Related Papers:

Neural Network Semantics / Semantic Encodings.

Classic Papers. [17]

Conditional Logic (Feedforward Net). [2], [14], [15], [8] (soundness), [9] (model-building) [Any other relevant work by the Garcez lab?]

Description Logic w. Typicality. [10], [11] [Any other relevant work by the Giordano lab?]

Modal Logic w. Typicality. [13]

[Any other big trends I'm missing? See the new survey by Odense + Garcez!]

Miscellaneous. [5], [6]

Surveys. [18] [1], [20], [12], [16], [3], [21] (the first few sections are a great introduction to Neural Network Semantics)

Help with Technical Details.

Neighborhood Models. [19]

Temporal Logic Rules. [7]

Nominals (Hybrid Logic). [4]

Step 4. Write up my new definitions & proof in the Texmacs file. Again, should be a *very* straightforward extension, and the proof (proofs are just unit-tests for definitions) shouldn't take up too much room at all (1-2 pages, including defs)

1 Interpreted Neural Nets

1.1 Basic Definitions

DEFINITION 1.1. An **interpreted ANN** (Artificial Neural Network) is a pointed directed graph $\mathcal{N} = \langle N, E, W, A, O, V \rangle$, where

- N is a finite nonempty set (the set of **neurons**)
- $E \subseteq N \times N$ (the set of **excitatory neurons**)
- $W : E \to \mathbb{R}$ (the **weight** of a given connection)
- A is a function which maps each $n \in N$ to $A^{(n)} : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ (the **activation function** for n, where k is the indegree of n)
- O is a function which maps each $n \in N$ to $O^{(n)} : \mathbb{R} \to \{0, 1\}$ (the **output function** for n)
- V: propositions $\cup$ nominals $\to \mathcal{P}(N)$ is an assignment of nominals to individual neurons (the **valuation function**). If i is a nominal, we require $|V(i)| = 1$, i.e. a singleton.

DEFINITION 1.2. A **BFNN** (Binary Feedforward Neural Network) is an interpreted ANN $\mathcal{N} = \langle N, E, W, A, O, V \rangle$ that is

- **Feed-forward**: E does not contain any cycles
- **Binary**: the output of each neuron is in $\{0, 1\}$
- $O^{(n)} \circ A^{(n)}$ is **zero at zero** in the first parameter: $O^{(n)}(A^{(n)}(\vec{0}, \vec{w})) = 0$
- $O^{(n)} \circ A^{(n)}$ is **strictly monotonically increasing** in the second parameter: for all $\vec{x}, \vec{w}_1, \vec{w}_2 \in \mathbb{R}^k$, if $\vec{w}_1 < \vec{w}_2$ then $O^{(n)}(A^{(n)}(\vec{x}, \vec{w}_1)) < O^{(n)}(A^{(n)}(\vec{x}, \vec{w}_2))$. We will more often refer to the equivalent condition:

$$\vec{w}_1 \leqslant \vec{w}_2 \quad \text{iff} \quad O^{(n)}(A^{(n)}(\vec{x}, \vec{w}_1)) \leqslant O^{(n)}(A^{(n)}(\vec{x}, \vec{w}_2))$$

DEFINITION 1.3. Given a BFNN $\mathcal{N}$, $\mathsf{Set} = \mathcal{P}(N) = \{S | S \subseteq N\}$

DEFINITION 1.4. For $S \in \mathsf{Set}$, let $\chi_S : N \to \{0, 1\}$ be given by $\chi_S = 1$ iff $n \in S$

We write W_{ij} to mean $W(i, j)$ for $(i, j) \in E$. To keep the notation from getting really messy, I'll also define:

DEFINITION 1.5. Let $S \in \mathsf{Set}$, $\vec{m} = m_1, \ldots, m_k$ be a sequence where each $m_i \in N$, and let $n \in N$. Then:

$$\mathsf{Activates}_S(\vec{m}, n) = O^{(n)}(A^{(n)}((\chi_S(m_1), \ldots, \chi_S(m_k)); (W(m_1, n), \ldots, W(m_k, n))))$$

i.e. the $m_i \in S$ subsequently "activate" n.

PROPOSITION 1.6. Let $S_1, S_2 \in \mathsf{Set}$, $\vec{m} = m_1, \ldots, m_k$ be a sequence where each $m_i \in N$, and let $n \in N$. Suppose that S_1 and S_2 agree on all m_i, i.e. for all $1 \le i \le k$, $m_i \in S_1$ iff $m_i \in S_2$. Then

$$\mathsf{Activates}_{S_1}(\vec{m}, n) = \mathsf{Activates}_{S_2}(\vec{m}, n)$$

Proof. We have:

$$\begin{aligned}
\mathsf{Activates}_{S_1}(\vec{m}, n) &= O^{(n)}(A^{(n)}((\chi_{S_1}(m_1), \ldots, \chi_{S_1}(m_k)); (W(m_1, n), \ldots, W(m_k, n)))) \\
&= O^{(n)}(A^{(n)}((\chi_{S_2}(m_1), \ldots, \chi_{S_2}(m_k)); (W(m_1, n), \ldots, W(m_k, n)))) \\
&= \mathsf{Activates}_{S_2}(\vec{m}, n)
\end{aligned}$$

□

1.2 Prop and Reach

DEFINITION 1.7. (Adapted from [14, Definition 3.4]) Let $\mathsf{Prop}: \mathsf{Set} \to \mathsf{Set}$ be defined recursively as follows: $n \in \mathsf{Prop}(S)$ iff either

Base Case. $n \in S$, or

Constructor. For those $\vec{m} = m_1, \ldots, m_k$ such that $(m_i, n) \in E$, $\mathsf{Activates}_{\mathsf{Prop}(S)}(\vec{m}, n) = 1$.

DEFINITION 1.8. Let $\mathsf{Reach}: \mathsf{Set} \to \mathsf{Set}$ be defined recursively as follows: $n \in \mathsf{Reach}(S)$ iff either

Base Case. $n \in S$, or

Constructor. There is an $m \in \mathsf{Reach}(S)$ such that $(m, n) \in E$.

PROPOSITION 1.9. **(Alternate characteriz. of Reach)** Let $n \in N$, $S \in \mathsf{Set}$. Then $n \in \mathsf{Reach}(S)$ iff there is a path from some $m \in S$ to n in E.

Proof. [] □

PROPOSITION 1.10. Let $\mathcal{N} \in \mathsf{Net}$. For all $S, S_1, S_2 \in \mathsf{Set}$, $n, m \in N$, Reach is

(Inclusive). $S \subseteq \mathsf{Reach}(S)$

(Idempotent). $\mathsf{Reach}(S) = \mathsf{Reach}(\mathsf{Reach}(S))$

(Acyclic). If $S_1 \subseteq \mathsf{Reach}(S_2)$ and $S_2 \subseteq \mathsf{Reach}(S_1)$ then $S_1 = S_2$.

(Monotonic). If $S_1 \subseteq S_2$ then $\mathsf{Reach}(S_1) \subseteq \mathsf{Reach}(S_2)$

Proof. We check each in turn:

(Inclusive). If $n \in S$, then $n \in \mathsf{Reach}(S)$ by the base case of Reach.

(Idempotent). The $(\subseteq)$ direction is just Inclusion. As for $(\supseteq)$, let $n \in \mathsf{Reach}(\mathsf{Reach}(S))$, and proceed by induction on the outer Reach.

Base Step. $n \in \mathsf{Reach}(S)$, and so we are done.

Inductive Step. There is an $m \in \mathsf{Reach}(\mathsf{Reach}(S))$ such that $(m, n) \in E$. by inductive hypothesis, $m \in \mathsf{Reach}(S)$. And so by definition, $n \in \mathsf{Reach}(S)$.

