e., \mathbf{A} can’t mail the letter) then $\cup_{s \ni h'} E(s, h', \mathbf{A}) \subseteq \llbracket V \rrbracket$, so burning the letter won’t ameliorate things for \mathbf{A} . In fact maybe what \mathbf{A} should do is mail it in the next state; if it is burnt, then no mailing can happen in the future. So burning could make things worse since it would lead to the inevitability of another violation state.

For us, not fulfilling a duty is being in a violation state. However, once a violation state has been reached, that doesn’t mean that just anything will make the situation better. An agent being in a violation state is a property of *implementations* of codes, not the codes themselves. That is partly why we see a distinction between the ought of ‘what should I do’ and institutional duty. The distinctions that we make between institutional duty and the ought of practical reason, and between a code and its implementation take away the bite of Ross’s paradox even given Carmo and Jones’s revamping of the worry.

9.1.3 Grossi’s Formalization of Searle

In Davide Grossi’s doctoral thesis *Designing Invisible Handcuffs* (Grossi, 2007), he offers a formal account of Searle’s conception of institutions:

“Institutions” are systems of constitutive rules. Every institutional fact is underlain by a (system of) rule(s) of the form “ X counts-as Y in context C ” (Searle, 1969, pp.51-52).

So clearly his project is very similar to ours, but there are some crucial differences. What we will do first is discuss Grossi’s theory in this section and then compare our project to his in the next. We have taken some inspiration from Grossi’s work, so we are looking to notice some points where the projects differ, particularly with respect to the notion of in forceness and our interpretation of Searle.

Grossi’s goal is to give a semantics for counts-as statements ‘ X counts-as Y in context C ’. A dog counts-as a mammal in every context, for instance. But also in certain places a person’s hands count-as weapons in the context where the person in question is a trained boxer. Such a classification does not hold in every context, e.g., where there are boxers but not weapons laws. Or in the context where a person has hands, but is not a boxer. The count-as statements are expressions of subsumption of the X concept under the Y concept in a context C . But in propositional languages these conceptual classifications must take the form of a relation between sentences. So we must change each count-as statement in to one of the form: ‘ φ -states count-as ψ -states in the context C ’. Although, the examples above are not of that form we can use the usual circumlocutions to capture what needs to be expressed. If we are to think about particular cases, e.g., where φ is ‘Jim is waving his hands in such and such circumstances’ and ψ is ‘Jim is directing traffic’, then $\varphi \supset \psi$ says ‘If Jim is waving his hands in such and such a way, in such and such a context and Jim is a police officer, then Jim is directing traffic’. So our sentence about Jim would represent: ‘Jim waving his hands in such and such a way counts-as Jim directing traffic in the context where he is a police officer and such and such’. But these kinds of count-as sentences could be interpreted as having the form ‘ $\varphi \supset \psi$ ’ is true in C ’, where C is to pick out some context. Sentences of the form ‘ $\varphi \supset \psi$ ’ is true in C ’ are called count as statements, and the $\varphi \supset \psi$ part is called a subsumption statement.

To give a semantics for counts-as sentences we need to interpret (1) contexts, and (2) the subsumption statements. Grossi, building on an idea of Stalnaker, uses possible worlds models $\mathfrak{M} = \langle W, R, v \rangle$. Contexts, in a possible worlds model, are simply a set of possible worlds, i.e., a subset of the domain W . To interpret the subsumption statements Grossi interprets material conditionals as true throughout a context.

So now we have the two components required by 1 and 2 above. A context C is represented by a subset of the domain $W_C \subseteq W$, and the subsumption statements are material conditionals $\varphi \supset \psi$. To say that ‘ φ counts-as ψ in C ’ is to say that $\varphi \supset \psi$ is satisfied throughout W_C in \mathcal{M} . To refer to these contexts we use a set of context variables **Ctx**. To express that a sentence is true throughout a context, i.e., some subset of possible worlds, Grossi introduces unary sentence operators $[C]$ one for each context variable **C** in a set of context variables **Ctx**. We will come back to the context variables after giving the conditions for satisfaction. The language is specified as follows:

$$\varphi := \perp \mid \mathbf{p} \mid \neg\varphi \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \varphi \supset \varphi \mid [C]\varphi$$

So a context frame for a set of context variables **Ctx** is $\mathfrak{F} = \langle W, \{W_C\}_{C \in \text{Ctx}} \rangle$ and a model $\mathfrak{M} = \langle \mathfrak{F}, v \rangle$ where $v : \mathbf{At} \rightarrow \mathcal{P}(W)$. We can then define the truth conditions for \mathfrak{M} and $w \in W$.

- $\mathfrak{M}, w \models \mathbf{p}$ iff $w \in V(\mathbf{p})$;
- $\mathfrak{M}, w \not\models \perp$;
- $\mathfrak{M}, w \models \neg\varphi$ iff $\mathfrak{M}, w \not\models \varphi$;
- $\mathfrak{M}, w \models \varphi \supset \psi$ iff $\mathfrak{M}, w \not\models \varphi$ or $\mathfrak{M}, w \models \psi$;
- $\mathfrak{M}, s \models [C](\varphi)$ iff $\forall x \in W_C, \mathfrak{M}, x \models \varphi$

The **C** are to range over the contexts in a context frame. The operator $[C]$ is used to pick out the subset W_C , i.e., the context that corresponds to **C**. Notice that if a $[C]$ -sentence—a sentence

whose primary connective is $[C]$ —is satisfied *somewhere*, it is so *everywhere*. Thus, $\mathfrak{M}, w \Vdash [C](\varphi \supset \psi)$ implies $\mathfrak{M} \Vdash [C](\varphi \supset \psi)$, i.e., $\forall w \in W, \mathfrak{M}, w \Vdash [C](\varphi \supset \psi)$.

Notice that if $\varphi \supset \psi$ is a classical tautology, then $\varphi \supset \psi$ will be true throughout W_C for each C , so φ states are classified as ψ states throughout W_C , *a fortiori*. Each count as statement is given relative to a model, since it will depend on the model whether $[C](\varphi \supset \psi)$ is true throughout W_C .

Grossi’s account of *genuine* count as statements is captured by *proper classification* count-as sentences. These sentences are those that are true throughout W_C , but false *somewhere* in the model. The intuition is that proper institutional classifications are not metaphysical necessities, but particular to our institutions. To express that relationship requires an operator that quantifies over the whole domain W . This is a unary operator \Box with the following satisfaction condition:

$$\mathfrak{M}, w \Vdash \Box \varphi \iff \forall w' \in W, \mathfrak{M}, w' \Vdash \varphi$$

This is simply the universal modality on the model \mathfrak{M} . So a proper classifying count-as sentence is true in \mathfrak{M} when $[C](\varphi \supset \psi) \wedge \neg \Box(\varphi \supset \psi)$ is true in the model. We denote the proper count-as conditional as $\varphi \Rightarrow^C \psi$.

According to Grossi, however, the “real” constitutive norms are not just proper classifications, but proper classifications that are part of a set of sentences. This is because constitutive norms must be considered relative to a system or code of norms. Moreover this set of sentences must *define* a context. A set of sentences Γ defines a context W_C in a model \mathfrak{M} , when $\mathfrak{M}, w \Vdash \Gamma$ iff $w \in W_C$. So for any $w' \notin W_C$, $\mathfrak{M}, w' \not\Vdash \Gamma$, i.e., at least one of the members of Γ fails at each world outside of W_C . The rationale behind a set defining a context is because for a count-as statement $[C](\varphi \supset \psi)$ to be a constitutive count-as, it must be a member of some code. The way to guarantee this semantically is for W_C to be defined by the code.

Grossi then shows how to define regulative norms in terms of count as statements. Using Anderson’s (1958) reduction of deontic logic to modal logic and an adaptation of it from Meyer (1988), Grossi reduces regulative norms to constitutive norms. We will recall the description of the reduction from section 5.1.3 and the special symbol V that is used to denote the set of

violation or “bad” states, in some sense. In regular modal logic under Anderson’s reduction, φ is forbidden at $w \in W$ iff $\Box(\varphi \supset V)$ is true at w . Obligation is then represented as $\Box(\neg\varphi \supset V)$. This is a forbidden-not-to-*be* version of obligation. It states that all $\neg\varphi$ states are violation states. So Grossi contextualizes this definition: φ is forbidden relative to the code Ω in \mathfrak{M} iff $\varphi \Rightarrow^C V$ is true in \mathfrak{M} and $\llbracket \Omega \rrbracket = W_C$.

In Grossi’s formalism we can then express ‘ φ counts-as a violation in the context picked out by C ’ as $[C](\varphi \supset V)$. Regulative rules—at least on Searle’s account—regulate existing states or actions:

Where the rule is purely regulative, behaviour which is in accordance with the rule could be given the same description or specification (the same answer to the question “What did he do?”) whether or not the rule existed, provided the description or specification makes no explicit reference to the rule. But where the rule (or system of rules) is constitutive, behaviour which is in accordance with the rule can receive specifications or descriptions which it could not receive if the rule did not exist (Searle, 1969, p. 35).

The regulative norms say what can and cannot be done in respect to the normative system or code, but the actions or states of affairs could still have the same description without the count-as statements contained in a system of rules. Walking on the grass can be described as walking on the grass even in the absence of any rules that prohibit walking on the grass. More formally, a φ -state is a φ -state even after a norm is introduced that classifies φ -states as violations. But classifying φ -states as violation could not exist without some constitutive norms. But then regulative norms can be represented as constitutive norms by providing classifications according to whether actions or states are in violation of the norms. So norms on Grossi’s account define the logical space in which they hold, i.e., constitutive norms define the context in which they hold. However, it is lacking an important ingredient thus far for norms: an account of in forceness.

