Notes on the Completeness of Neural Net Models in a Modal Language

Step 0. Find a paper I enjoy, and read it. Try to understand its ideas, with an eye towards extending it/altering it.

This paper is inspired by Hannes Leitgeb's Nonmonotonic Reasoning by Inhibition Nets, which proves completeness for the neuro-symbolic interface suggested by Balkenius and Gärdenfors' Nonmonotonic Inferences in Neural Networks.

Step 1. Look for an extension/open problem that makes me think "What the fuck? That's still open? No way, this shit is low-hanging fruit, free paper here I come." i.e. something *easy* and *straightforward*, *without complications*.

Hannes Leitgeb showed that feed-forward neural networks are complete with respect to certain conditional laws of $\Rightarrow$. But $\varphi \Rightarrow \psi$ just reads "$\psi \subseteq \mathsf{Prop}(\varphi)$" (i.e. ψ is in the propagation of the signal φ), which we can re-write in modal language as $\mathbf{T}\varphi \to \psi$. In the same way that Hannes shows that feed-forward nets and preferential-conditional models are equivalent, it shouldn't be too hard at all to show that feed-forwards nets and neighborhood models are equivalent. (Note that it is well-known that neighborhood models are a generalization of preferential models.)

I also think I should be able to throw in a $\mathbf{K}$ modality (graph-reachability) in there, almost for free.

Step 2. Follow-up question (only answer after Step 1): Is the extension *interesting* or *surprising*? What do we learn by extending the result?

Why bother with completeness?. In formal specifications (of AI agents, or otherwise), we're often content with just listing some sound rules or behaviors that the agent will always follow. And it's definitely cool to see that neural networks satisfy some sound logical axioms. But if we want to fundamentally bridge the gap between logic and neural networks, we should set our aim higher: Towards *complete* logical characterizations of neural networks.

A more practical reason: Completeness gives us model-building, i.e. given a specification Γ, we can *build* a neural network $\mathcal{N}$ satisfying Γ.

Why bother with this modal language?. Almost all of the previous work bridging logic and neural networks has focused on neural net models of *conditionals*. In some sense, doing this in modal language is just a re-write of this old work. But this previous work hasn't addressed how *learning* or *update* in neural networks can be cast in logical terms. This is not merely due to circumstance — integrating conditionals with update is a long-standing controversial issue. So instead, we believe that it is more natural to work with modalities (instead of conditionals), because

Modal language natively supports update.

In other words, our modal setting sets us up to easily cast update operators (e.g. neural network learning) as modal operators in our logic.

Also this gives me an excuse to title a paper *Neural Network Models à la Mode* :-) (This is a play on both modal logic and also bringing some old work back in style!)

And LOL I can name a section "Learning: The Cherry on Top"

Step 3. Two things to do at this point:

- **Make a new Texmacs file named “PAPERNAME-master-notes.tm”. Transcribe the key definitions, examples, lemmas, and results from the paper. This makes it easier to later copy-paste parts of proofs, and also ensures that I don't reinvent the wheel later (it's tempting to redefine everything yourself!)**
- **Go to `https://www.connectedpapers.com/` and download any major nearby papers. Upload the papers to paperless-ngx and make a point to read them (understanding context helps a lot!).**

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 $\mathcal{N} = \langle N, E, W, T, A, V\rangle$, 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 $\mathcal{N} = \langle N, E, W, T, A, V\rangle$ that is

- **Feed-forward**, i.e. E does not contain any cycles

- **Binary**, i.e. the output of each neuron is in $\{0, 1\}$
- $O^{(n)} \circ A^{(n)}$ is **zero at zero** in the first parameter, i.e.

$$O^{(n)}(A^{(n)}(\vec{0}, \vec{w})) = 0$$

- $O^{(n)} \circ A^{(n)}$ is **strictly monotonically increasing** in the second parameter, i.e. 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 $\mathcal{N}$, $\mathsf{Set} = \mathcal{P}(N) = \{S \mid S \subseteq N\}$