(Acyclic). Suppose $S_1 \subseteq \mathsf{Reach}(S_2)$ and $S_2 \subseteq \mathsf{Reach}(S_1)$. We will show $S_1 \subseteq S_2$ (the other direction is similar). [Todo]

(Monotonic). Let $n \in \mathsf{Reach}(S_1)$. We proceed by induction on $\mathsf{Reach}(S_1)$.

Base Step. $n \in S_1$. So $n \in S_2 \subseteq \mathsf{Reach}(S_2)$.

Inductive Step. There is an $m \in \mathsf{Reach}(S_1)$ such that $(m, n) \in E$. By inductive hypothesis, $m \in \mathsf{Reach}(S_2)$. And so by definition, $n \in \mathsf{Reach}(S_2)$.

□

PROPOSITION 1.11. (Adapted from [14, Remark 4]) Let $\mathcal{N} \in \mathsf{Net}$. For all $S, S_1, S_2 \in \mathsf{Set}$, Prop is

(Inclusive). $S \subseteq \mathsf{Prop}(S)$

(Idempotent). $\mathsf{Prop}(S) = \mathsf{Prop}(\mathsf{Prop}(S))$

(Contained in Reach). $\mathsf{Prop}(S) \subseteq \mathsf{Reach}(S)$

Proof. We check each in turn:

(Inclusive). Similar to the proof of Inclusion for Reach.

(Idempotent). The ($\subseteq$) direction is just Inclusion. As for ($\supseteq$), let $n \in \mathsf{Prop}(\mathsf{Prop}(S))$, and proceed by induction on $\mathsf{Prop}(\mathsf{Prop}(S))$.

Base Step. $n \in \mathsf{Prop}(S)$, and so we are done.

Inductive Step. For those $\vec{m} = m_1, \ldots, m_k$ such that $(m_i, n) \in E$,

$$\mathsf{Activates}_{\mathsf{Prop}(\mathsf{Prop}(S))}(\vec{m}, n) = 1$$

By inductive hypothesis, $m_i \in \mathsf{Prop}(\mathsf{Prop}(S))$ iff $m_i \in \mathsf{Prop}(S)$. By Proposition 1.6, $\mathsf{Activates}_{\mathsf{Prop}(S)}(\vec{m}, n) = 1$, and so $n \in \mathsf{Prop}(S)$.

(Contained in Reach). Let $n \in \mathsf{Prop}(S)$, and proceed by induction on Prop.

Base Step. $n \in S$. So $n \in \mathsf{Reach}(S)$.

Inductive Step. For those $\vec{m} = m_1, \ldots, m_k$ such that $(m_i, n) \in E$,

$$\mathsf{Activates}_{\mathsf{Prop}(S)}(\vec{m}, n) = 1$$

Since $O \circ A$ is zero at zero, we have $m_i \in \mathsf{Prop}(S)$ for *some* $m = m_i$. By inductive hypothesis, $m \in \mathsf{Reach}(S)$. And since $(m, n) \in E$, by definition of Reach, $n \in \mathsf{Reach}(S)$. □

PROPOSITION 1.12. The Cumulative and Loop properties from [14] [The KLM Cumulative & Loop properties, actually], i.e.

(Cumulative). If $S_1 \subseteq S_2 \subseteq \mathsf{Prop}(S_1)$ then $\mathsf{Prop}(S_1) \subseteq \mathsf{Prop}(S_2)$

(Loop). If $S_1 \subseteq \mathsf{Prop}(S_0), \ldots, S_n \subseteq \mathsf{Prop}(S_{n-1})$ and $S_0 \subseteq \mathsf{Prop}(S_n)$, then $\mathsf{Prop}(S_i) = \mathsf{Prop}(S_j)$ for all $i, j \in \{0, \ldots, n\}$

follow from the properties of Prop and Reach above.

Proof. [Todo – note that (Cumulative) actually follows from (Loop). Use acyclic property of Reach to get (Loop)] □

1.3 Neural Network Semantics

DEFINITION 1.13. Formulas of our language $\mathcal{L}$ are given by

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathbf{K}\varphi \mid \mathbf{T}\varphi$$

where p is any propositional variable, and i is any nominal (denoting a neuron). Material implication $\varphi \to \psi$ is defined as $\neg\varphi \vee \psi$. We define $\bot, \vee, \leftrightarrow, \Leftrightarrow$, and the dual operators $\langle\mathbf{K}\rangle, \langle\mathbf{T}\rangle$ in the usual way.

DEFINITION 1.14. Let $\mathcal{N} \in \mathsf{Net}$. The semantics $[\![\cdot]\!] : \mathcal{L} \to \mathsf{Set}$ for $\mathcal{L}$ are defined recursively as follows:

$[\![p]\!]$	$=$	$V(p) \in \mathsf{Set}$
$[\![\neg\varphi]\!]$	$=$	$[\![\varphi]\!]^{\mathsf{C}}$
$[\![\varphi \wedge \psi]\!]$	$=$	$[\![\varphi]\!] \cap [\![\psi]\!]$
$[\![\langle\mathbf{K}\rangle\varphi]\!]$	$=$	$\mathsf{Reach}([\![\varphi]\!])$
$[\![\langle\mathbf{T}\rangle\varphi]\!]$	$=$	$\mathsf{Prop}([\![\varphi]\!])$

DEFINITION 1.15. **(Truth at a neuron)** $\mathcal{N}, n \Vdash \varphi$ iff $n \in [\![\varphi]\!]_{\mathcal{N}}$.

DEFINITION 1.16. **(Truth in a net)** $\mathcal{N} \models \varphi$ iff $\mathcal{N}, n \Vdash \varphi$ for all $n \in N$.

DEFINITION 1.17. **(Entailment)** $\Gamma \models_{\mathrm{BFNN}} \varphi$ if for all BFNNs $\mathcal{N}$ for all neurons $n \in N$, if $\mathcal{N}, n \models \Gamma$ then $\mathcal{N}$, $n \models \varphi$.

2 Neighborhood Models

2.1 Basic Definitions

DEFINITION 2.1. [19, Definition 1.9] A **neighborhood frame** is a pair $\mathcal{F} = \langle W, f \rangle$, where W is a non-empty set of **worlds** and $f: W \to \mathcal{P}(\mathcal{P}(W))$ is a **neighborhood function**. A **multi-frame** may have more than one neighborhood function, but to keep things simple I won't distinguish between frames and multi-frames.

DEFINITION 2.2. [19, Section 1.1] Let $\mathcal{F} = \langle W, f \rangle$ be a neighborhood frame, and let $w \in W$. The set $\bigcap_{X \in f(w)} X$ is called the **core of $f(w)$**, abbreviated $\cap f(w)$.

DEFINITION 2.3. [19, Definition 1.4] Let $\mathcal{F} = \langle W, f \rangle$ be a frame. $\mathcal{F}$ is a **proper filter** iff:

- f is **closed under finite intersections**: for all $w \in W$, if $X_1, \ldots, X_n \in f(w)$ then their intersection $\bigcap_{i=1}^{k} X_i \in f(w)$
- f is **closed under supersets**: for all $w \in W$, if $X \in f(w)$ and $X \subseteq Y \subseteq W$, then $Y \in f(w)$
- f **contains the unit**: iff $W \in f(w)$

PROPOSITION 2.4. [19, Corollary 1.1] If $\mathcal{F} = \langle W, f \rangle$ is a filter, and W is finite, then $\mathcal{F}$ contains its core.

Proof. [Todo] □

PROPOSITION 2.5. [19, [Which?]] If $\mathcal{F} = \langle W, f \rangle$ is a proper filter, then for all $w \in W$, $Y^{\mathsf{C}} \in f(w)$ iff $Y \notin f(w)$.