9.1.4 Some Interpretation of Grossi's Work

Grossi contends that we can treat, at least extensionally, all norms as constitutive count-as *statements*. So although it may not be that all norms *really are* constitutive count-as statements, the logic of count-as statements is not effected when they are analyzed as such. But if this were a definition of norms, then norms are really linguistic entities, which is Searle's view. So Grossi is taking a more philosophically neutral position on the matter.

The semantics of norms is given as a special kind of proposition, propositions about what count-as what and what counts-as a violation. One of the novel results of Grossi's thesis is that regulative rules can be treated as constitutive rules. But how can Grossi's formalism be interpreted with respect to the notion of a norm being in force? And can Grossi's formalism provide a logic of norms? To get clear on these questions the former will be discussed first.

Grossi says that

...when statements "[φ] counts-as [ψ] in the context [C] of normative system Γ " are read as constitutive rules, what is meant is that the classification of [φ] under [ψ] is considered to be an explicit promulgation of the normative system Γ defining [C]. (Grossi, 2007, p. 81)

Grossi is saying that we are to interpret the truth of constitutive count-as statements as the constitutive norm being in force. The relation of a context model to the actual world is not clear in Grossi, we must do our best to give a reading of it. Suppose that a local legislator in some other possible world w has promulgated all of the count-as statements in Γ . Presumably these count-as statements would now be in force. But that would mean that the sentences in this set are all true in w . That means that there is a context which they define, but the actual world does not have to be a member of the worlds in the context to make all of the count-as statements true since they are contextualized. Indeed, $\varphi \Rightarrow^C \psi$ will be true at every world if it is true anywhere. By definition of the set Γ being a set of constitutive count-as statements, for some $\varphi \Rightarrow^C \psi \in \Gamma$, $\varphi \supset \psi$ must be false at the actual world since the actual world isn't in the

context defined by Γ . Intuitively, then, at the actual world we don't count φ -states as ψ -states. But $\varphi \Rightarrow^C \psi$ is true at the actual world anyway, thus it is in force at the actual world. And that is counterintuitive.

This presents a problem between the relation of norms that are in force and the truths of the actual world. What Grossi was after was a model of how norms are imposed on the world, but this model doesn't seem to capture this view. Related to this problem, not every context model contains every possible context. So it may be possible for a set of sentences Γ to be true throughout a subset of the domain of some context model, and false in the right places, but the set of sentences would not be in force because the subset of the domain of the model that Γ defines is not one of the contexts of the frame on which the model is based. I.e., Γ might be a set of uncontextualized sentences, might define a set of worlds in a model \mathfrak{M} , but that set of worlds $\llbracket \Gamma \rrbracket$ isn't one of the W_C in the model. So we can't interpret a code as in force in just any model, it has to have the right kind of contexts so that it can interpret Γ as in force.

Also, since Grossi assumes that a set of constitutive count-as statements is the set of explicitly promulgated constitutive count as statements it cannot be deductively closed, otherwise there would have to be implicit norms as well. But maybe that can be put aside. More importantly, as per our discussion in section 3.2 the context should really just be W , not a proper subset. This is because W is supposed to represent all of the ways the world might be, and when a code is in force, it is imposed on all of the ways the world could be. That is at odds with the interpretation of proper classifications. A proper classification must be true throughout the context that is defined by its set of constitutive norms, but it must also be false *somewhere*. When W is the context, then those two conditions are impossible to satisfy.

So let's suppose, for the sake of argument, that proper count-as statements that follow from sets of proper count-as statements that have been explicitly promulgated are also norms in force, but as long as the set of statements define a context. A proper count-as might just happen to be true throughout a context defined by Γ , but since it is not in Γ it fails to be constitutive. Thus, if only constitutive count-as statements are to be considered in force, this account gives an austere

logic of norms, i.e., just the explicitly promulgated count as statements. So the consequence relation would be trivial. I.e., $\mathbb{C}(\Gamma) = \Gamma$.

Grossi's theory does have some virtues. There are no new semantic entities entering into Grossi's formalism. There are contexts, and count-as sentences that are interpreted as conditionals, but no new semantic value for norm-formulations. So it has a parsimonious effect on the semantics of norms. Also, if we (contra Grossi) interpret proper count-as statements as norms also in force, that is, the explicit count-as that are promulgated and what follow from them are in force, then we might get a non-trivial logic of norms. Let's look at this notion precisely.

First, we should notice that the only institutional facts in Grossi's system are count as statements. But generally speaking there are other institutional facts that might not have the simple form of a subsumption statement: free standing Y terms (recall section 3.2). Our system allows for institutional facts of general forms. So let's suppose further that a general institutional fact φ can be included in a code. Also we have to realize that a code shouldn't be represented in terms of a set of contextualized formulas, i.e., formulas of the form $[\mathbf{C}]\varphi$ for φ a boolean formula. It should be just a set of purely Boolean formulas, the contextualized version is used to represent whether a norm is in force in a model. Also, if a code was represented by a set of contextualized formulas, then it couldn't be false anywhere in a model. Now we can offer a definition for normative consequence for Grossi's system: \models_N^G ('G' for 'Grossi').

A code of pure Boolean formulas Γ can normatively entail another pure Boolean formula ψ : $\Gamma \models_N^G \psi$ iff in all models in which the context determined by Γ , also makes ψ true, but ψ is false somewhere, and each member of Γ is false somewhere in the model. This is in line with Grossi's interpretation of proper count as statements: they are true throughout a context, but false somewhere. Or, in any model in which Γ determines the context \mathbf{C} , and no $w \in W \setminus W_{\mathbf{C}}$ satisfies all of Γ , each $w \in W_{\mathbf{C}}$ makes ψ true and ψ is false somewhere. We can put this more formally as:

Definition 9.1.2. $\Gamma \models_N^G \psi$, iff for every \mathfrak{M} , s.t.

1. if $\forall x \in W (x \in W_{\mathbf{C}} \text{ iff } \mathfrak{M}, x \Vdash \Gamma)$, and

2. for each $\gamma \in \Gamma$ there is $w \in W$ such that $\mathfrak{M}, w \not\models \gamma$, then
3. $W_C \subseteq \llbracket \psi \rrbracket_{\mathfrak{M}}$, and
4. $\exists w' \in W$ such that $\mathfrak{M}, w' \not\models \psi$.

So we can now see that our interpretation of Grossi's system makes his rather different from our own. First, it is impossible to have any tautology be an institutional fact since tautologies are never false at any world in any model. But what is more interesting is that we can take any set of pure boolean formulas Γ and any classical consequence of Γ , ψ , that is neither a tautology nor logically stronger than any member of Γ , and show that $\Gamma \not\models_N^G \psi$.

Proposition 9.1.1. *If Γ is consistent, and $\Gamma \models_{CL} \psi$, where $\not\models_{CL} \psi \supset \gamma$ for each $\gamma \in \Gamma$, and $\not\models_{CL} \psi$, then $\Gamma \not\models_N^G \psi$.*

Proof. We will build a model \mathfrak{M} out of classically maximally consistent sets. Let W_C be the set of \models_{CL} -maximally consistent sets Δ such that $\Gamma \subseteq \Delta$. Note that there are such maxi sets since Γ is consistent. Since $\Gamma \models_{CL} \psi$, $W_C \subseteq \llbracket \psi \rrbracket$, where $\llbracket \psi \rrbracket$ is the set of maximally consistent sets that contain ψ .

For each $\gamma \in \Gamma$, $\{\neg\gamma, \psi\}$ is classically consistent since none of the members of Γ is entailed by ψ . So it can be extended to a maxi set, Δ_γ .

Define $W = W_C \cup \{\Delta_\gamma : \gamma \in \Gamma\}$, and $v(\mathbf{p}) = \{\Delta \in W : \mathbf{p} \in \Delta\}$. But then we would have a model where 1, 2, and 3 from definition 9.1.2 are met, but 4 isn't. Therefore $\Gamma \not\models_N^G \psi$. □

So Grossi's norm consequence, or our reconstruction of it, is weaker than ours. There are no norm consequences that are strictly weaker than the individual norms in a code on Grossi's system.

To sum up, there are some very good ideas in Grossi's work. We can account for constitutive norms, and capture regulative norms as special kinds of constitutive norms. However, there are some failures for the application that we have in mind. There are some complications with

how Grossi's models relate to "the world". The missing element is a coherent story about how norms relate to the world, i.e., when the norms are in force. What is missing is a proper representation of the actual world in the models. His models should be thought of as hovering over the world, defining the logical space of how things look relative to different codes. But then interpreting how that relates to the world is left unresolved by Grossi. Also, from our philosophical position Grossi's interpretation of norm consequence is too demanding. As we can see from the result above. The problem for norm consequence comes from the need for the content of an institutional fact to be false somewhere in the model. On our view, what is in force at a world is a relation that concerns what is true at that world. Whether something is in force depends on whether it was imposed on that world, not whether it isn't true somewhere else.

9.2 Philosophical Evaluation

In this section we want to answer two questions of philosophical interest in relation to this project:

1. How general is the logic of institutions and does the system generalize to normative systems broadly construed?
2. We have ignored defeasibility: why and is that right?

9.2.1 Applicability to other Institutions and Systems of Norms

How general is this account of institutions, and does it generalize to normative systems broadly construed? For other institutions there are a few cases to consider. First are cases of institutions that derive their existence *from* law. These institutions are simply another part of law. But there are institutions that don't, e.g., universities. Of course, it will take empirical investigation to make sure that Searle's theory applies to all things we might consider social institutions. But that isn't our project; our project follows a methodology where we take a particular philosophical position and see where it can go.