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

1.2 Prop and Reach

Definition 1.5. (Adapted from [14]) 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 $m_1, \ldots, m_k$ such that $(m_i, n) \in E$ we have

$$O^{(n)}(A^{(n)}(\vec{\chi}_{\mathsf{Prop}(S)}(m_i), \vec{W}(m_i, n))) = 1$$

Proposition 1.6. [14] Let $\mathcal{N} \in \mathsf{Net}$. For all $S, S_1, S_2 \in \mathsf{Set}$, Prop satisfies

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

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

(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 $i, j \in \{0, \ldots, n\}$

Definition 1.7. 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 1.8. Let $\mathcal{N} \in \mathsf{Net}$. For all $S, S_1, S_2 \in \mathsf{Set}$, $n, m \in N$, Reach satisfies

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

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

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

Definition 1.9. For all $n \in N$, $\mathsf{Reach}^{-1}(n) = \bigcap_{n \notin \mathsf{Reach}(X)} X^{\mathsf{C}}$

Proposition 1.10. For all $n \in N$, $\mathsf{Reach}^{-1}(n) = \{m \mid \text{there is an } E\text{-path from } m \text{ to } n\}$

Proposition 1.11. Reach^{-1} is acyclic in the following sense: For $n_1, \ldots, n_k \in N$, if

$$n_1 \in \mathsf{Reach}^{-1}(n_2), \ldots, n_{k-1} \in \mathsf{Reach}^{-1}(n_k), n_k \in \mathsf{Reach}^{-1}(n_1)$$

Then each $n_i = n_j$.

PROPOSITION 1.12. **(Minimal Cause)** For all $n \in N$, if $n \in \mathsf{Prop}(S)$ then $n \in \mathsf{Prop}(S \cap \mathsf{Reach}^{-1}(n))$

1.3 Neural Network Semantics

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

$$\varphi :: = i | p | \neg\varphi | \varphi \wedge \varphi | \mathbf{K}\varphi | \mathbf{K}^{\leftarrow} i | \mathbf{T}\varphi |$$

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

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

$$\boxed{\begin{array}{lll}
[\![i]\!] & = & V(i) \in \mathsf{Set} \\
[\![p]\!] & = & V(p) \in \mathsf{Set} \\
[\![\neg\varphi]\!] & = & [\![\varphi]\!]^{\mathsf{C}} \\
[\![\varphi \wedge \psi]\!] & = & [\![\varphi]\!] \cap [\![\psi]\!] \\
[\![\langle\mathbf{K}\rangle\varphi]\!] & = & \mathsf{Reach}([\![\varphi]\!]) \\
[\![\langle\mathbf{K}^{\leftarrow}\rangle\varphi]\!] & = & \{n | \exists m \in [\![\varphi]\!]^{\mathsf{C}} \text{ such that } n \in \mathsf{Reach}^{-1}(m)\} \\
[\![\langle\mathbf{T}\rangle\varphi]\!] & = & \mathsf{Prop}([\![\varphi]\!])
\end{array}}$$

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)** $\Gamma \models_{\text{BFNN}} \varphi$ 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] A **neighborhood frame** is a pair $\mathcal{F} = \langle W, f \rangle$, where W is a non-empty set of **worlds** and $f: W \rightarrow \mathcal{P}(\mathcal{P}(W))$ 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] 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 $f(w)$. We often abbreviate this by $\cap f(w)$.

DEFINITION 2.3. [19] 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 $W \in f(w)$
- f **does not contain the empty set**: $\emptyset \notin f(w)$

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

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

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
- 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 $\{u \mid X \in f(u)\} \in f(w)$ (similarly for g)
- f is the **skeleton** of g: for all $w \in W$, if $X \cup (\cap f(w))^{\mathsf{C}} \in g(w)$ then $X \in g(w)$.