Proof. $(\to)$ Suppose for contradiction that $Y^{\mathsf{C}} \in f(w)$ and $Y \in f(w)$. Since $\mathcal{F}$ is closed under intersection, $Y^{\mathsf{C}} \cap Y = \emptyset \in f(w)$, which contradicts the fact that $\mathcal{F}$ is proper.

$(\leftarrow)$ Suppose for contradiction that $Y \notin f(w)$, yet $Y^{\mathsf{C}} \notin f(w)$. Since $\mathcal{F}$ is closed under intersection, $\cap f(w) \in f(w)$. Moreover, since $\mathcal{F}$ is closed under superset we must have $\cap f(w) \nsubseteq Y$ and $\cap f(w) \nsubseteq Y^{\mathsf{C}}$. But this means $\cap f(w) \nsubseteq Y \cap Y^{\mathsf{C}} = \emptyset$, i.e. there is some $x \in \cap f(w)$ such that $x \in \emptyset$. This contradicts the definition of the empty set. □

DEFINITION 2.6. Let $\mathcal{F} = \langle W, f, g \rangle$ be a frame. $\mathcal{F}$ is a **preferential filter** iff:

- W is finite
- $\langle W, f \rangle$ forms a proper filter, and g contains the unit
- f is **acyclic**: for all $u_1, \ldots, u_n \in W$, if $u_1 \in \cap f(u_2), \ldots, u_{n-1} \in \cap f(u_n), u_n \in \cap f(u_1)$ then all $u_i = u_j$.
- f, g are **reflexive**: for all $w \in W$, $w \in \cap f(w)$ (similarly for g)
- f, g are **transitive**: for all $w \in W$, if $X \in f(w)$ then $\{u | X \in f(u)\} \in f(w)$ (similarly for g)
- g contains f: for all $w \in W$, if $X \in f(w)$ then $X \in g(w)$.

PROPOSITION 2.7. Let $\mathcal{F} = \langle W, f \rangle$ be a frame. Suppose f is reflexive, transitive, and **asymmetric**, i.e. $u_1 \in \cap f(u_2)$ and $u_2 \in \cap f(u_1)$ implies $u_1 = u_2$. Then f is acyclic.

Proof. Let $u_1, \ldots, u_n \in W$, and suppose $u_1 \in \cap f(u_2), \ldots, u_{n-1} \in \cap f(u_n), u_n \in \cap f(u_1)$. WLOG we will show that $u_1 = u_n$. [Todo] □

2.2 Neighborhood Semantics

DEFINITION 2.8. [19, Definition 1.11] Let $\mathcal{F} = \langle W, f, g\rangle$ be a neighborhood frame. A **neighborhood model** based on $\mathcal{F}$ is $\mathcal{M} = \langle W, f, g, V\rangle$, where $V : \mathcal{L} \to \mathcal{P}(W)$ is a valuation function.

DEFINITION 2.9. [19, Definition 1.12] Let $\mathcal{M} = \langle W, f, g, V\rangle$ be a model based on $\mathcal{F} = \langle W, f, g\rangle$ The (neighborhood) semantics for $\mathcal{L}$ are defined recursively as follows:

$\mathcal{M}, w \Vdash p$	iff	$w \in V(p)$
$\mathcal{M}, w \Vdash \neg\varphi$	iff	$\mathcal{M}, w \nVdash \varphi$
$\mathcal{M}, w \Vdash \varphi \wedge \psi$	iff	$\mathcal{M}, w \Vdash \varphi$ and $\mathcal{M}, w \Vdash \psi$
$\mathcal{M}, w \Vdash \mathbf{K}\varphi$	iff	$\{u \mid \mathcal{M}, u \Vdash \varphi\} \in f(w)$
$\mathcal{M}, w \Vdash \mathbf{T}\varphi$	iff	$\{u \mid \mathcal{M}, u \Vdash \varphi\} \in g(w)$

In neighborhood semantics, the operators $\mathbf{K}$, and $\mathbf{T}$ are more natural to interpret. But when we gave our neural semantics, we instead interpreted the *duals* $\langle\mathbf{K}\rangle$, and $\langle\mathbf{T}\rangle$. Since we need to relate the two, I'll write the explicit neighborhood semantics for the duals here:

$$
\begin{array}{lll}
\mathcal{M}, w \Vdash \langle\mathbf{K}\rangle\varphi & \text{iff} & \{u \mid \mathcal{M}, u \nVdash \varphi\} \notin f(w) \\
\mathcal{M}, w \Vdash \langle\mathbf{T}\rangle\varphi & \text{iff} & \{u \mid \mathcal{M}, u \nVdash \varphi\} \notin g(w)
\end{array}
$$

DEFINITION 2.10. [19, Definition 1.13] **(Truth in a model)** $\mathcal{M} \models \varphi$ iff $\mathcal{M}, w \Vdash \varphi$ for all $w \in W$.

DEFINITION 2.11. [19, Definition 2.32] **(Entailment)** Let F be a collection of neighborhood frames. $\Gamma \models_{\mathsf{F}} \varphi$ if for all models $\mathcal{M}$ based on a frame $\mathcal{F} \in \mathsf{F}$ and for all worlds $w \in W$, if $\mathcal{M}, w \models \Gamma$ then $\mathcal{M}, w \models \varphi$.

Note. This is the *local* consequence relation in modal logic.

3 From Nets to Frames

This is the easy ("soundness") direction!

DEFINITION 3.1. Given a BFNN $\mathcal{N}$, its **simulation frame** $\mathcal{F}^\bullet = \langle W, f, g\rangle$ is given by:

- $W = N$
- $f(w) = \{S \subseteq W \mid w \notin \mathsf{Reach}(S^{\complement})\}$
- $g(w) = \{S \subseteq W \mid w \notin \mathsf{Prop}(S^{\complement})\}$

Moreover, the **simulation model** $\mathcal{M}^\bullet = \langle W, f, g, V\rangle$ based on $\mathcal{F}^\bullet$ has:

- $V_{\mathcal{M}^\bullet}(p) = V_{\mathcal{N}}(p)$;
- $V_{\mathcal{M}^\bullet}(i) = V_{\mathcal{N}}(i)$

THEOREM 3.2. Let $\mathcal{N}$ be a BFNN, and let $\mathcal{M}^\bullet$ be the simulation model based on $\mathcal{F}^\bullet$. Then for all $w \in W$,

$$\mathcal{M}^\bullet, w \Vdash \varphi \quad \text{iff} \quad \mathcal{N}, w \Vdash \varphi$$

Proof. By induction on φ. The propositional, $\neg\varphi$, and $\varphi \wedge \psi$ cases are trivial.