On the other hand there are other types of normative systems. The norms of morality and etiquette may not be institutions like the law. Of course, whether they are depends on the philosophy of those normative systems. Etiquette may rely on common recognition, but it doesn't come about through declarations. Indeed, it has been argued that institutions don't work the way Searle thinks they do, but we aren't considering that here. The basic idea for the logic would still work in the same way: the logic for the conditions of in forceness is the foundation of the logic of norms. What we offer here is one account of what that logic is and, more importantly, why it is that way. If a notion of entailment could be made sense of for the relation of common recognition, then that could be a notion of normative consequence for etiquette.

The logic developed here is limited in its scope of application. It can only be applied to normative systems that function in the way Searle's account of institutions works. But it might not work for all accounts of how institutions work. And how widely Searle's system applies is beyond the scope of the current project.

9.2.2 Norms: Interpretation and Defeasibility

Interpretation of norm formulations is always a bit of a problem, the legal profession is built around it. Part of the problem stems from the ambiguity that exists in any natural language. Of course logical formalism is supposed to remove any ambiguity and only express "what is really meant". But that may be problematic with norms. Some theorists (e.g., Hart) assert that norms are opened textured, i.e., there are certain cases that will defy classification. Any concept is always too wide and too narrow, thus human interpretation must always play a part in what a norm "really means". So to ignore this would be in error. In formalization there must always be some idealization of natural language and human reasoning. The approach we have taken is to say the logic developed applies to the norms *after* the interpretation is done, i.e., it applies to a particular master system rather than to a master book. The idealization that we have made is that we can pick a particular master system at all.

A deontic logic, to be an accurate representation of actual normative reasoning, must include some version of defeasible obligation. The reason can be illustrated by the following story. Suppose that Tammy is a student at the, fictional, university U of C, and there is a norm in the code of conduct of U of C that says that each student has a duty to go to every class that she/he has registered for. So Tammy has a duty to go to every class she has registered for. However, if there were a death in Tammy's family, and the funeral and a class coincided, it would not follow, intuitively, that she has a duty, in this instance, to go to that class. The duty to pay respect at the funeral of a family member can override the institutional duty to go to class, and thus absolves Tammy of her duty to attend class in this circumstance.⁵

So we have to be able to account for this kind of situation in the formalism. On the one hand we do recognize this kind of situation, but we dismiss it. We always interpret a master system as what is being represented by the formulas. This means that we are looking at norms that have all exceptions explicitly represented. At the real University of Calgary, there are norms that allow students to defer exams because of a death in the family. This is represented as an explicit exception to the norms. The point is that nothing can be implicit in the formulas.

The need for defeasibility arises when trying to capture or represent common sense reasoning. Common sense reasoning is always done from an epistemically limited position. But we have assumed that there isn't such a limited position. Again, we could add this kind of defeasibility to better represent actual common sense reasoning, but we haven't done that.⁶ Our goal was to capture norm consequence, and from an objective standpoint where we only consider the norms in force of an institution, with all exceptions represented, we only look at what institutional facts follow, duties among them.

Other authors have argued that defeasibility isn't necessary in the representation of deontic reasoning, cf. Hurtig (2007). But we take a weaker stance since we just argue that for institutional norm consequence, defeasibility is irrelevant. Also, we are not saying that what the

⁵This phenomena is connected to the idea of a *prima facie* duty from Ross (1930).

⁶Systems of these kinds exist in the work of Prakken (1997), Hage (1997), Governatori and Rotolo (2004) and Horty (1994).

norms of the institution say ought to be done *is* what ought to be done. That is a further question about the priority of institutional duties, e.g., legal duties, over other sorts of duties. We are simply asking what institutional facts follow from the the set of explicit institutional facts in force for that institution. What duties there are in general is another question, and what duties an agent has in a particular situation is a question about the relationship between the particular implementation of that institution, and what other facts (brute and institutional) there are in that situation. That latter question is one of practical reasoning which we have left largely undeveloped.

So now that we have discussed some of the philosophical issues, we can look at some questions to do with the formalism.

9.3 Logical Evaluation

In this section we look at how aphiosophical our formal system is, and we summarize the future directions of this research based on the current project.

9.3.1 Future Directions

Our future work is to focus on formalizing consequence relations that correspond to different conceptions of in forceness, further explore the notion of normative consistency developed above, and properly delve into the semantics of ‘ought’ within this framework. In relation to other conceptions of in forceness, we want to explore how to reason about institutions conceived in a game-theoretic manner. This is to look at institutions from the perspective of economics.

The notion of normative inconsistency developed above is rather minimal. We want to look closer and compare our notion of inconsistency to those discussed by Hamblin (1972). At the moment, our notion of consistency is similar to his notion of “minimal consistency”. For the semantics of ‘ought’, we are interested in deriving an account of ought that is axiological, but doesn’t introduce an unanalysed notion of value. We want to derive the account of value from the existence and non-existence of violations in a history.

We would also like to apply this formalism in relation to designing codes with certain goals. We want to look at properties of models as goals, i.e., the intent behind making a certain norm. Generally these are called *policies*. Using our formalism we would like to find ways to construct codes so that models that realize those codes will have those properties.

9.3.2 Aphilosophicality

The logic developed in chapter 6 was based on a philosophical position given in section 4.4. Our system bases norm entailment on strong implication, and it could undermine much of this project, if it turned out to be a faulty foundation for the work. The reason that many formal logicians try to make their work aphilosophical is to insulate the formal theories developed from shifting philosophical grounding. We would like to take this section to point out where our theory is fixed to the philosophical foundations developed in part I.

The primary way that our account is connected to the Searlean account of institutions is via the conception of norms. It is only with a conception of norms like Searle's that something like Anderson's reduction of institutional duty can be defended. Of course, we are not saying that Searle's theory is the only way to defend this conception of norms. Any Searle-independent philosophical theory using this conception of norms would permit our logic. What this view couldn't withstand is a theory that said that there must be some basic and irreducible account of obligation used to express norms for institutional codes. That would undermine a great deal of the work in part 2. The other issue to worry about is the notion of norm consequence we used, and related to that the notion of in forceness.

As long as the extension of the norm consequence relation remains untouched, the logic will work. However, if a new notion of in forceness is used, and so the extension of norm consequence is changed, that will cause some disruption. As long as the relata of norm consequence are the same, but just a different set of pairs are related, we can mediate the damage caused.

The work in chapter 8 will remain untouched pretty much under any change to the extension to norm consequence. For the results concerning consistency, \vdash_N isn't used, it is all defined

relative to a \vdash_{xp}^{Ω} consequence relation since it depends on the idea of a code being followable relative to an implementation, and that has little to do with the norm consequence, i.e., \vdash_N . So any change to norm consequence will not affect the idea norm consistency.

Although our methodology is to take the fork in the road, the particular extension of the fork doesn't play a huge role in the applications of the technical work. But that is the best that we could hope for in this kind of project.

9.4 Conclusion

We've come to the end of the story. In the end we have answered our two questions: α) Is a logic of institutional norms possible?, and β) Given that a logic of institutional norms is possible, what does it look like? Our answer to α came from the discussion in part I, and the answer to β came from chapter 6, particularly section 6.5.

The answer to α was given by taking a substantial theory of institutions, then showing how to conceive of a logic for that theory. Again, we return to the methodology briefly outlined in the introduction. We don't want to say: yes, this is *THE* logic of norms! Nor even *THE* logic of institutional norms. That isn't how we are thinking about the project of philosophical logic. However, we do want to say that this is the logic of institutional norms *on Searle's conception*. Nevertheless, as long as Searle's theory has a chance at being correct, our discussion shows that a logic of institutional norms is possible, epistemically speaking.

For question β , we get a sense of what the consequence relation is like for norms on Searle's view from chapter 4. But we get a *formal* characterization of it from chapter 6. Thus we have a mathematically precise way to investigate what the logic looks like. Although we didn't look closely at the consequence relation, our formal characterization gives us the ability to formulate precise questions about the relation and evaluate arguments about what follows from the norms of an institution. As an added bonus we have a few other interesting fallouts from the discussion.

The highlights of this essay have been the representation of the distinction between institutions and the world and the discussion about normative consistency. The distinction between

the institutional facts and the brute facts via a code Ω and its implementation $\mathfrak{I}(\Omega)$ is an important distinction that is often left unrepresented in logical systems. Normative consistency hasn't been given much direct attention in recent years, but is becoming more important with the rise in computer science of artificial institutions. So this will be a fruitful line for future inquiry.

A final comment to bring things to a close. The quotation from *The Little Prince* above shows us how not to run an institution, but more important is the expression: it's funny because it's true. Indeed, it captures the idea that there is a strong distinction between brute facts and institutional facts. But the only facts that we can have the kind of control over via our ability to represent the world are the institutional facts. As Searle's theory is wont to show, our command over the institutional world is absolute, but our command over the brute world is by luck. The best philosophical theories are justifiable by a joke.