2.2 Neighborhood Semantics

Definition 2.7. [19] 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 2.8. [19] 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:

$$\boxed{\begin{array}{lll}
\mathcal{M}, w \Vdash i & \text{iff} & V(i) = \{w\} \\
\mathcal{M}, w \Vdash p & \text{iff} & w \in V(p) \\
\mathcal{M}, w \Vdash \neg\varphi & \text{iff} & \mathcal{M}, w \nVdash \varphi \\
\mathcal{M}, w \Vdash \varphi \wedge \psi & \text{iff} & \mathcal{M}, w \Vdash \varphi \text{ and } \mathcal{M}, w \Vdash \psi \\
\mathcal{M}, w \Vdash \mathbf{K}\varphi & \text{iff} & \{u \mid \mathcal{M}, u \Vdash \varphi\} \in f(w) \\
\mathcal{M}, w \Vdash \mathbf{K}^{\leftarrow}\varphi & \text{iff} & \forall u \in W, \text{ if } w \in \cap f(u) \text{ then } \mathcal{M}, u \Vdash \varphi \\
\mathcal{M}, w \Vdash \mathbf{T}\varphi & \text{iff} & \{u \mid \mathcal{M}, u \Vdash \varphi\} \in g(w)
\end{array}}$$

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

$$\begin{array}{lll}
\mathcal{M}, w \Vdash \langle \mathbf{K} \rangle\varphi & \text{iff} & \{u \mid \mathcal{M}, u \nVdash \varphi\} \notin f(w) \\
\mathcal{M}, w \Vdash \langle \mathbf{K}^{\leftarrow} \rangle\varphi & \text{iff} & \exists u \in W \text{ such that } w \in \cap f(u) \text{ and } \mathcal{M}, u \nVdash \varphi \\
\mathcal{M}, w \Vdash \langle \mathbf{T} \rangle\varphi & \text{iff} & \{u \mid \mathcal{M}, u \nVdash \varphi\} \notin g(w)
\end{array}$$

Definition 2.9. [19] **(Truth in a model)** $\mathcal{M} \models \varphi$ iff $\mathcal{M}, w \Vdash \varphi$ for all $w \in W$.

Definition 2.10. [19] **(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:

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

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 nominal, propositional, $\neg\varphi$, and $\varphi \wedge \psi$ cases are trivial.

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

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

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

$$\begin{aligned}
\mathcal{M}^\bullet, w \Vdash \langle \mathbf{K}^{\leftarrow} \rangle \varphi \quad & \text{iff} \quad \exists u \text{ such that } w \in \cap f(u) \text{ and } \mathcal{M}^\bullet, u \nVdash \varphi && \text{(by definition)} \\
& \text{iff} \quad \exists u \text{ such that } w \in \cap f(u) \text{ and } u \notin [\![\varphi]\!] && \text{(IH)} \\
& \text{iff} \quad \exists u \in [\![\varphi]\!]^{\mathsf{C}} \text{ such that } w \in \bigcap_{X \in f(u)} X && \\
& \text{iff} \quad \exists u \in [\![\varphi]\!]^{\mathsf{C}} \text{ such that } w \in \bigcap_{u \notin \mathsf{Reach}(X^{\mathsf{C}})} X && \text{(by choice of } f\text{)} \\
& \text{iff} \quad \exists u \in [\![\varphi]\!]^{\mathsf{C}} \text{ such that } w \in \mathsf{Reach}^{-1}(u) && \\
& \text{iff} \quad \mathcal{N}, w \Vdash \langle \mathbf{K}^{\leftarrow} \rangle \varphi && \text{(by definition)}
\end{aligned}$$

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

$$\begin{aligned}
\mathcal{M}^\bullet, w \Vdash \langle \mathbf{T} \rangle \varphi \quad & \text{iff} \quad \{u \mid \mathcal{M}^\bullet, w \nVdash \varphi\} \notin g(w) && \text{(by definition)} \\
& \text{iff} \quad \{u \mid u \notin [\![\varphi]\!]\} \notin g(w) && \text{(IH)} \\
& \text{iff} \quad [\![\varphi]\!]^{\mathsf{C}} \notin g(w) && \\
& \text{iff} \quad w \in \mathsf{Prop}([\![(\varphi^{\mathsf{C}})^{\mathsf{C}}]\!]) && \text{(by choice of } g\text{)} \\
& \text{iff} \quad w \in \mathsf{Prop}([\![\varphi]\!]) && \\
& \text{iff} \quad w \in [\![\langle \mathbf{T} \rangle \varphi]\!] && \text{(by definition)} \\
& \text{iff} \quad \mathcal{N}, w \Vdash \langle \mathbf{T} \rangle \varphi && \text{(by definition)}
\end{aligned}$$

□

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 $W \in f(w)$.