$\langle\mathbf{K}\rangle\varphi$ case:

$$
\begin{array}{llll}
\mathcal{M}^\bullet, w \Vdash \langle\mathbf{K}\rangle\varphi & \text{iff} & \{u \mid \mathcal{M}^\bullet, w \nVdash \varphi\} \notin f(w) & \text{(by definition)} \\
& \text{iff} & \{u \mid u \notin [\![\varphi]\!]\} \notin f(w) & \text{(IH)} \\
& \text{iff} & [\![\varphi]\!]^{\complement} \notin f(w) & \\
& \text{iff} & w \in \mathsf{Reach}([\![(\varphi^{\complement})^{\complement}]\!]) & \text{(by choice of } f) \\
& \text{iff} & w \in \mathsf{Reach}([\![\varphi]\!]) & \\
& \text{iff} & w \in [\![\langle\mathbf{K}\rangle\varphi]\!] & \text{(by definition)} \\
& \text{iff} & \mathcal{N}, w \Vdash \langle\mathbf{K}\rangle\varphi & \text{(by definition)}
\end{array}
$$

$\langle \mathbf{T} \rangle \varphi$ case:

$$\begin{array}{llll} \mathcal{M}^\bullet, w \Vdash \langle \mathbf{T} \rangle \varphi & \text{iff} & \{u \mid \mathcal{M}^\bullet, w \not\Vdash \varphi\} \notin g(w) & \text{(by definition)} \\ & \text{iff} & \{u \mid u \notin [\![\varphi]\!]\} \notin g(w) & \text{(IH)} \\ & \text{iff} & [\![\varphi]\!]^{\mathsf{C}} \notin g(w) & \\ & \text{iff} & w \in \mathsf{Prop}([\![(\varphi^{\mathsf{C}})^{\mathsf{C}}]\!]) & \text{(by choice of } g\text{)} \\ & \text{iff} & w \in \mathsf{Prop}([\![\varphi]\!]) & \\ & \text{iff} & w \in [\![\langle \mathbf{T} \rangle \varphi]\!] & \text{(by definition)} \\ & \text{iff} & \mathcal{N}, w \Vdash \langle \mathbf{T} \rangle \varphi & \text{(by definition)} \end{array}$$

□

COROLLARY 3.3. $\mathcal{M}^\bullet \models \varphi$ iff $\mathcal{N} \models \varphi$.

THEOREM 3.4. $\mathcal{F}^\bullet$ is a preferential filter.

Proof. We show each in turn:

***W* is finite.** This holds because our BFNN is finite.

***f* is closed under finite intersection.** Suppose $X_1,\ldots,X_n \in f(w)$. By definition of f, $w \notin \bigcup_i \mathsf{Reach}(X_i^{\mathsf{C}})$ for all i. Since Reach is monotonic, [Make this a lemma!] we have $\bigcup_i \mathsf{Reach}(X_i^{\mathsf{C}}) = \mathsf{Reach}(\bigcup_i X_i^{\mathsf{C}}) = \mathsf{Reach}((\bigcap_i X_i)^{\mathsf{C}})$. So $w \notin \mathsf{Reach}((\bigcap_i X_i)^{\mathsf{C}})$. But this means that $\bigcap_i X_i \in f(w)$.

***f* is closed under superset.** Suppose $X \in f(w), X \subseteq Y$. By definition of f, $w \notin \mathsf{Reach}(X^{\mathsf{C}})$. Note that $Y^{\mathsf{C}} \subseteq X^{\mathsf{C}}$, and so by monotonicity of Reach we have $w \notin \mathsf{Reach}(Y^{\mathsf{C}})$. But this means $Y \in f(w)$, so we are done.

***f* contains the unit.** Note that for all $w \in W$, $w \notin \mathsf{Reach}(\emptyset) = \mathsf{Reach}(W^{\mathsf{C}})$. So $W \in f(w)$.

***g* contains the unit.** Same as the proof for f, except that we use the fact that for all w, $w \notin \mathsf{Prop}(\emptyset)$

***f* is acyclic.** Suppose $u_1, \ldots, u_n \in W$, with $u_1 \in \cap f(u_2), \ldots, u_{n-1} \in \cap f(u_n), u_n \in \cap f(u_1)$. That is, each $u_i \in \bigcap_{X \in f(u_{i+1})} X$. By choice of f, each $u_i \in \bigcap_{u_{i+1} \notin \mathsf{Reach}(X^{\mathsf{C}})} X$. Substituting X^{C} for X we get $u_i \in \bigcap_{u_{i+1} \notin \mathsf{Reach}(X)} X^{\mathsf{C}}$. In other words, $u_1 \in \mathsf{Reach}^{-1}(u_2), \ldots, u_{n-1} \in \mathsf{Reach}^{-1}(n), u_n \in \mathsf{Reach}^{-1}(u_1)$. [Update!] By Proposition ?, each $u_i = u_j$.

***f* is reflexive.** We want to show that $w \in \cap f(w)$. Well, suppose $X \in f(w)$, i.e. $w \notin \mathsf{Reach}(X^{\mathsf{C}})$ (by definition of f). Since for all S, $S \subseteq \mathsf{Reach}(S)$, we have $w \notin X^{\mathsf{C}}$. But this means $w \in X$, and we are done.

***g* is reflexive.** Same as the proof for f, except we use the fact that for all S, $S \subseteq \mathsf{Prop}(S)$.

***f* is transitive.** Suppose $X \in f(w)$, i.e. $w \notin \mathsf{Reach}(X^{\mathsf{C}})$. Well,

$$\begin{array}{lll} \mathsf{Reach}(X^{\mathsf{C}}) & = & \mathsf{Reach}(\mathsf{Reach}(X^{\mathsf{C}})) \quad \text{(by Idempotence of Reach)} \\ & = & \mathsf{Reach}(\{u \mid u \in \mathsf{Reach}(X^{\mathsf{C}})\}) \\ & = & \mathsf{Reach}(\{u \mid u \notin \mathsf{Reach}(X^{\mathsf{C}})\}^{\mathsf{C}}) \\ & = & \mathsf{Reach}(\{u \mid X \in f(u)\}^{\mathsf{C}}) \quad \text{(by definition of } f\text{)} \end{array}$$

So by definition of f, $\{u \mid X \in f(u)\} \in f(w)$.

***g* is transitive.** Same as the proof for f, except we use the fact that Prop is idempotent.

***g* contains *f*.** Suppose $X \in f(w)$, i.e. $w \notin \mathsf{Reach}(X^{\mathsf{C}})$. Since for all S, $\mathsf{Prop}(S) \subseteq \mathsf{Reach}(S)$, we have $w \notin \mathsf{Prop}(X^{\mathsf{C}})$. And so $X \in g(w)$, and we are done.

□

4 From Frames to Nets

This is the harder ("completeness") direction!

DEFINITION 4.1. Suppose we have net $\mathcal{N}$ and node $n \in N$ with incoming nodes $m_1, \ldots, m_k$, $(m_i, n) \in E$. Let $\mathsf{hash}\colon \mathcal{P}(\{m_1, \ldots, m_k\}) \times \mathbb{N}^k \to \mathbb{N}$ be defined by

$$\mathsf{hash}(S, \vec{w}) = \prod_{m_i \in S} w_i$$

PROPOSITION 4.2. $\mathsf{hash}(S, \vec{W}(m_i, n)) \colon \mathcal{P}(\{m_1, \ldots, m_k\}) \to P_k$, where

$$P_k = \{n \in \mathbb{N} \mid n \text{ is the product of some subset of primes } \{p_1, \ldots, p_k\}\}$$

is bijective (and so has a well-defined inverse hash^{-1}).

DEFINITION 4.3. Let $\mathcal{M}$ be a model based on preferential filter $\mathcal{F} = \langle W, f, g \rangle$. Its **simulation net** $\mathcal{N}^\bullet = \langle N, E, W, A, O, V \rangle$ is the BFNN given by:

- $N = W$
- $(u, v) \in E$ iff $u \in \cap f(v)$

Now let $m_1, \ldots, m_k$ list those nodes such that $(m_i, n) \in E$.