Bibliography

- Ågotnes, T., W. van der Hoek, J. A. Rodríguez-Aguilar, C. Sierra, and M. Wooldridge (2007). On the logic of normative systems. In G. Boella, L. W. N. van der Torre, and H. Verhagen (Eds.), *Normative Multi-agent Systems*, Volume 07122 of *Dagstuhl Seminar Proceedings*. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.
- Alchourrón, C. E. (1996). On law and logic. *Ratio Juris* 9(4), 331–48.
- Alchourrón, C. E. and E. Bulygin (1981). The expressive conception of norms. In R. Hilpinen (Ed.), *New Studies In Deontic Logic: Norms, Actions, and the Foundations of Ethics*, Volume 152 of *Studies in Epistemology, Logic, Methodology, and the Philosophy of Science*, Chapter 4, pp. 95–124. Dordrecht: D. Reidel Publishing Company.
- Alchourrón, C. E., P. Gärdenfors, and D. Makinson (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50(2), 510–530.
- Anderson, A. R. (1958, January). A reduction of deontic logic to alethic modal logic. *Mind* 67(265), 100–3.
- Anderson, A. R. (1967, Dec). Some nasty problems in the formal logic of ethics. *Noûs* 1(4), 345–60.
- Anscombe, G. E. M. (1958). On brute facts. *Analysis* 18, 69–72.
- Austin, J. L. (1962). *How to do Things With Words* (2 ed.). Cambridge, MA: Harvard University Press.
- Bartha, P. (1993). Conditional obligation, deontic paradoxes, and the logic of agency. *Ann. Math. Artif. Intell.* 9(1-2), 1–23.
- Belnap, N. and M. Perloff (1992). The way of the agent. *Studia Logica* 51(3/4), 463–484.

- Bicchieri, C. (1997). *Rationality and Coordination*. Cambridge Studies in Probability, Induction and Decision Theory. Cambridge, UK: Cambridge University Press.
- Binmore, K. (2010). Game theory and institutions. *Journal of Comparative Economics* 38(3), 245 – 252. [jce:title;Symposium: The Dynamics of Institutions;/ce:title;.](#)
- Braybrooke, D., B. Brown, and P. Schotch (1995). *Logic on the Track of Social Change*. Clarendon Library of Logic and Philosophy. Oxford: Clarendon Press.
- Broersen, J. and J.-J. C. Meyer (2011). A stit logical study into choice, failure and free will action. In J.-J. C. Meyer (Ed.), *Epistemology, Context and Formalism*. Springer. Forthcoming.
- Caleiro, C., C. Sernadas, and A. Sernadas (1999). Parameterisation of logics. In J. Fiadeiro (Ed.), *Recent Trends in Algebraic Development Techniques*, Volume 1589 of *Lecture Notes in Computer Science*, pp. 48–62. Springer.
- Carmo, J. and A. Jones (2002). *Deontic Logic and Contrary-to-Duties* (2 ed.), Volume 8 of *Handbook of Philosophical Logic*, pp. 265–343. Dordrecht: Kluwer.
- Castañeda, H.-N. (1975). *Thinking and Doing: The Philosophical Foundations of Institutions*, Volume 7 of *Philosophical Studies Series in Philosophy*. Dordrecht, Holland: D. Reidel.
- Chellas, B. F. (1969). *The Logical Form of Imperatives*. Phd, Stanford University, Stanford. Published by Perry lane press.
- Cherniak, C. (1981a). Feasible inferences. *Philosophy of Science* 48(2), 248–68.
- Cherniak, C. (1981b). Minimal rationality. *Mind* 90(358), 161–83.
- Cherniak, C. (1986). *Minimal Rationality*. Cambridge, Mass.: MIT Press.
- Church, A. (2009). Alonzo church: Referee reports on fitch’s “a definition of value”. In J. Salreno (Ed.), *New Essays on the Knowability Paradox*, Chapter 1, pp. 13–18. Oxford University Press.

- Ciuni, R. and A. Zanardo (2010). Completeness of a branching-time logic with possible choices. *Studia Logica* 96(3), 393–420.
- Davidson, D. (1970). How is weakness of the will possible? In J. Feinberg (Ed.), *Moral Concepts*, pp. 93–113. New York: Oxford University Press.
- Dretske, F. (1981, May). *Knowledge and the Flow of Information*. The MIT Press.
- Dummett, M. (1991). *The Logical Basis of Metaphysics*. Harvard University Press.
- Dworkin, R. (1978). *Taking Rights Seriously* (2nd ed.). Cambridge, Mass: Harvard University Press.
- Encyclopdia Britannica Online, s. v. “baptism”. Online. accessed June 29, 2012, <http://www.britannica.com/EBchecked/topic/52311/Baptism>.
- Field, H. (2009, June). Pluralism in logic. *Review of Symbolic Logic* 2(2), 342–59.
- Fine, K. and G. Schurz (1996). Transfer theorems for multimodal logics. In J. Copeland (Ed.), *Logic and Reality: Essays on the Legacy of Arthur Prior*, pp. 169–213. Oxford: Oxford University Press.
- Fitch, F. B. (1967). A revision of holfrld’s theory of legal concepts. *Logique et Analyse* 10(3), 269–76.
- Goldenrowley (May 5, 2009). Hyletics. Wiktionary. <http://en.wiktionary.org/wiki/hyletics> May 2, 2011.
- Governatori, G. and A. Rotolo (2004). Defeasible logic: Agency, intention and obligation. In A. Lomuscio and D. Nute (Eds.), *DEON*, Volume 3065 of *Lecture Notes in Computer Science*, pp. 114–128. Springer.

- Grossi, D. (2007, September). *Designing Invisible Handcuffs*. Ph. D. thesis, University of Utrecht. SIKS Dissertation Series No. 2007-16.
- Hage, J. C. (1997). *Reasoning with Rules: An Essay on Legal Reasoning and Its Underlying Logic*, Volume 27 of *Law and Philosophy Library*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Hamblin, C. L. (1972). Quandries and the logic of rules. *Journal of Philosophical Logic* 1(1), 74–85.
- Hamblin, C. L. (1987). *Imperatives*. Oxford, UK: Basil Blackwell.
- Hansson, S. O. (2001). *The Structure of Values and Norms*. Cambridge Studies in Probability, Induction and Decision Theory. Cambridge: Cambridge University Press.
- Hare, R. M. (1952). *The Language of Morals*. Oxford University Press.
- Harman, G. (2002). Internal critique: A logic is not a theory of reasoning and a theory of reasoning is not a logic. In D. M. Gabbay, R. Johnson, H. Ohlbach, and J. Woods (Eds.), *Handbook of the Logic of Argument and Inference: The Turn Towards the Practical*, Volume 1 of *Studies in Logic and Practical Reasoning*, pp. 171–86. Amsterdam: Elsevier Science B.V.
- Herzig, A. and F. Schwarzentruher (2008). Properties of logics of individual and group agency. In *Advances in Modal Logic*, Volume 7, pp. 133–49.
- Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Contemporary Philosophy. Ithaca and London: Cornell University Press.
- Hofstadter, A. and J. C. C. McKinsey (1939). On the logic of imperatives. *Philosophy of Science* 6(4), 446–457.
- Holfeld, W. N. (1920). *Fundamentals of Legal Concepts*. New Haven, Conn: Yale University Press. Reprinted 1964.

- Horty, J. (1994). Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic* 23, 35–65.
- Horty, J. (2001). *Agency and Deontic Logic*. Oxford, UK: Oxford University Press.
- Hurtig, K. (2007). On prima facie obligations and nonmonotonicity. *Journal of Philosophical Logic* 36(5), 599–604.
- Jørgensen, J. (1937). Imperatives and logic. *Erkenntnis* 7, 288–96.
- Kanger, S. and H. Kanger (1966). Rights and parliamentarism. *Theoria* 32, 85–115.
- Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard University Press.
- Lorini, E., D. Longin, B. Gaudou, and A. Herzig (2009). The logic of acceptance: Grounding institutions on agents' attitudes. *J. Log. Comput.* 19(6), 901–940.
- Makinson, D. (1986). On the formal representation of rights relations. *Journal of Philosophical Logic* 15, 403–25.
- Makinson, D. and L. van der Torre (2000). Input-output logics. *Journal of Philosophical Logic* 29, 383–408.
- Mally, E. (1926). *Grundgesetze des Sollens. Elemente der Logik des Willens*. Graz: Leuschner & Leubensky. Reprinted in *Ernst Mally, Logische Schriften: Großes Logikfragment, Grundgesetze des Sollens*, Karl Wolf and Paul Weingartner (eds.), Dordrecht: D. Reidel, 1971, 227–324.
- Marcus, R. B. (1980). Moral dilemmas and consistency. *The Journal of Philosophy* 77(3), 121–36.
- Meijers, A. (2007). Collective speech acts. In S. Tsohatzidis (Ed.), *Intentional Acts and Institutional Facts*, pp. 93–110. ?: Springer.

- Meyer, J.-J. C. (1988). A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic* 29(1), 109–136.
- of Statutes, R. (2009, August). Maine legislative drafting manual. Online.
- Ogilvy, C. S. (1956). *Through the Mathscope*. New York: Oxford University Press.
- Pauly, M. (2001). *Logics for Social Software*. Ph. D. thesis, University of Amsterdam, Amsterdam. Dissertation Series DS-2001-10.
- Peczenik, A. (2009). *On Law and Reason*, Volume 8 of *Law and Philosophy Library*. Houten: Springer.
- Pörn, I. (1977). *Action Theory and Social Science: Some Formal Models*, Volume 120 of *Synthese Library*. Dordrecht, Holland: D. Reidel Publishing Company.
- Postema, G. (1982, Jan). Coordination and convention at the foundations of law. *The Journal of Legal Studies* 11(1), 165–203.
- Postema, G. (1994, Aug). Implicit law. *Law and Philosophy* 13(3), 361–87.
- Prakken, H. (1997). *Logical Tools for Modelling Legal Argument: A study of Defeasible Reasoning in Law*, Volume 32 of *Law and Philosophy Library*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Reynolds, M. (2002). Axioms for branching time. *J. Log. Comput.* 12(4), 679–697.
- Ricoeur, P. (1988). *Time and Narrative*, Volume 3. Chicago: University of Chicago Press.
- Ross, A. (1944, Jan). Imperatives and logic. *Theoria* 11(1), 30–46.
- Ross, A. (1958). *On Law and Justice*. London: Stevens.
- Ross, W. D. (1930). *The Right and the Good*. Oxford, UK: Oxford University Press.
- Sagan, C. (1980). *Cosmos*. New York: Random House.