***f* does not contain ∅.** Similarly, for all $w \in W$, $w \in \mathsf{Reach}(W) = \mathsf{Reach}(\emptyset^{\mathsf{C}})$. So $\emptyset \notin f(w)$.

***f* is acyclic.** Suppose $u_1, \dots, u_n \in W$, with $u_1 \in \cap f(u_2), \dots, u_{n-1} \in \cap f(u_n), u_n \in \cap f(u_1)$. 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, $u_1 \in \mathsf{Reach}^{-1}(u_2), \ldots, u_{n-1} \in \mathsf{Reach}^{-1}(n), u_n \in \mathsf{Reach}^{-1}(u_1)$. By Proposition 1.11, 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,

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

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

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

***f* is the skeleton of *g*.** Suppose $X \cup (\cap f(w))^{\mathsf{C}} \in g(w)$. By choice of g, $w \notin \mathsf{Prop}([X \cup (\cap f(w))^{\mathsf{C}}]^{\mathsf{C}})$. Distributing the outer complement, we have $w \notin \mathsf{Prop}(X^{\mathsf{C}} \cap (\cap f(w)))$, i.e. $w \notin \mathsf{Prop}(X^{\mathsf{C}} \cap (\bigcap_{Y \in f(w)} Y))$. By choice of f, $w \notin \mathsf{Prop}(X^{\mathsf{C}} \cap (\bigcap_{w \notin \mathsf{Reach}(Y^{\mathsf{C}})} Y))$. Substituting Y^{C} for Y, we get $w \notin \mathsf{Prop}(X^{\mathsf{C}} \cap (\bigcap_{w \notin \mathsf{Reach}(Y)} Y^{\mathsf{C}}))$. By definition of Reach^{-1}, $w \notin \mathsf{Prop}(X^{\mathsf{C}} \cap \mathsf{Reach}^{-1}(w))$. From (Minimal Cause), we conclude that $w \notin \mathsf{Prop}(X^{\mathsf{C}})$, i.e. $X \in g(w)$.

□

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 $\mathsf{hash}: \mathcal{P}(\{m_1, \dots, m_k\}) \times \mathbb{N}^k \to \mathbb{N}$ be defined by

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

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

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

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

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

Claim 4.4. $\mathcal{N}^\bullet$ 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 4.5. $\mathsf{Reach}_{\mathcal{N}^\bullet}(S) = \{v | S^{\mathsf{C}} \notin f(v)\}$

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

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

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

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

Lemma 4.6. $\mathsf{Prop}_{\mathcal{N}^\bullet}(S) = \{v | S^{\mathsf{C}} \notin g(v)\}$

Proof. For the $(\supseteq)$ direction, suppose $S^{\mathsf{C}} \notin g(v)$. Since f is the skeleton of g, we have $S^{\mathsf{C}} \cup (\cap f(v))^{\mathsf{C}} \notin g(v)$, i.e. $[S \cap (\cap f(v))]^{\mathsf{C}} \notin g(v)$. But $S \cap (\cap f(v)) = \{u | u \in S$ and $(u, v) \in E\} = \mathsf{hash}^{-1}(\mathsf{hash}(\overrightarrow{\chi}_{\mathsf{Prop}_{\mathcal{N}^\bullet}(S)}(u), \vec{W}(u, v)))[0]$, and so

$$(\mathsf{hash}^{-1}(\mathsf{hash}(\overrightarrow{\chi}_{\mathsf{Prop}_{\mathcal{N}^\bullet}(S)}(u), \vec{W}(u, v)))[0])^{\mathsf{C}} \notin g(v)$$

i.e. $O^{(v)}(A^{(v)}(\overrightarrow{\chi}_{\mathsf{Prop}_{\mathcal{N}^\bullet}(S)}(u), \vec{W}(u, v))) = 1$, and we conclude that $v \in \mathsf{Prop}_{\mathcal{N}^\bullet}(S)$.