- $W(m_i, n) = p_i$, the ith prime number.
- $A^{(n)}(\vec{x}, \vec{w}) = \mathsf{hash}(\{m_i | x_i = 1\}, \vec{w})$
- $O^{(w)}(x) = 1$ iff $(\mathsf{hash}^{-1}(x)[0])^{\mathsf{C}} \notin g(n)$
- $V_{\mathcal{N}^\bullet}(p) = V_{\mathcal{M}}(p)$

LEMMA 4.4. Let $\vec{m} = m_1, \ldots, m_k$ be those nodes such that $(m_i, n) \in E$. Then

$$\mathsf{Activates}_S(\vec{m}, n) = 1 \quad \text{iff} \quad \{m_i | m_i \in S\}^{\mathsf{C}} \notin g(n)$$

Proof. $\mathsf{Activates}_S(\vec{m}, n) = 1$ iff:

$$\begin{array}{ll} & O^{(n)}(A^{(n)}((\chi_S(m_1), \ldots, \chi_S(m_k)); (W(m_1, n), \ldots, W(m_k, n)))) = 1 \\ \text{iff} & \mathsf{hash}^{-1}(\mathsf{hash}(\{m_i | m_i \in S\}; (W(m_1, n), \ldots, W(m_k, n)))[0])^{\mathsf{C}} \notin g(n) \\ \text{iff} & \{m_i | m_i \in S\}^{\mathsf{C}} \notin g(n) \end{array}$$

□

CLAIM 4.5. $\mathcal{N}^\bullet$ is a BFNN.

Proof. Clearly $\mathcal{N}^\bullet$ is a binary ANN. We check the rest of the conditions:

$\mathcal{N}^\bullet$ is feed-forward. Suppose for contradiction that E contains a cycle, i.e. distinct $u_1, \ldots, u_n \in N$ such that $u_1 E u_2, \ldots, u_{n-1} E u_n, u_n E u_1$. Then we have $u_1 \in \cap f(u_2), \ldots, u_{n-1} \in \cap f(u_{n-1}), u_n \in \cap f(u_1)$, which contradicts the fact that f is acyclic.

$O^{(n)} \circ A^{(n)}$ is zero at zero. Suppose for contradiction that $O^{(v)}(A^{(v)}(\vec{0}, \vec{w})) = 1$. Then $(\mathsf{hash}^{-1}(\mathsf{hash}(\emptyset)))^{\mathsf{C}} = \emptyset^{\mathsf{C}} = W \notin g(v)$, which contradicts the fact that f contains the unit.

$O^{(n)} \circ A^{(n)}$ is monotonically increasing. Let $\vec{w}_1, \vec{w}_2$ be such that hash is well-defined (i.e. are vectors of prime numbers), and suppose $\vec{w}_1 < \vec{w}_2$. If $O^{(v)}(A^{(v)}(\vec{x}, \vec{w_1})) = 1$, then $(\mathsf{hash}^{-1}(\mathsf{hash}(\vec{x}, \vec{w_1}))[0])^{\mathsf{C}} \notin g(v)$. But this just means $\{m_i | x_i = 1\}^{\mathsf{C}} \notin g(v)$. And so $(\mathsf{hash}^{-1}(\mathsf{hash}(\vec{x}, \vec{w_2}))[0])^{\mathsf{C}} \notin g(v)$. But then $O^{(n)}(A^{(n)}(\vec{x}, \vec{w_2})) = 1$.

The main point here is just that $\vec{w_1}$ and $\vec{w_2}$ are just indexing the set in question, and their actual values don't affect the final output (we don't need the $\vec{w}_1 < \vec{w}_2$ hypothesis!). The real work happens within $g(v)$. □

LEMMA 4.6. $\mathsf{Reach}_{\mathcal{N}^\bullet}(S) = \{n | S^{\mathsf{C}} \notin f(n)\}$

Proof. For the $(\supseteq)$ direction, let $n \in N$ be such that $S^{\mathsf{C}} \notin f(n)$. By Proposition 2.5 and the fact that $\langle W, f \rangle$ forms a proper filter, $S \in f(n)$. By the definition of core, $\cap f(n) \subseteq S$. f is reflexive, which means in particular that $n \in \cap f(n) \subseteq S$. By the base case of Reach, we have $n \in \mathsf{Reach}_{\mathcal{N}^\bullet}(S)$.

Now for the ($\subseteq$) direction. Suppose $n \in \mathsf{Reach}(S)$, and proceed by induction on Reach.

Base step. $n \in S$. Suppose for contradiction that $S^{\mathsf{C}} \in f(n)$. By definition of core, $\cap f(n) \subseteq S^{\mathsf{C}}$. But since $\mathcal{F}$ is reflexive, $n \in \cap f(n)$. So $n \in S^{\mathsf{C}}$, which contradicts $n \in S$.

Inductive step. There is $m \in \mathsf{Reach}_{\mathcal{N}^\bullet}(S)$ such that $(m,n) \in E$ (and so $m \in \cap f(n)$). By inductive hypothesis, $S^{\mathsf{C}} \notin f(m)$. Now suppose for contradiction that $S^{\mathsf{C}} \in f(n)$. Since f is transitive, $\{t \mid S^{\mathsf{C}} \in f(t)\} \in f(n)$. By definition of core, $\cap f(n) \subseteq \{t \mid S^{\mathsf{C}} \in f(t)\}$. Since $m \in \cap f(n)$, $S^{\mathsf{C}} \in f(m)$. But this contradicts $S^{\mathsf{C}} \notin f(m)$! $\square$

LEMMA 4.7. $\mathsf{Prop}_{\mathcal{N}^\bullet}(S) = \{n \mid S^{\mathsf{C}} \notin g(n)\}$

Proof. For the ($\supseteq$) direction, let $n \in N$, suppose $S^{\mathsf{C}} \notin g(n)$. Since g contains f, $S^{\mathsf{C}} \notin f(n)$. By Proposition 2.5 and the fact that $\langle W, f\rangle$ forms a proper filter, $S \in f(n)$. By the definition of core, $\cap f(n) \subseteq S$. f is reflexive, which means in particular that $n \in \cap f(n) \subseteq S$. By the base case of Prop, $n \in \mathsf{Prop}_{\mathcal{N}^\bullet}(S)$.

As for the ($\subseteq$) direction, suppose $n \in \mathsf{Prop}_{\mathcal{N}^\bullet}(S)$, and proceed by induction on Prop.

Base step. $n \in S$. Suppose for contradiction that $S^{\mathsf{C}} \in g(n)$. Since $\mathcal{G}$ is reflexive, $n \in \cap g(n)$. By definition of core, we have $\cap g(n) \subseteq S^{\mathsf{C}}$. But then $n \in \cap g(n) \subseteq S^{\mathsf{C}}$, i.e. $n \in S^{\mathsf{C}}$, which contradicts $n \in S$.

Inductive step. Let $\vec{m} = m_1, \ldots, m_k$ list those nodes such that $(u_i, v) \in E$. We have

$$\mathsf{Activates}_{\mathsf{Prop}_{\mathcal{N}^\bullet}(S)}(\vec{m}, n) = 1$$

By Lemma 4.4, this means that $\{m_i \mid m_i \in \mathsf{Prop}_{\mathcal{N}^\bullet}(S)\}^{\mathsf{C}} \notin g(n)$. But by our inductive hypothesis, $\{m_i \mid m_i \in \mathsf{Prop}_{\mathcal{N}^\bullet}(S)\} = \{m_i \mid S^{\mathsf{C}} \notin g(n)\}$. For convenience, let T be this latter set, i.e. $T = \{m_i \mid S^{\mathsf{C}} \notin g(n)\}$. So we have $T^{\mathsf{C}} \notin g(n)$.

We would like to show that $S^{\mathsf{C}} \notin g(n)$. Suppose for contradiction that $S^{\mathsf{C}} \in g(n)$. Notice that, by definition of T, $T^{\mathsf{C}} = \{u_i \mid S^{\mathsf{C}} \in g(u_i)\}$. Since $S^{\mathsf{C}} \in g(v)$ and $\mathcal{G}$ is transitive, $T^{\mathsf{C}} \in g(v)$, which contradicts $T^{\mathsf{C}} \notin g(v)$.