- Sauro, L., J. Gerbrandy, W. van der Hoek, and M. Wooldridge (2006). Reasoning about action and cooperation. In *AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, New York, NY, USA, pp. 185–192. ACM.
- Sayre-McCord, G. (1986, June). Deontic logic and the priority of moral theory. *Noûs* 20(2), 179–97.
- Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge UK: Cambridge University Press.
- Searle, J. (1995). *The Construction of Social Reality*. The Free Press.
- Searle, J. (2005). What is an institution? *Journal of Institutional Economics* 1(1), 1–22.
- Searle, J. (2010). *Making the Social world*. Oxford, UK: Oxford University Press.
- Searle, J. R. and D. Vanderveken (1985, July). *Foundations of Illocutionary Logic*. Cambridge University Press.
- Segeberg, K. (2009, December). Blueprint for a dynamic deontic logic. *Journal of Applied Logic* 7(4), 388–402.
- Stolpe, A. (2008a). Normative consequence: The problem of keeping it whilst giving it up. In R. van der Meyden and L. van der Torre (Eds.), *DEON*, Volume 5076 of *Lecture Notes in Computer Science*, pp. 174–188. Springer.
- Stolpe, A. (2008b). *Norms and Norm-System Dynamics*. Department of philosophy, University of Bergen, Norway.
- Thomason, R. (1984). Combinations of tense and modality. In D. Gabbay (Ed.), *Handbook of Philosophical Logic* (2 ed.), Volume 7, pp. 205–234. Dordrecht, The Netherlands: Kluwer.
- Tuomela, R. and W. Balzer (1999). Collective acceptance and collective social notions. *Synthese* 117, 175205.

- van Benthem, J. (1983). Logical semantics as an empirical science. *Studia Logica* 42(2–3), 299–313.
- van Fraassen, B. C. (1973). Values and the hart’s command. *Journal of Philosophical Logic* 70(1), 5–19.
- Vanderveken, D. (1990). *Meaning and Speech Acts: Principles of language use*, Volume 1. Cambridge, UK: Cambridge University Press.
- Vanderveken, D. (1991). *Meaning and Speech Acts: Formal Semantics of success and satisfaction*, Volume 2. Cambridge, UK: Cambridge University Press.
- Vanderveken, D. and M. Nowak (1995, May). A complete minimal logic of the propositional contents of thought. *Studia Logica* 54(3), 391–410.
- von Wright, G. H. (1951). Deontic logic. *Mind* 60(237), 1–15.
- von Wright, G. H. (1963). *Norm and Action*. London: Routledge.
- von Wright, G. H. (1991, December). Is there a logic of norms? *Ratio Juris* 4(3), 265–83.
- Vranas, P. (2008). New foundations for imperative logic i: Logical connectives, consistency, and quantifiers. *Noûs* 42, 529–72.
- Zanardo, A. (1996). Branching-time logic with quantification over branches: The point of view of modal logic. *The Journal of Symbolic Logic* 61(1), 1–39.
- Zanardo, A. and J. Carmo (1993). Ockhamist computational logic: Past-sensitive necessitation in ctl^* . *J Logic Computation* 3(3), 249–68.

Appendix A

Languages, Models, and Logics

A.1 Languages

$\mathcal{L}_{\text{xstit}}$:

$$\varphi := \mathbf{p} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [\mathbf{A} \text{ xstit}] \varphi \mid X\varphi$$

Where $\mathbf{p} \in \mathbf{At}_B$.

\mathcal{L} :

$$\varphi := \mathbf{p} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [\mathbf{A} \text{ xstit}] \varphi \mid X\varphi \mid P\varphi$$

Where $\mathbf{p} \in \mathbf{At}_B \cup \mathbf{At}_I$

\mathcal{L}^B :

$$\varphi := \mathbf{p} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [\mathbf{A} \text{ xstit}] \varphi \mid X\varphi \mid P\varphi$$

Where $\mathbf{p} \in \mathbf{At}_B$

\mathcal{L}_Ω^B :

$$\varphi := \mathbf{p} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [\mathbf{A} \text{ xstit}] \varphi \mid X\varphi \mid P\varphi$$

Where $\mathbf{p} \in \mathbf{At}_B \cup \text{at}(\Omega)$

\mathcal{L}^I :

$$\varphi := \mathbf{p} \mid V \mid (\varphi \wedge \varphi) \mid \neg(\varphi) \mid (\varphi \supset \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \equiv \varphi) \mid X\varphi \mid P\varphi \mid [\mathbf{R} \text{ xstit}] \varphi$$

Where $\mathbf{R} \subseteq \mathbf{Rol}$, and $\mathbf{p} \in \mathbf{At}_B \cup \mathbf{At}_I$

\mathcal{L}_\in^I :

$$\varphi := \mathbf{p} \mid V \mid (\varphi \wedge \varphi) \mid \neg(\varphi) \mid (\varphi \supset \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \equiv \varphi) \mid X\theta \mid P\theta \mid [\mathbf{R} \text{ xstit}] \theta \mid \theta \in \theta'$$

Where $\mathbf{p} \in \mathbf{At}_I \cup \mathbf{At}_B$ $\mathbf{R} \subseteq \mathbf{Rol}$, and $\theta, \theta' \in \mathcal{L}^I$.

A.2 Models

A regular \mathcal{L} -frame is a triple $\mathfrak{F} = \langle S, H, E \rangle$ such that:

1. $S \neq \emptyset$, are the static states.
2. $H \neq \emptyset$ is a set of orders $\langle h, <_h \rangle$ such that for each $h \in H$

- (a) $h \subseteq S$ and $\langle h, <_h \rangle$ is isomorphic to \mathbb{Z} with its usual order, and
 - (b) if $s \in h \cap h'$, then $\{s' : s' <_h s\} = \{s' : s' <_{h'} s\}$. Since each order is isomorphic with \mathbb{Z} , there is a unique successor and predecessor for each $s \in h$, we refer to these by $\text{lub}(s, h)$ and $\text{glb}(s, h)$ respectively. We can generalise these concepts in the following way: $\text{glb}(s) = \{s' : \exists h \text{ glb}(s, h) = s'\}$ and $\text{lub}(s) = \{s' : \exists h \text{ lub}(s, h) = s'\}$. These give the set of successors and predecessors of s , respectively.
3. $E : S \times H \times \mathcal{P}(\mathbf{Ag}) \rightarrow \mathcal{P}(S)$ is called an h -effectivity function. The effectivity function provides a set of states that, relative to a history h a group of agents is effective in ensuring from a given state s . The function E must obey the following conditions:
- (a) if $s \notin h$, then $E(s, h, \mathbf{A}) = \emptyset$
 - (b) if $s' \in E(s, h, \mathbf{A})$, then $s' \in \text{lub}(s)$
 - (c) if $s \in h$, $\text{lub}(s, h) \in E(s, h, \mathbf{A})$
 - (d) $E(s, h, \emptyset) = \text{lub}(s)$, if $s \in h$
 - (e) if $s \in h$, then $E(s, h, \mathbf{Ag}) = \{\text{lub}(s, h)\}$
 - (f) if $\mathbf{A} \subsetneq \mathbf{B}$, then $E(s, h, \mathbf{B}) \subseteq E(s, h, \mathbf{A})$
 - (g) if $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $s \in h \cap h'$, then there is h'' with $s \in h''$ and $E(s, h'', \mathbf{A})$ and $E(s, h'', \mathbf{B})$ are contained in $E(s, h, \mathbf{A})$ and $E(s, h', \mathbf{B})$, respectively.

A universal regular frame is a regular frame such that $\bigcap H \neq \emptyset$.

A neutral (**NU**) \mathcal{L} -frame is a triple $\mathfrak{F} = \langle S, H, E, \approx \rangle$ such that:

1. $S \neq \emptyset$, are the static states.
2. \approx is an equivalence relation on S
3. $H \neq \emptyset$ is a set of triples $h = \langle h, f_h, b_h \rangle$ with $h \subseteq S$ such that
 - H1 each $\langle h, f_h, b_h \rangle \in H$ is an injective **DDL**F,
 - H2 if $s \in h$ and $s' \in h'$ with $s \approx s'$, then for each $n \in \mathbb{N}$, $b_h^n(s) \approx b_{h'}^n(s')$.
4. $\text{Id}_{\mathfrak{F}} = \{(s, h) \in S \times H : s \in h\}$
5. Again $\text{lub}_{\mathfrak{F}}(s, h) = f_h(s)$ and $\text{glb}_{\mathfrak{F}}(s, h) = b_h(s)$ but now
6. $\text{lub}_{\mathfrak{F}}(s) = \{s^* \in S : \exists h', s' \text{ w/ } s \approx s' \ \& \ f_{h'}(s') = s^*\}$, and
7. $\text{glb}_{\mathfrak{F}}(s) = \{s^* \in S : \exists h', s' \text{ w/ } s \approx s' \ \& \ b_{h'}(s') = s^*\}$.
8. $E : S \times H \times \mathcal{P}(\mathbf{Ag}) \rightarrow \mathcal{P}(S)$ is called an h -effectivity function. The effectivity function provides a set of states that, relative to a history h a group of agents is effective in ensuring from a given state s . The function E must obey the following conditions:
 - (a) if $s \notin h$, then $E(s, h, \mathbf{A}) = \emptyset$

- (b) if $s' \in E(s, h, \mathbf{A})$, then $s' \in \text{lub}(s)$
- (c) if $s \in h$, $\text{lub}(s, h) \in E(s, h, \mathbf{A})$
- (d) $E(s, h, \emptyset) = \text{lub}(s)$
- (e) if $s \in h$, then $E(s, h, \mathbf{A}\mathbf{g}) = \{s' : s' \approx \text{lub}(s, h)\}$
- (f) if $\mathbf{A} \subsetneq \mathbf{B}$, then $E(s, h, \mathbf{B}) \subseteq E(s, h, \mathbf{A})$
- (g) For all \mathbf{A}, \mathbf{B} $(s, h), (s', h'), (s'', h'') \in \mathbb{D}_{\mathfrak{F}}$, if $\mathbf{A} \cap \mathbf{B} = \emptyset$ and $s' \approx s \approx s''$, then there is $(s''', h''') \in \mathbb{D}_{\mathfrak{F}}$ such that $s''' \approx s$ with $E(s''', h''', \mathbf{A})$ and $E(s''', h''', \mathbf{B})$ contained in $E(s', h', \mathbf{A})$ and $E(s'', h'', \mathbf{B})$, respectively.