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

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

Inductive step. Let $u_1, \ldots, u_k$ list those nodes such that $(u_i, v) \in E$. We have

$$O^{(v)}(A^{(v)}(\overrightarrow{\chi}_{\mathsf{Prop}_{\mathcal{N}^\bullet}(S)}(u_i), \vec{W}(u_i, v))) = 1$$

Let $T = \{u_i | S^{\mathsf{C}} \notin g(u_i)\}$. By our inductive hypothesis,

$$O^{(v)}(A^{(v)}(\overrightarrow{\chi}_T(u_i), \vec{W}(u_i, v))) = 1$$

By choice of O and A,

$$(\mathsf{hash}^{-1}(\mathsf{hash}(\overrightarrow{\chi}_T(u_i), \vec{W}(u_i, v)))[0])^{\mathsf{C}} \notin g(v)$$

i.e. $T^{\complement} \notin g(v)$. We would like to show that $S^{\complement} \notin g(v)$. Suppose for contradiction that $S^{\complement} \in g(v)$. Recall that $T = \{u_i | S^{\complement} \notin g(u_i)\}$, i.e. $T^{\complement} = \{u_i | S^{\complement} \in g(u_i)\}$. Since $S^{\complement} \in g(v)$ and $\mathcal{G}$ is transitive, $T^{\complement} \in g(v)$, which contradicts $T^{\complement} \notin g(v)$.

□

THEOREM 4.7. 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 nominal, propositional, $\neg\varphi$, and $\varphi \wedge \psi$ cases are trivial.

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

$$\begin{array}{lll}
\mathcal{M}, w \Vdash \langle \mathbf{K} \rangle \varphi \text{ iff} & \{u | \mathcal{M}, w \nVdash \varphi\} \notin f(w) & \text{(by definition)} \\
\text{iff} & \{u | u \notin \llbracket \varphi \rrbracket_{\mathcal{N}^{\bullet}}\} \notin f(w) & \text{(Inductive Hypothesis)} \\
\text{iff} & \llbracket \varphi \rrbracket^{\complement}_{\mathcal{N}^{\bullet}} \notin g(w) & \\
\text{iff} & w \in \mathsf{Reach}_{\mathcal{N}^{\bullet}}(\llbracket \varphi \rrbracket) & \text{(by Lemma 4.5)} \\
\text{iff} & w \in \llbracket \langle \mathbf{K} \rangle \varphi \rrbracket_{\mathcal{N}^{\bullet}} & \text{(by definition)} \\
\text{iff} & \mathcal{N}^{\bullet}, w \Vdash \langle \mathbf{K} \rangle \varphi & \text{(by definition)}
\end{array}$$

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

$$\begin{array}{lll}
\mathcal{M}, w \Vdash \langle \mathbf{K}^{\leftarrow} \rangle \varphi \text{ iff} & \exists u \text{ such that } w \in \cap f(u) \text{ and } \mathcal{M}, u \nVdash \varphi & \text{(by definition)} \\
\text{iff} & \exists u \text{ such that } w \in \cap f(u) \text{ and } u \notin \llbracket \varphi \rrbracket_{\mathcal{N}^{\bullet}} & \text{(IH)} \\
\text{iff} & \exists u \in \llbracket \varphi \rrbracket^{\complement}_{\mathcal{N}^{\bullet}} \text{ such that } w \in \displaystyle\bigcap_{X \in f(u)} X & \\
 & \exists u \in \llbracket \varphi \rrbracket^{\complement}_{\mathcal{N}^{\bullet}} \text{ such that } w \in \displaystyle\bigcap_{u \notin \mathsf{Reach}_{\mathcal{N}^{\bullet}}(X^{\complement})} X & \text{(by Lemma 4.5)} \\
 & \exists u \in \llbracket \varphi \rrbracket^{\complement}_{\mathcal{N}^{\bullet}} \text{ such that } w \in \mathsf{Reach}^{-1}_{\mathcal{N}^{\bullet}}(u) & \\
\text{iff} & \mathcal{N}^{\bullet}, w \Vdash \langle \mathbf{K}^{\leftarrow} \rangle \varphi & \text{(by definition)}
\end{array}$$