$\square$

THEOREM 4.8. Let $\mathcal{M}$ be a model based on a preferential filter $\mathcal{F}$, and let $\mathcal{N}^\bullet$ be the corresponding simulation net. We have, for all $w \in W$,

$$\mathcal{M}, w \Vdash \varphi \quad \text{iff} \quad \mathcal{N}^\bullet, w \Vdash \varphi$$

Proof. By induction on φ. Again, the propositional, $\neg\varphi$, and $\varphi \wedge \psi$ cases are trivial.

$\langle \mathbf{K} \rangle \varphi$ case:

$$\begin{array}{rlll} \mathcal{M}, w \Vdash \langle \mathbf{K} \rangle \varphi & \text{iff} & \{u \mid \mathcal{M}, w \nVdash \varphi\} \notin f(w) & \text{(by definition)} \\ & \text{iff} & \{u \mid u \notin [\![\varphi]\!]_{\mathcal{N}^\bullet}\} \notin f(w) & \text{(Inductive Hypothesis)} \\ & \text{iff} & [\![\varphi]\!]^{\mathsf{C}}_{\mathcal{N}^\bullet} \notin g(w) & \\ & \text{iff} & w \in \mathsf{Reach}_{\mathcal{N}^\bullet}([\![\varphi]\!]) & \text{(by Lemma 4.6)} \\ & \text{iff} & w \in [\![\langle \mathbf{K} \rangle \varphi]\!]_{\mathcal{N}^\bullet} & \text{(by definition)} \\ & \text{iff} & \mathcal{N}^\bullet, w \Vdash \langle \mathbf{K} \rangle \varphi & \text{(by definition)} \end{array}$$

$\langle \mathbf{T} \rangle \varphi$ case:

$$\begin{array}{rlll} \mathcal{M}, w \Vdash \langle \mathbf{T} \rangle \varphi & \text{iff} & \{u \mid \mathcal{M}, u \nVdash \varphi\} \notin g(w) & \text{(by definition)} \\ & \text{iff} & \{u \mid u \notin [\![\varphi]\!]_{\mathcal{N}^\bullet}\} \notin g(w) & \text{(Inductive Hypothesis)} \\ & \text{iff} & [\![\varphi]\!]^{\mathsf{C}}_{\mathcal{N}^\bullet} \notin g(w) & \\ & \text{iff} & w \in \mathsf{Prop}_{\mathcal{N}^\bullet}([\![\varphi]\!]) & \text{(by Lemma 4.7)} \\ & \text{iff} & w \in [\![\langle \mathbf{T} \rangle \varphi]\!]_{\mathcal{N}^\bullet} & \text{(by definition)} \\ & \text{iff} & \mathcal{N}^\bullet, w \Vdash \langle \mathbf{T} \rangle \varphi & \text{(by definition)} \end{array}$$

$\square$

COROLLARY 4.9. $\mathcal{M} \models \varphi$ iff $\mathcal{N}^\bullet \models \varphi$.

5 Completeness

5.1 The Base Modal Logic

DEFINITION 5.1. Our logic $\mathbf{L}$ is the smallest set of formulas in $\mathcal{L}$ containing the axioms

(K). $\mathbf{K}(\varphi \to \psi) \to (\mathbf{K}\varphi \to \mathbf{K}\psi)$
($\mathbf{T_K}$). $\mathbf{K}\varphi \to \varphi$
($\mathbf{4_K}$). $\mathbf{K}\varphi \to \mathbf{KK}\varphi$
(Grz). $\mathbf{K}(\mathbf{K}(\varphi \to \mathbf{K}\varphi) \to \varphi) \to \varphi$

($\mathbf{T_T}$). $\mathbf{T}\varphi \to \varphi$
($\mathbf{4_T}$). $\mathbf{T}\varphi \to \mathbf{TT}\varphi$
(K-T). $\mathbf{K}\varphi \to \mathbf{T}\varphi$

that is closed under:

(Necessitation). If $\varphi \in \mathbf{L}$ then $\Box\varphi \in \mathbf{L}$ for $\Box \in \{\mathbf{K}, \mathbf{T}\}$

DEFINITION 5.2. [19, Definition 2.30] **(Deduction for L)** $\vdash \varphi$ iff either φ is an axiom, or φ follows from some previously obtained formula by one of the inference rules. If $\Gamma \subseteq \mathcal{L}$ is a set of formulas, $\Gamma \vdash \varphi$ whenever there are finitely many $\psi_1, \ldots, \psi_k \in \Gamma$ such that $\vdash \psi_1 \wedge \ldots \wedge \psi_k \to \varphi$.

DEFINITION 5.3. [19, Definition 2.36] Γ is **consistent** iff $\Gamma \nvdash \bot$. Γ is **maximally consistent** if Γ is consistent and for all $\varphi \in \mathcal{L}$ either $\varphi \in \Gamma$ or $\varphi \notin \Gamma$.

LEMMA 5.4. [19, Lemma 2.19] ("Lindenbaum's Lemma") We can extend any set Γ to a maximally consistent set $\Delta \supseteq \Gamma$.

DEFINITION 5.5. [19, Definition 2.36] **(Proof Set)** $|\varphi|_{\mathbf{L}} = \{\Delta \mid \Delta \text{ is maximally consistent and } \varphi \in \Delta\}$

PROPOSITION 5.6. Let Δ be maximally consistent, and let $\Box \in \{\mathbf{K}, \mathbf{T}\}$. We have $\Box\varphi \in \Delta$ iff

$$\forall \Sigma \text{ maximally consistent, if } \forall \psi, \Box\psi \in \Delta \text{ implies } \psi \in \Sigma, \text{ then } \varphi \in \Sigma$$

Proof. The $(\to)$ direction is straightforward. As for the $(\leftarrow)$ direction, suppose contrapositively that $\Box\varphi \notin \Delta$, and let $\Sigma = \{\psi \mid \Box\psi \in \Delta\}$ [why is Σ maximally consistent?]. Then by construction, for all ψ $\Box\psi \in \Delta$ implies $\psi \in \Sigma$, but $\varphi \notin \Sigma$ (since $\Box\varphi \notin \Delta$). □

5.2 Soundness

THEOREM 5.7. **(Soundness)** If $\Gamma \vdash \varphi$ then $\Gamma \models_{\mathrm{BFNN}} \varphi$

Proof. Suppose $\Gamma \vdash \varphi$, and let $\mathcal{N}, n \models \Gamma$ We just need to check that each of the axioms and rules of inference are sound, from which we can conclude that $\mathcal{N}, n \models \varphi$. We can do this either by the semantics of BFNNs, or instead by checking them in an equivalent preferential frame $\mathcal{M}^{\bullet} = \langle W, f, g, V\rangle$:

To show soundness of:	Use:	Alternative:
(K)	Monotonicity of Reach	$\langle W, f\rangle$ forms a filter
($\mathrm{T_K}$)	Inclusion of Reach	Reflexivity of f
($\mathrm{4_K}$)	Idempotence of Reach	Transitivity of f
(Grz)	Proposition ?[Check! – and update, since the def changed]	f is acyclic [Check!]
($\mathrm{T_T}$)	Inclusion of Prop	Reflexivity of g
($\mathrm{4_T}$)	Idempotence of Prop	Transitivity of g
($\mathbf{T{-}K}$)	Reach contains Prop	g contains f
(Necessitation)	$\forall w, w \notin \mathsf{Reach}(\emptyset), \mathsf{Prop}(\emptyset)$	f, g contain the unit

□

5.3 Model Building

Given a set $\Gamma \subseteq \mathcal{L}$, I will show that we can build a net $\mathcal{N}$ that models Γ. Since preferential filters are equivalent to BFNNs (over $\mathcal{L}$), I will focus instead on building a preferential filter $\mathcal{F}$. This is the same strategy taken by [14], who constructs KLM cumulative-ordered models in order to build a neural net.