An $\mathcal{L}_{\text{xstit}}$, or \mathcal{L} neutral model \mathfrak{M} , is an neutral xstit frame \mathfrak{F} with a valuation $v : \mathbf{At} \rightarrow \mathcal{P}(S)$ such that if $s \approx s'$ and $s \in v(\mathbf{p})$, then $s' \in v(\mathbf{p})$.

If we want to interpret this for \mathcal{L}^I or $\mathcal{L}_{\subseteq}^I$, we remove the requirement of condition g, and replace \mathbf{A} with \mathbf{R} and \mathbf{B} with \mathbf{R}' .

A.3 Logics

Axioms for \vdash_x

Assume that $\mathbf{A}, \mathbf{B} \subseteq \mathbf{Ag}$ $\mathbf{p} \in \mathbf{At}$ and $\varphi, \psi \in \mathcal{L}_{\text{xstit}}$,

(p) $\mathbf{p} \supset \Box \mathbf{p}$

S5 for \Box :

- $\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$
- $\Box\varphi \supset \varphi$
- $\Box\varphi \supset \Box\Box\varphi$
- $\varphi \supset \Box\neg\Box\neg\varphi$

KD for each $[\mathbf{A} \text{ xstit}] \varphi$ and X :

- $[\mathbf{A} \text{ xstit}](\varphi \supset \psi) \supset ([\mathbf{A} \text{ xstit}]\varphi \supset [\mathbf{A} \text{ xstit}]\psi)$
- $[\mathbf{A} \text{ xstit}]\varphi \supset \neg[\mathbf{A} \text{ xstit}]\neg\varphi$
- $X(\varphi \supset \psi) \supset (X\varphi \supset X\psi)$
- $X\varphi \supset \neg X\neg\varphi$

(DetX) $\neg X\neg\varphi \supset X\varphi$

(\emptyset =SettX) $[\emptyset \text{ xstit}]\varphi \equiv \Box X\varphi$

(\mathbf{Ag} =XSett) $[\mathbf{Ag} \text{ xstit}]\varphi \equiv X\Box\varphi$

(C-mon) $[\mathbf{A} \text{ xstit}]\varphi \supset [\mathbf{A} \cup \mathbf{B} \text{ xstit}]\varphi$

(Indep-G) $\Diamond[\mathbf{A} \text{ xstit}]\varphi \wedge \Diamond[\mathbf{B} \text{ xstit}]\psi \supset \Diamond([\mathbf{A} \text{ xstit}]\varphi \wedge [\mathbf{B} \text{ xstit}]\psi)$ where $\mathbf{A} \cap \mathbf{B} = \emptyset$.

Axioms for the Logic SI (\vdash_{SI}):

Axioms for Classical Propositional Logic (CL)

- CL1 $\varphi \supset (\psi \supset \varphi)$
- CL2 $(\varphi \supset (\psi \supset \theta)) \supset ((\varphi \supset \psi) \supset (\psi \supset \theta))$
- CL3 $(\varphi \wedge \psi) \supset \psi$
- CL4 $(\varphi \wedge \psi) \supset \varphi$
- CL5 $(\varphi \supset \psi) \supset ((\varphi \supset \theta) \supset (\varphi \supset \psi \wedge \theta))$
- CL6 $\varphi \supset (\varphi \vee \psi)$
- CL7 $\psi \supset (\varphi \vee \psi)$
- CL8 $(\varphi \supset \psi) \supset ((\theta \supset \psi) \supset (\varphi \vee \theta \supset \psi))$
- CL9 $(\psi \supset \neg\varphi) \supset (\varphi \supset \neg\psi)$
- CL10 $\neg(\psi \supset \psi) \supset \varphi$
- CL11 $\varphi \vee \neg\varphi$
- CL12 $(\varphi \wedge \neg\varphi) \supset \perp$

Axioms for Propositional Containment

- PC1 $A \in A$
- PC2 $(B \in A) \supset ((C \in B) \supset (C \in A))$
- PC3 $(\mathbf{p}_i \in \mathbf{p}_j) \supset (\mathbf{p}_j \in \mathbf{p}_i)$
- PC4 $A \in (A \wedge B)$
- PC5 $B \in (A \wedge B)$
- PC6 $(B \in A) \supset ((C \in A) \supset ((C \wedge B) \in A))$
- PC7 $A \in \neg A$
- PC8 $\neg A \in A$
- PC9 $(\mathbf{p}_i \in (A \wedge B)) \supset ((\mathbf{p}_i \in A) \vee (\mathbf{p}_i \in B))$
- PC10 $\perp \in A$

Rules

- MP If $\vdash_{SI} \varphi \supset \psi$ and $\vdash_{SI} \varphi$, then $\vdash_{SI} \psi$

Axioms for \vdash_{xp} .

1. Axioms for classical logic

(p) $p \supset \Box p, p \in \mathbf{At}_B \cup \mathbf{At}_I$

S5 for \Box :

$$\mathbf{K} \quad \Box(\theta \supset \theta') \supset (\Box\theta \supset \Box\theta')$$

$$\mathbf{T} \quad \Box\theta \supset \theta$$

$$4 \quad \Box\theta \supset \Box\Box\theta$$

$$\mathbf{B} \quad \theta \supset \Box\neg\Box\neg\theta$$

KD for each $[\mathbf{A} \text{ xstit}] \theta, \mathbf{A} \subseteq \mathbf{Ag}, P$ and X :

$$\mathbf{KA} \quad [\mathbf{A} \text{ xstit}](\theta \supset \theta') \supset ([\mathbf{A} \text{ xstit}] \theta \supset [\mathbf{A} \text{ xstit}] \theta')$$

$$\mathbf{DA} \quad [\mathbf{A} \text{ xstit}] \theta \supset \neg[\mathbf{A} \text{ xstit}] \neg\theta$$

$$\mathbf{KX} \quad X(\theta \supset \theta') \supset (X\theta \supset X\theta')$$

$$\mathbf{DX} \quad X\theta \supset \neg X\neg\theta$$

$$\mathbf{KP} \quad P(\theta \supset \theta') \supset (P\theta \supset P\theta')$$

$$\mathbf{DP} \quad P\theta \supset \neg P\neg\theta$$

$$[(\mathbf{DetX})] \neg X\neg\theta \supset X\theta$$

$$[(\mathbf{DetP})] \neg P\neg\theta \supset P\theta$$

$$[(\mathbf{XP})] XP\theta \equiv \theta$$

$$[(\mathbf{PX})] \theta \equiv PX\theta$$

$$[(\mathbf{NP})] P\Box\theta \supset \Box P\theta$$

$$[(\mathbf{SettX})] [\emptyset \text{ xstit}] \theta \equiv \Box X\theta$$

$$[(\mathbf{XSett})] [\mathbf{Ag} \text{ xstit}] \theta \equiv X\Box\theta$$

$$[(\mathbf{C-mon})] [\mathbf{A} \text{ xstit}] \theta \supset [\mathbf{A} \cup \mathbf{B} \text{ xstit}] \theta$$

$$[(\mathbf{Indep-G})] \Diamond [\mathbf{A} \text{ xstit}] \theta \wedge \Diamond [\mathbf{B} \text{ xstit}] \theta' \supset \Diamond ([\mathbf{A} \text{ xstit}] \theta \wedge [\mathbf{B} \text{ xstit}] \theta') \text{ where } \mathbf{A} \cap \mathbf{B} = \emptyset.$$

2. Rules: MP and Nec \clubsuit for $\clubsuit \in \{\Box, X, P, [\mathbf{A} \text{ xstit}] : \mathbf{A} \subseteq \mathbf{Ag}\}$