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

$$\begin{array}{lll}
\mathcal{M}, w \Vdash \langle \mathbf{T} \rangle \varphi \text{ iff} & \{u | \mathcal{M}, u \nVdash \varphi\} \notin g(w) & \text{(by definition)} \\
\text{iff} & \{u | u \notin \llbracket \varphi \rrbracket_{\mathcal{N}^{\bullet}}\} \notin g(w) & \text{(Inductive Hypothesis)} \\
\text{iff} & \llbracket \varphi \rrbracket^{\complement}_{\mathcal{N}^{\bullet}} \notin g(w) & \\
\text{iff} & w \in \mathsf{Prop}_{\mathcal{N}^{\bullet}}(\llbracket \varphi \rrbracket) & \text{(by Lemma 4.6)} \\
\text{iff} & w \in \llbracket \langle \mathbf{T} \rangle \varphi \rrbracket_{\mathcal{N}^{\bullet}} & \text{(by definition)} \\
\text{iff} & \mathcal{N}^{\bullet}, w \Vdash \langle \mathbf{T} \rangle \varphi & \text{(by definition)}
\end{array}$$

□

COROLLARY 4.8. $\mathcal{M} \models \varphi$ 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). $\mathbf{K}(\varphi \to \psi) \to (\mathbf{K}\varphi \to \mathbf{K}\psi)$

(K←). $\mathbf{K}^{\leftarrow}(\varphi \to \psi) \to (\mathbf{K}^{\leftarrow}\varphi \to \mathbf{K}^{\leftarrow}\psi)$

($\mathbf{T_K}$). $\mathbf{K}\varphi \to \varphi$

($\mathbf{4_K}$). $\mathbf{K}\varphi \to \mathbf{K}\mathbf{K}\varphi$

(Asym). $i \to \neg\langle \mathbf{K} \rangle \langle \mathbf{K} \rangle i$

(Back). $\varphi \to \mathbf{K}\langle \mathbf{K}^{\leftarrow}\rangle\varphi$
(Forth). $\varphi \to \mathbf{K}^{\leftarrow}\langle \mathbf{K}\rangle\varphi$

(T$_\mathbf{T}$). $\mathbf{T}\varphi \to \varphi$
(4$_\mathbf{T}$). $\mathbf{T}\varphi \to \mathbf{T}\mathbf{T}\varphi$
(Skel). $i \wedge \mathbf{T}(\langle \mathbf{K}^{\leftarrow}\rangle i \to \varphi) \to \mathbf{T}\varphi$
that is closed under **(Necessitation)**, i.e. if $\varphi \in \mathbf{L}$ then $\Box\varphi \in \mathbf{L}$ for $\Box \in \{\mathbf{K}, \mathbf{K}^{\leftarrow}, \mathbf{T}\}$.

The first group of axioms say that **K** characterizes a monotonic, reflexive, transitive, acyclic graph. The second group are axioms relating **K** and $\mathbf{K}^{\leftarrow}$ — these are from the minimal Tense Logic in temporal logic (**K** is "looking into the future", $\mathbf{K}^{\leftarrow}$ is "looking into the past"). See the SEoP page for more details.

The third group characterizes **T** as a non-monotonic preference relation in terms of **K** and $\mathbf{K}^{\leftarrow}$.

DEFINITION 5.2. [19] **(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] 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] ("Lindenbaum's Lemma") We can extend any set Γ to a maximally consistent set $\Delta \supseteq \Gamma$.