The following are the standard canonical construction and facts for neighborhood models (see Eric Pacuit's book). Adapting these to our logic of $\mathbf{K}, \mathbf{K}^{\downarrow}, \mathbf{T}$ is a straightforward exercise in modal logic.

LEMMA 5.8. [19, Lemma 2.12 & Definition 2.37] We can build a **canonical** neighborhood model for $\mathbf{L}$, i.e. a model $\mathcal{M}^C = \langle W^C, f^C, g^C, V^C \rangle$ such that:

- $W^C = \{\Delta | \, \Delta \text{ is maximally consistent}\}$
- For each $\Delta \in W^C$ and each $\varphi \in \mathcal{L}$, $|\varphi|_{\mathbf{L}} \in f^C(\Delta)$ iff $\mathbf{K}\varphi \in \Delta$
- For each $\Delta \in W^C$ and each $\varphi \in \mathcal{L}$, $|\varphi|_{\mathbf{L}} \in g^C(\Delta)$ iff $\mathbf{T}\varphi \in \Delta$
- $V^C(p) = |p|_{\mathbf{L}}$

Note. This is where the Necessitation rules come into play — we need them in order to guarantee that we can actually build this model!

LEMMA 5.9. [19, Lemma 2.13] **(Truth Lemma)** We have, for canonical model $\mathcal{M}^C$,

$$\{\Delta | \, \mathcal{M}^C, \Delta \Vdash \varphi\} = |\varphi|_{\mathbf{L}}$$

Proof. By induction on φ. The propositional, and boolean cases are straightforward.

K case.

$$\begin{array}{llll} \mathcal{M}^C, \Delta \Vdash \mathbf{K}\varphi & \text{iff} & \{u | \, \mathcal{M}^C, \Sigma \Vdash \varphi\} \in f^C(\Delta) & \text{(by definition)} \\ & \text{iff} & |\varphi|_{\mathbf{L}} \in f^C(\Delta) & \text{(by IH)} \\ & \text{iff} & \mathbf{K}\varphi \in \Delta & \text{(since } \mathcal{M}^C \text{ is canonical)} \\ & \text{iff} & \Delta \in |\mathbf{K}\varphi|_{\mathbf{L}} & \text{(by definition)} \end{array}$$

T case.

$$\begin{array}{llll} \mathcal{M}^C, \Delta \Vdash \mathbf{T}\varphi & \text{iff} & \{u | \, \mathcal{M}^C, \Sigma \Vdash \varphi\} \in g^C(\Delta) & \text{(by definition)} \\ & \text{iff} & |\varphi|_{\mathbf{L}} \in g^C(\Delta) & \text{(by IH)} \\ & \text{iff} & \mathbf{T}\varphi \in \Delta & \text{(since } \mathcal{M}^C \text{ is canonical)} \\ & \text{iff} & \Delta \in |\mathbf{T}\varphi|_{\mathbf{L}} & \text{(by definition)} \end{array}$$

□

THEOREM 5.10. [State that our logic has the finite model property]

Proof. [Prove it by the usual filtration construction — the fact that the filtration is closed under $\cap, \subseteq$, reflexive, and transitive are all shown in Pacuit's book. So I just need to show that the same is true of the acyclic & skeleton properties.] □

PROPOSITION 5.11. If $\mathcal{M}$ is finite and satisfies the Truth Lemma, then $\mathcal{M}$ is a preferential filter.

Proof. W^C is finite by assumption. Since $\mathbf{L}$ contains all instances of **(K)**, **(T)**, **(4)**, **(T)**, **(4)** it follows that f^C is a reflexive, transitive, proper filter, and g^C is reflexive and transitive (this is another classical result, see Pacuit's book). The only things left to show are that f^C is acyclic and f^C is the skeleton of g^C.

W^C is finite. Holds by assumption.

f^C is closed under finite intersection. It's enough to show that f^C is closed under binary intersections. $\mathbf{L}$ contains all instances of **(K)**, from which we can derive all instances of:

$$\textbf{(C)} \quad \mathbf{K}\varphi \wedge \mathbf{K}\psi \rightarrow \mathbf{K}(\varphi \wedge \psi)$$

Suppose $|\varphi|_{\mathbf{L}}, |\psi|_{\mathbf{L}} \in f^C(\Delta)$. By definition of f^C, $\mathbf{K}\varphi \in \Delta$ and $\mathbf{K}\psi \in \Delta$. So $\mathbf{K}\varphi \wedge \mathbf{K}\psi \in \Delta$. Applying **(C)**, $\mathbf{K}(\varphi \wedge \psi) \in \Delta$. So $|\varphi \wedge \psi|_{\mathbf{L}} = |\varphi|_{\mathbf{L}} \cap |\psi|_{\mathbf{L}} \in \Delta$.

f^C is closed under superset. $\mathbf{L}$ contains all instances of **(K)** and the necessitation rule, from which we can derive:

$$\textbf{(RM)} \quad \text{If } \varphi \to \psi \in \mathbf{L} \text{ then } \mathbf{K}\varphi \to \mathbf{K}\psi \in \mathbf{L}$$

Suppose $|\varphi|_{\mathbf{L}} \in f^C(\Delta)$, and $|\varphi|_{\mathbf{L}} \subseteq |\psi|_{\mathbf{L}}$. The former fact gives us $\mathbf{K}\varphi \in \Delta$. The latter gives us, for all maximally consistent Δ, if $\varphi \in \Delta$ then $\psi \in \Delta$, i.e. $\varphi \to \psi \in \mathbf{L}$ [Is this correct? Probably not; we need to close the canonical model under superset]. By **(RM)**, we have $\mathbf{K}\psi \in \Delta$, i.e. $|\psi|_{\mathbf{L}} \in f^C(\Delta)$.

f^C contains the unit. $\mathbf{L}$ is closed under necessitation for $\mathbf{K}$, from which we can derive:

$$\textbf{(N)} \quad \mathbf{K}\top$$

That is, $\mathbf{K}\top \in \Delta$ for all maximally consistent Δ. So $|\top|_{\mathbf{L}} \in f^C(\Delta)$, i.e. $W^C \in f^C(\Delta)$.

f^C is reflexive. First, let $\Delta \in W^C$, and suppose $|\varphi|_{\mathbf{L}} \in f^C(\Delta)$. By definition of f^C, $\mathbf{K}\varphi \in \Delta$. By $(\mathbf{T_K})$, $\varphi \in \Delta$. Since φ was chosen arbitrarily, we have for all φ, if $|\varphi|_{\mathbf{L}} \in f^C(\Delta)$ then $\varphi \in \Delta$. In other words, $\Delta \in \bigcap_{|\varphi|_{\mathbf{L}} \in f^C(\Delta)} |\varphi|_{\mathbf{L}} = \cap f^C(\Delta)$.

f^C is transitive. Suppose $|\varphi|_{\mathbf{L}} \in f^C(\Delta)$. By definition of f^C, $\mathbf{K}\varphi \in \Delta$. By the $(\mathbf{4_K})$ axiom, $\mathbf{K}\mathbf{K}\varphi \in \Delta$. But this means that $|\mathbf{K}\varphi|_{\mathbf{L}} \in f^C(\Delta)$. By definition of proof set, we have $\{\Sigma | \mathbf{K}\varphi \in \Sigma\} \in f^C(\Delta)$. That is, $\{\Sigma \mid |\varphi|_{\mathbf{L}} \in f^C(\Sigma)\} \in f^C(\Delta)$, and we are done.

f^C is acyclic. [Update! – no more nominals, acyclic rule has changed to Grz!] Since f^C is reflexive and transitive, by Proposition 2.7 it's enough to show that f^C is asymmetric. Suppose $\Delta_1 \in \cap f^C(\Delta_2)$ and $\Delta_2 \in \cap f^C(\Delta_1)$. By definition of core, $\Delta_1 \in \bigcap_{|\varphi|_{\mathbf{L}} \in f^C(\Delta_2)} |\varphi|_{\mathbf{L}}$ and $\Delta_2 \in \bigcap_{|\varphi|_{\mathbf{L}} \in f^C(\Delta_1)} |\varphi|_{\mathbf{L}}$, i.e. we have both of the following:

1. $\forall \varphi$, if $\mathbf{K}\varphi \in \Delta_2$ then $\varphi \in \Delta_1$
2. $\forall \varphi$, if $\mathbf{K}\varphi \in \Delta_1$ then $\varphi \in \Delta_2$

We want to show that $\Delta_1 = \Delta_2$. I'll show the $(\subseteq)$ direction (the other direction is similar). Suppose for contradiction that $\varphi \in \Delta_1$, but $\varphi \notin \Delta_2$ (i.e. $\neg\varphi \in \Delta_2$).