Axioms for \vdash_{xp}^I

1. Axioms for classical logic

(p) $p \supset \Box p, p \in \mathbf{At}_B \cup \mathbf{At}_I$

S5 for \Box :

$$\mathbf{K} \Box(\theta \supset \theta') \supset (\Box\theta \supset \Box\theta')$$

$$\mathbf{T} \Box\theta \supset \theta$$

$$4 \Box\theta \supset \Box\Box\theta$$

$$\mathbf{B} \theta \supset \Box\neg\Box\neg\theta$$

KD for each $[\mathbf{R} \text{ xstit}] \theta$, $\mathbf{R} \subseteq \mathbf{Rol}$, P and X :

$$\mathbf{KR} [\mathbf{R} \text{ xstit}](\theta \supset \theta') \supset ([\mathbf{R} \text{ xstit}] \theta \supset [\mathbf{R} \text{ xstit}] \theta')$$

$$\mathbf{DR} [\mathbf{R} \text{ xstit}] \theta \supset \neg [\mathbf{R} \text{ xstit}] \neg\theta$$

$$\mathbf{KX} X(\theta \supset \theta') \supset (X\theta \supset X\theta')$$

$$\mathbf{DX} X\theta \supset \neg X\neg\theta$$

$$\mathbf{KP} P(\theta \supset \theta') \supset (P\theta \supset P\theta')$$

$$\mathbf{DP} P\theta \supset \neg P\neg\theta$$

$$[(\mathbf{DetX})] \neg X\neg\theta \supset X\theta$$

$$[(\mathbf{DetP})] \neg P\neg\theta \supset P\theta$$

$$[(\mathbf{XP})] XP\theta \equiv \theta$$

$$[(\mathbf{PX})] \theta \equiv PX\theta$$

$$[(\mathbf{NP})] P\Box\theta \supset \Box P\theta$$

$$[(\mathbf{SettX})] [\emptyset \text{ xstit}] \theta \equiv \Box X\theta$$

$$[(\mathbf{XSett})] [\mathbf{Rol} \text{ xstit}] \theta \equiv X\Box\theta$$

$$[(\mathbf{C-mon})] [\mathbf{R} \text{ xstit}] \theta \supset [\mathbf{R} \cup \mathbf{R}' \text{ xstit}] \theta$$

2. Rules: MP and Nec \clubsuit for $\clubsuit \in \{\Box, X, P, [\mathbf{R} \text{ xstit}] : \mathbf{R} \subseteq \mathbf{Rol}\}$

Axioms for $\vdash_{\text{xp}}^\Omega$ are dependant on what language \mathcal{L}_Ω^B is being used, notice that Indep-G is there:

- Axioms for classical logic

$$(\mathbf{p}) \mathbf{p} \supset \Box\mathbf{p}, \mathbf{p} \in \mathbf{At}_B \cup \text{at}(\Omega)$$

S5 for \Box :

$$\mathbf{K} \Box(\theta \supset \theta') \supset (\Box\theta \supset \Box\theta')$$

$$\mathbf{T} \Box\theta \supset \theta$$

$$4 \Box\theta \supset \Box\Box\theta$$

$$\mathbf{B} \theta \supset \Box\neg\Box\neg\theta$$

KD for each $[\mathbf{A} \text{ xstit}] \theta$, $\mathbf{A} \subseteq \mathbf{Ag}$, P and X :

KA $[\mathbf{A} \text{ xstit}](\theta \supset \theta') \supset ([\mathbf{A} \text{ xstit}] \theta \supset [\mathbf{A} \text{ xstit}] \theta')$

DA $[\mathbf{A} \text{ xstit}] \theta \supset \neg [\mathbf{A} \text{ xstit}] \neg \theta$

KX $X(\theta \supset \theta') \supset (X\theta \supset X\theta')$

DX $X\theta \supset \neg X\neg\theta$

KP $P(\theta \supset \theta') \supset (P\theta \supset P\theta')$

DP $P\theta \supset \neg P\neg\theta$

(DetX) $\neg X\neg\theta \supset X\theta$

(DetP) $\neg P\neg\theta \supset P\theta$

(XPPX) $XP\theta \equiv \theta \equiv PX\theta$

(NP) $P\Box\theta \supset \Box P\theta$

(SettX) $[\emptyset \text{ xstit}] \theta \equiv \Box X\theta$

XSett $[\mathbf{Ag} \text{ xstit}] \theta \equiv X\Box\theta$

(C-mon) $[\mathbf{A} \text{ xstit}] \theta \supset [\mathbf{A} \cup \mathbf{B} \text{ xstit}] \theta$

(Indep-G) $\Diamond [\mathbf{A} \text{ xstit}] \theta \wedge \Diamond [\mathbf{B} \text{ xstit}] \theta' \supset \Diamond ([\mathbf{A} \text{ xstit}] \theta \wedge [\mathbf{B} \text{ xstit}] \theta')$ where $\mathbf{A} \cap \mathbf{B} = \emptyset$.

Axioms of \vdash_{Ixp} .

1. First we include all axioms for classical logic (Group CL axioms) where $\varphi, \psi \in \mathcal{L}_{\subseteq}^I$:

CL1 $\varphi \supset (\psi \supset \varphi)$

CL2 $(\varphi \supset (\psi \supset \theta)) \supset ((\varphi \supset \psi) \supset (\psi \supset \theta))$

CL3 $(\varphi \wedge \psi) \supset \psi$

CL4 $(\varphi \wedge \psi) \supset \varphi$

CL5 $(\varphi \supset \psi) \supset ((\varphi \supset \theta) \supset (\varphi \supset \psi \wedge \theta))$

CL6 $\varphi \supset (\varphi \vee \psi)$

CL7 $\psi \supset (\varphi \vee \psi)$

CL8 $(\varphi \supset \psi) \supset ((\theta \supset \psi) \supset (\varphi \vee \theta \supset \psi))$

CL9 $(\psi \supset \neg\varphi) \supset (\varphi \supset \neg\psi)$

CL10 $\neg(\psi \supset \psi) \supset \varphi$

CL11 $\varphi \vee \neg\varphi$

CL12 $(\varphi \wedge \neg\varphi) \supset \perp$

2. We extend the axioms for propositional containment (group PC axioms) where $A, B, C \in \mathcal{L}^I$, $\mathbf{s}, \mathbf{s}' \in \mathbf{At}_I \cup \mathbf{At}_B \cup \{V\}$ and $\mathbf{R} \cup \{\mathbf{r}, \mathbf{r}'\} \subseteq \mathbf{Rol}$:

- PC1 $A \in A$
 PC2 $(B \in A) \supset ((C \in B) \supset (C \in A))$
 PC3 $(s \in s') \supset (s' \in s)$
 PC4 $A \in (A \wedge B)$
 PC5 $B \in (A \wedge B)$
 PC6 $(B \in A) \supset ((C \in A) \supset ((C \wedge B) \in A))$
 PC7 $A \in \neg A$
 PC8 $\neg A \in A$
 PC9 $(s \in (A \wedge B)) \supset ((s \in A) \vee (s \in B))$
 PC10 $\perp \in A$
 PC11 $A \in (A \vee B)$
 PC12 $B \in (A \vee B)$
 PC12A $(A \vee B) \in (A \wedge B)$
 PC13 $A \in (A \supset B)$
 PC14 $B \in (A \supset B)$
 PC14A $(A \supset B) \in (A \wedge B)$
 PC15 $A \in (A \equiv B)$
 PC16 $B \in (A \equiv B)$
 PC16A $(A \equiv B) \in (A \wedge B)$
 PC17 $\top \in A$
 PCX1 $A \in \Box A$
 PCX2 $XA \in A$
 PCX3 $PA \in XA$
 PCX4 $\Box A \in PA$
 PCX5 $A \in [\mathbf{r} \text{ xstit}] A$
 PCX6 $\mathbf{R} \in [\mathbf{R} \text{ xstit}] A$
 PCX7 $\perp \in \mathbf{r}$
 PCX8 $(\{\mathbf{r}\} \in \{\mathbf{r}'\}) \supset (\{\mathbf{r}'\} \in \{\mathbf{r}\})$
 PCX9 $\{\mathbf{r}\} \in \mathbf{R} \text{ for } \mathbf{r} \in \mathbf{R} \subseteq \mathbf{R} \mathbf{ol}$
 PCX10 $\neg(\{\mathbf{r}\} \in \mathbf{p}) \wedge \neg(\mathbf{p} \in \{\mathbf{r}\})$

3. We then extend the axioms for xstit (we call these the XPstit-group) by the following for $\theta, \theta' \in \mathcal{L}_{\in}^I$,

$$(\mathbf{p}) \mathbf{p} \supset \Box \mathbf{p}$$

S5 for \Box :

$$\mathbf{K} \quad \Box(\theta \supset \theta') \supset (\Box\theta \supset \Box\theta')$$

$$\mathbf{T} \quad \Box\theta \supset \theta$$

$$\mathbf{4} \quad \Box\theta \supset \Box\Box\theta$$

$$\mathbf{B} \quad \theta \supset \Box\neg\Box\neg\theta$$

KD for each $[\mathbf{R} \text{ xstit}] \theta$, $\mathbf{R} \subseteq \mathbf{Rol}$, P and X :

$$\mathbf{KR} \quad [\mathbf{R} \text{ xstit}](\theta \supset \theta') \supset ([\mathbf{R} \text{ xstit}] \theta \supset [\mathbf{R} \text{ xstit}] \theta')$$

$$\mathbf{DR} \quad [\mathbf{R} \text{ xstit}] \theta \supset \neg [\mathbf{R} \text{ xstit}] \neg\theta$$

$$\mathbf{KX} \quad X(\theta \supset \theta') \supset (X\theta \supset X\theta')$$

$$\mathbf{DX} \quad X\theta \supset \neg X\neg\theta$$

$$\mathbf{KP} \quad P(\theta \supset \theta') \supset (P\theta \supset P\theta')$$

$$\mathbf{DP} \quad P\theta \supset \neg P\neg\theta$$

$$(\mathbf{DetX}) \quad \neg X\neg\theta \supset X\theta$$

$$(\mathbf{DetP}) \quad \neg P\neg\theta \supset P\theta$$

$$(\mathbf{XP}) \quad XP\theta \equiv \theta$$

$$(\mathbf{PX}) \quad \theta \equiv PX\theta$$

$$(\mathbf{NP}) \quad P\Box\theta \supset \Box P\theta$$

$$(\mathbf{SettX}) \quad [\emptyset \text{ xstit}] \theta \equiv \Box X\theta$$

$$(\mathbf{XSett}) \quad [\mathbf{Rol} \text{ xstit}] \theta \equiv X\Box\theta$$

$$(\mathbf{C-mon}) \quad [\mathbf{R} \text{ xstit}] \theta \supset [\mathbf{R} \cup \mathbf{R}' \text{ xstit}] \theta \text{ where } \mathbf{R}' \cup \mathbf{R} \subseteq \mathbf{Rol}$$