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

5.2 Soundness

THEOREM 5.6. **(Soundness)** If $\Gamma \vdash \varphi$ then $\Gamma \models_{\text{BFNN}} \varphi$

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
(K←)	Definition of Reach^{-1}	Definition of $\mathbf{K}^{\leftarrow}$
(T$_\mathbf{K}$)	Inclusion of Reach	Reflexivity of f
(4$_\mathbf{K}$)	Idempotence of Reach	Transitivity of f
(Asym)	Proposition 1.11	f is acyclic
(T$_\mathbf{T}$)	Inclusion of Prop	Reflexivity of g
(4$_\mathbf{T}$)	Idempotence of Prop	Transitivity of g
(Back)	[CHECK]	[CHECK]
(Forth)	[CHECK]	[CHECK]
(Skel)	Minimal Cause	f is the skeleton of g

□

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 [Leitgeb 2001], who constructs [what] 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 $\mathbf{K}, \mathbf{K}^{\leftarrow}, \mathbf{T}$ is a straightforward exercise in modal logic.

LEMMA 5.7. [19] 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:

- $W^C = \{\Delta \mid \Delta \text{ is maximally consistent}\}$
- 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 5.8. [19] **(Truth Lemma)** We have, for canonical model $\mathcal{M}^C$,

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

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

K case.

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

K$^{\leftarrow}$ case.

$$\begin{array}{rlll}
\mathcal{M}^C, \Delta \Vdash \mathbf{K}^{\leftarrow}\varphi & \text{iff} & \forall \Sigma \in W^C, \text{ if } \Delta \in \cap f(\Sigma) \text{ then } \mathcal{M}, \Sigma \Vdash \varphi & \text{(by definition)} \\
& \text{iff} & \forall \Sigma \in W^C, \text{ if } \Delta \in \cap f(\Sigma) \text{ then } \Sigma \in |\varphi|_{\mathbf{L}} & \text{(by IH)} \\
& \text{iff} & \forall \Sigma \in W^C, \text{ if } \Delta \in \cap f(\Sigma) \text{ then } \varphi \in \Sigma & \\
& & \text{[CHECK]} & \\
& \text{iff} & \mathbf{K}^{\leftarrow}\varphi \in \Delta & \\
& \text{iff} & \Delta \in |\mathbf{K}^{\leftarrow}\varphi|_{\mathbf{L}} & \text{(by definition)}
\end{array}$$

T case.

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

□

THEOREM 5.9. [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 5.10. 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 **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.

f^C is acyclic. [CHECK]

[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^C is the skeleton of g^C. [CHECK]

[for all $w \in W$, if $X \cup (\cap f(w))^C \in g(w)$ then $X \in g(w)$.] □

THEOREM 5.11. **(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 **L** guaranteed by Lemma 5.7. By the Truth Lemma (Lemma 5.8), $\mathcal{M}^C, \Delta \models \Delta$. So in particular, $\mathcal{M}^C, \Delta \models \Gamma$.

By the Finite Model Property (Lemma 5.9), we can construct a finite model $\mathcal{M}'$ satisfying exactly the same formulas at all worlds. By Proposition 5.10, $\mathcal{M}'$ 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 4.7). And so $\mathcal{N}^\bullet, \Delta \models \Gamma$. □

THEOREM 5.12. **(Completeness)** For all consistent $\Gamma \subseteq \mathcal{L}$, if $\Gamma \models_{\text{BFNN}} \varphi$ then $\Gamma \vdash \varphi$

Proof. Suppose contrapositively that $\Gamma \not\vdash \varphi$. This means that $\Gamma \cup \{\neg\varphi\}$ is consistent, i.e. by Theorem 5.11 we can build a BFNN $\mathcal{N}$ and neuron n such that $\mathcal{N}, n \models \Gamma \cup \{\neg\varphi\}$. In particular, $\mathcal{N}, n \not\models \varphi$. But then we must have $\Gamma \not\models \varphi$. □

TODO:

- Incorporate *named* maximally consistent sets in completeness proof
- 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)
- Make drawings in Tikz
- Make corrections Saul gave

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.

[21] Dongran Yu, Bo Yang, Dayou Liu, and Hui Wang. A survey on neural-symbolic systems. *ArXiv preprint arXiv:2111.08164*, 2021.