Since Δ_1 is named, some $i \in \Delta_1$. By **(Antisym)**, $\mathbf{K}(\langle \mathbf{K} \rangle i \to i) \in \Delta_1$. By (2), $\langle \mathbf{K} \rangle i \to i \in \Delta_2$. Rewriting, we get $\neg i \to \mathbf{K}\neg i \in \Delta_2$. [What next?]

g^C contains the unit. Similar to the proof for f^C, but apply necessitation for $\mathbf{T}$ instead of $\mathbf{K}$.

g^C is reflexive. Similar to the proof for f^C, but apply $(\mathbf{T_T})$ instead of $(\mathbf{T_K})$.

g^C is transitive. Similar to the proof for f^C, but apply $(\mathbf{4_T})$ instead of $(\mathbf{4_K})$.

g^C contains f^C. Suppose $|\varphi|_{\mathbf{L}} \in f^C(\Delta)$. By definition of f^C, $\mathbf{K}\varphi \in \Delta$. By the $(\mathbf{K{-}T})$ axiom, $\mathbf{T}\varphi \in \Delta$. And so $|\varphi|_{\mathbf{L}} \in f^C(\Delta)$.

□

THEOREM 5.12. **(Model Building)** Given any consistent $\Gamma \subseteq \mathcal{L}$, we can construct a BFNN $\mathcal{N}$ and neuron $n \in N$ such that $\mathcal{N}, n \models \Gamma$.

Proof. Extend Γ to maximally consistent Δ using Lemma 5.4. Let $\mathcal{M}^C$ be a canonical model for $\mathbf{L}$ guaranteed by Lemma 5.8. By the Truth Lemma (Lemma 5.9), $\mathcal{M}^C, \Delta \models \Delta$. So in particular, $\mathcal{M}^C, \Delta \models \Gamma$.

By the Finite Model Property (Lemma 5.10), we can construct a finite model $\mathcal{M}'$ satisfying exactly the same formulas at all worlds. By Proposition 5.11, $\mathcal{M}'$ is a preferential filter.

From here, we can build our net $\mathcal{N}^\bullet$ as before, satisfying exactly the same formulas as $\mathcal{M}$ at all neurons (by Theorem 4.8). And so $\mathcal{N}^\bullet, \Delta \models \Gamma$. □

THEOREM 5.13. **(Completeness)** For all consistent $\Gamma \subseteq \mathcal{L}$, if $\Gamma \models_{\text{BFNN}} \varphi$ then $\Gamma \vdash \varphi$

Proof. Suppose contrapositively that $\Gamma \nvdash \varphi$. This means that $\Gamma \cup \{\neg\varphi\}$ is consistent, i.e. by Theorem 5.12 we can build a BFNN $\mathcal{N}$ and neuron n such that $\mathcal{N}, n \models \Gamma \cup \{\neg\varphi\}$. In particular, $\mathcal{N}, n \nvDash \varphi$. But then we must have $\Gamma \nvDash \varphi$. □

TODO:

- Double-check properties for canonical model & completeness
- Do filtration/finite model property
- Get bound on the size of the finite model.
- Think about complexity of decidability of the logic (but only if it seems easy)
- Copy-paste flipping $\wedge, \vee, \neg$ considerations
- Write up fuzzy network considerations (in a crisp (non-fuzzy) language) — fuzzy nets satisfy *exactly* the same crisp formulas as binary nets
- Make drawings in Tikz
- Make corrections Saul gave
- Close the canonical model under superset
- Put the page number/theorem number for each result
- Rename the axioms to something more readable (($\mathbf{T_T}$) is confusing as hell)

References

[1] Sebastian Bader and Pascal Hitzler. Dimensions of neural-symbolic integration-a structured survey. *ArXiv preprint cs/0511042*, 2005.

[2] Christian Balkenius and Peter Gärdenfors. Nonmonotonic Inferences in Neural Networks. In *KR*, pages 32–39. 1991.

[3] Vaishak Belle. Logic Meets Learning: From Aristotle to Neural Networks. In *Neuro-Symbolic Artificial Intelligence: The State of the Art*, pages 78–102. IOS Press, 2021.

[4] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal logic: graph. Darst*, volume 53. Cambridge University Press, 2001.

[5] Reinhard Blutner. Nonmonotonic inferences and neural networks. In *Information, Interaction and Agency*, pages 203–234. Springer, 2004.

[6] Antony Browne and Ron Sun. Connectionist inference models. *Neural Networks*, 14(10):1331–1355, 2001.

[7] Dov M Gabbay, Ian Hodkinson, and Mark A Reynolds. Temporal logic: mathematical foundations and computational aspects. 1994.

[8] Artur S d'Avila Garcez, Krysia Broda, and Dov M Gabbay. Symbolic knowledge extraction from trained neural networks: a sound approach. *Artificial Intelligence*, 125(1-2):155–207, 2001.

[9] Artur S d'Avila Garcez, Luis C Lamb, and Dov M Gabbay. *Neural-symbolic cognitive reasoning*. Springer Science Business Media , 2008.

[10] Laura Giordano, Valentina Gliozzi, and Daniele Theseider Dupré. From common sense reasoning to neural network models through multiple preferences: An overview. *CoRR*, abs/2107.04870, 2021.

[11] Laura Giordano, Valentina Gliozzi, and Daniele Theseider DuprÉ. A conditional, a fuzzy and a probabilistic interpretation of self-organizing maps. *Journal of Logic and Computation*, 32(2):178–205, 2022.

[12] The Third AI Summer, AAAI Robert S. Engelmore Memorial Award Lecture. AAAI, 2020.

[13] Caleb Kisby, Saúl Blanco, and Lawrence Moss. The logic of hebbian learning. In *The International FLAIRS Conference Proceedings*, volume 35. 2022.

[14] Hannes Leitgeb. Nonmonotonic reasoning by inhibition nets. *Artificial Intelligence*, 128(1-2):161–201, 2001.

[15] Hannes Leitgeb. Nonmonotonic reasoning by inhibition nets II. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(supp02):105–135, 2003.

[16] Hannes Leitgeb. Neural Network Models of Conditionals. In *Introduction to Formal Philosophy*, pages 147–176. Springer, 2018.

[17] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[18] Simon Odense and Artur d'Avila Garcez. A semantic framework for neural-symbolic computing. *ArXiv preprint arXiv:2212.12050*, 2022.

[19] Eric Pacuit. *Neighborhood semantics for modal logic*. Springer, 2017.

[20] Md Kamruzzaman Sarker, Lu Zhou, Aaron Eberhart, and Pascal Hitzler. Neuro-symbolic artificial intelligence: current trends. *ArXiv preprint arXiv:2105.05330*, 2021.