4. Rules: MP and Nec \clubsuit for $\clubsuit \in \{\Box, X, P, [\mathbf{R} \text{ xstit}] : \mathbf{R} \subseteq \mathbf{Rol}\}$

[Strong Implication] A set of sentences Γ strongly implies φ ($\Gamma \vdash_S \varphi$) iff there are sentences $\gamma_1, \dots, \gamma_n$ in Γ such that $\vdash_{CL} \gamma_1 \wedge \dots \wedge \gamma_n \supset \varphi$, and there are $\gamma'_1, \dots, \gamma'_m$ in Γ such that $\vdash_{SI} \gamma'_1 \wedge \dots \wedge \gamma'_m \supset (\varphi \Subset \gamma_1 \wedge \dots \wedge \gamma_n)$

[Norm Consequence] Ω Norm Entails φ ($\Omega \vdash_N \varphi$) iff there are $\psi'_1, \dots, \psi'_m \in \Omega$ s.t.

$$\mathbf{NC1}: \Omega \vdash_{\text{Ixp}} \varphi,$$

$$\mathbf{NC2}: \vdash_{\text{Ixp}} \varphi \Subset (\psi'_1 \wedge \dots \wedge \psi'_m), \text{ and}$$

$$\mathbf{NC3}: \varphi \in IC(\mathcal{L}^I).$$

Appendix B

Some Background

B.1 Kripke Models

A standard modal language generated by the grammar

$$\varphi := \mathbf{p} \mid \varphi \wedge \varphi \mid \neg\varphi \mid \varphi \supset \varphi \mid \Box\varphi$$

where $\mathbf{p} \in \mathbf{At}$ the set of atomic sentences, can be interpreted on a Kripke model. Kripke models consist of two things: Kripke Frames and valuations.

Definition B.1.1. A *Kripke frame* is a pair $\langle W, R \rangle$ consisting of a non-empty set W and a relation on W , i.e., $R \subseteq W \times W$.

A Kripke model $\mathfrak{M} = \langle W, R, v \rangle$ is a Kripke frame along with a valuation $v : \mathbf{At} \rightarrow \mathcal{P}(W)$.

The semantics is defined as follows:

- $\mathfrak{M}, w \Vdash \mathbf{p}$ iff $w \in v(\mathbf{p})$;
- $\mathfrak{M}, w \Vdash \neg\varphi$ iff $\mathfrak{M}, w \not\Vdash \varphi$;
- $\mathfrak{M}, w \Vdash \varphi \wedge \psi$ iff $\mathfrak{M}, w \Vdash \varphi$ and $\mathfrak{M}, w \Vdash \psi$;
- $\mathfrak{M}, w \Vdash \varphi \supset \psi$ iff $\mathfrak{M}, w \not\Vdash \varphi$ or $\mathfrak{M}, w \Vdash \psi$;
- $\mathfrak{M}, w \Vdash \Box\varphi$ iff for all $w' \in W$, $\langle w, w' \rangle \in R$ only if $\mathfrak{M}, w' \Vdash \varphi$.

B.2 Boolean Algebras

Definition B.2.1. A boolean algebra $\langle B, 1, 0, \sqcap, \sqcup, -() \rangle$ is a set B with two binary operations \sqcap (meet), and \sqcup (join) along with a unary operation $-()$. There are also two special elements 0 and 1. These operations obey the following equations for $a, b, c \in B$:

- $a \sqcap \neg(a) = 0, a \sqcup \neg(a) = 1$
- $a \sqcap (b \sqcup c) = (a \sqcap b) \sqcup (a \sqcap c), a \sqcup (b \sqcap c) = (a \sqcup b) \sqcap (a \sqcup c),$
- $a \sqcap a = a, a \sqcup a = a,$
- $a \sqcup (a \sqcap b) = a, a \sqcap (a \sqcup b) = a,$
- $a \sqcap b = b \sqcap a, a \sqcup b = b \sqcup a.$

We can also define a partial order on the set B by defining $a \leq b$ iff $a \sqcap b = a$ (or iff $a \sqcup b = b$). The special element 1 is often referred to as the “top” and the 0 as the “bottom”; this is so since 1 is the largest element relative to \leq and 0 is the smallest element in B in the respective sense. These special elements also have the properties that $a \sqcap 1 = a, a \sqcup 0 = a, \neg(1) = 0$ and vice versa. There are also some things that follow about negation and other connectives. In addition, we have

1. double negation: $\neg(\neg(a)) = a$, and
2. the DeMorgan rules: $\neg(a \sqcap b) = \neg(a) \sqcup \neg(b), \neg(a \sqcup b) = \neg(a) \sqcap \neg(b).$

With the DeMorgan relationships it follows that if $a \leq b$, then $\neg(b) \leq \neg(a).$

B.3 Fusion of Logics

The fusion of logics is smallest logic extending all of the logics that are to be fused. Of course formally speaking the concept is a lot more complicated. Since what is of interest to us in this essay is the fusion of normal modal logics we will discuss that particular kind of fusion. The idea was introduced as a general method of combining logics in Thomason (1984), and it was extensively studied in Fine and Schurz (1996).

A modal logic consists of three things, first a modal language \mathcal{L}_i that consists of a set of formulas constructed recursively from a signature of atomic sentences \mathbf{At} , the set of boolean connectives, and a unary modal operator: \Box_i , in the usual way. Second, a Hilbert style axiomatisation Ax that includes the K axiom for \Box_i , the rule of modus ponens, and necessitation

for \Box_i . Third, there is a Kripke semantics for the logic which consists of a class of Kripke frames F_i , such that each $\mathfrak{F} \in F_i$ is like $\langle W, R \rangle$, where R is a relation on W and $W \neq \emptyset$. Here we are working with a monomodal logic, i.e., only one modal operator, but we could extend the definition so that it included multiple modal operators, the only difference is that there would be, for each \Box_i a corresponding R_i in each of the Kripke frames in F . Thus a modal logic L_i is a triple $\langle \mathcal{L}_i, Ax_i, F_i \rangle$ as just described.

If we have two such logics, L_1 and L_2 , then we can define the fusion of the two logics, denoted by $L_1 \oplus L_2$, as the pointwise fusion of each “part” of the component logics. That is, $L_1 \oplus L_2 = \langle \mathcal{L}_1 \oplus \mathcal{L}_2, Ax_1 \oplus Ax_2, F_1 \oplus F_2 \rangle$. The fusion of languages is done as follows. First note that each \mathcal{L}_i , $i = 1, 2$, contains the same boolean connectives and atomic sentences **At**. We take the union of the signatures in the \mathcal{L}_i s and recursively generate a new language with from **At**, the boolean connectives and the collection of the modal operators. So $\mathcal{L}_1 \oplus \mathcal{L}_2$ is

$$\varphi := \mathbf{p} \mid \Box_1\varphi \mid \Box_2\varphi \mid \neg\varphi \mid \varphi \supset \varphi \mid \varphi \wedge \varphi$$

For example, if the languages were \mathcal{L}_1 =boolean connectives, **At** and \Box_1 , and \mathcal{L}_2 =boolean connectives, **At** and \Box_2 , then we would get formulas like $\Box_1\Box_2(\mathbf{p} \supset \mathbf{q})$ and $\Box_1(\Box_2\mathbf{p} \vee \Box_1(\mathbf{q} \wedge \Box_2\mathbf{p}'))$.

$Ax_1 \oplus Ax_2$ is formed by taking the union of the sets of axiom schema.

Finally, the fusion of classes of Kripke semantics consists of “putting frames together”. That is,

$$F_1 \oplus F_2 = \{ \langle W, R_1, R_2 \rangle : \langle W, R_1 \rangle \in F_1 \ \& \ \langle W, R_2 \rangle \in F_2 \}.$$

Note that we can only fuse frames that have the same domain. An example: if we had a F_1 that is the class of frames where R_1 is transitive and reflexive, and F_2 is the class of frames where R_2 is serial, then $F_1 \oplus F_2$ is the class of frames $\langle W, R_1, R_2 \rangle$ such that R_1 is transitive and reflexive, and R_2 is serial.

Fusion models are then fused frames with a valuation on them, i.e., $v : \mathbf{At} \rightarrow \mathcal{P}(W)$, and the usual semantics such that for \mathfrak{M} , a fusion model of L_1 and L_2 , $w \in W$, $\varphi \in \mathcal{L}_1 \oplus \mathcal{L}_2$ and

$\mathbf{p} \in \mathbf{At}$,

- $\mathfrak{M}, w \Vdash \mathbf{p}$ iff $w \in v(\mathbf{p})$,
- Boolean clauses are standard,
- $\mathfrak{M}, w \Vdash \Box_1 \varphi$ iff for all $w' \in W$, $\langle w, w' \rangle \in R_1$ only if $\mathfrak{M}, w' \Vdash \varphi$, and
- $\mathfrak{M}, w \Vdash \Box_2 \varphi$ iff for all $w' \in W$, $\langle w, w' \rangle \in R_2$ only if $\mathfrak{M}, w' \Vdash \varphi$.

It turns out that if Ax_1 is sound (complete) wrt F_1 and the same holds for Ax_2 wrt F_2 , then $Ax_1 \oplus Ax_2$ is sound (complete) wrt $F_1 \oplus F_2$. Moreover, $L_1 \oplus L_2$ will be decidable if both L_1 and L_2 are (see Fine and Schurz, 1996).