Appendix A Helper Proofs

Proof. (of Proposition 1.6) We prove each in turn:

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

(Idempotence). 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 $m_1, \ldots, m_k$ such that $(m_i, n) \in E$,

$$O^{(n)}(A^{(n)}(\overrightarrow{\chi}_{\mathsf{Prop}(\mathsf{Prop}(S))}(m_i), \overrightarrow{W}(m_i, n))) = 1$$

By inductive hypothesis, $\chi_{\mathsf{Prop}(\mathsf{Prop}(S))}(m_i) = \chi_{\mathsf{Prop}(S)}(m_i)$. By definition, $n \in \mathsf{Prop}(S)$.

(Cumulative). For the ($\subseteq$) direction, let $n \in \mathsf{Prop}(S_1)$. We proceed by induction on $\mathsf{Prop}(S_1)$.

Base Step. Suppose $n \in S_1$. Well, $S_1 \subseteq S_2 \subseteq \mathsf{Prop}(S_2)$, so $n \in \mathsf{Prop}(S_2)$.

Inductive Step. For those $m_1, \ldots, m_k$ such that $(m_i, n) \in E$,

$$O^{(n)}(A^{(n)}(\overrightarrow{\chi}_{\mathsf{Prop}(S_1)}(m_i), \overrightarrow{W}(m_i, n))) = 1$$

By inductive hypothesis, $\chi_{\mathsf{Prop}(S_1)}(m_i) = \chi_{\mathsf{Prop}(S_2)}(m_i)$. By definition, $n \in \mathsf{Prop}(S_2)$.

Now consider the ($\supseteq$) direction. The Inductive Step holds similarly (just swap S_1 and S_2). As for the Base Step, if $n \in S_2$ then since $S_2 \subseteq \mathsf{Prop}(S_1)$, $n \in S_1$.

(Loop). Let $n \geq 0$ and suppose the hypothesis. Our goal is to show that for each i, $\mathsf{Prop}(S_i) \subseteq \mathsf{Prop}(S_{i-1})$, and additionally $\mathsf{Prop}(S_0) \subseteq \mathsf{Prop}(S_n)$. This will show that all $\mathsf{Prop}(S_i)$ contain each other, and so are equal. Let $i \in \{0, \ldots, n\}$ (if $i = 0$ then $i - 1$ refers to n), and let $e \in \mathsf{Prop}(S_i)$. We proceed by induction on $\mathsf{Prop}(S_i)$.

Base Step. $e \in S_i$, and since $S_i \subseteq \mathsf{Prop}(S_{i-1})$ by assumption, $e \in \mathsf{Prop}(S_{i-1})$.

Inductive Step. For those $m_1,\ldots,m_k$ such that $(m_i,n)\in E$,

$$O^{(e)}(A^{(e)}(\overrightarrow{\chi}_{\mathsf{Prop}(S_i)}(m_i),\overrightarrow{W}(m_i,e)))=1$$

By inductive hypothesis, $\chi_{\mathsf{Prop}(S_i)}(m_j)=\chi_{\mathsf{Prop}(S_{i-1})}(m_j)$. By definition, $n\in\mathsf{Prop}(S_{i-1})$. □

Proof. (of Proposition 1.8) We check each in turn:

(Inclusion). Similar to the proof of Inclusion for Prop.

(Idempotence). Similar to the proof of Idempotence for Prop.

(Monotonicity). 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, $n\in\mathsf{Reach}(S_2)$. □

Proof. (of Proposition 1.10) $(\rightarrow)$ Suppose $u\in\mathsf{Reach}^{-1}(n)$, i.e. for all X such that $n\notin\mathsf{Reach}(X)$, $u\in X^{\mathsf{C}}$. Consider in particular

$$X=\{m|\,\text{there is an}\,E\text{-path from}\,m\,\text{to}\,n\}^{\mathsf{C}}$$

Notice that $n\notin\mathsf{Reach}(X)$. And so $u\in X^{\mathsf{C}}$, i.e. there *is* an E-path from u to n.

$(\leftarrow)$ Suppose there is an E-path from u to n, and let X be such that $n\notin\mathsf{Reach}(X)$. By definition of Reach, there is no $m\in X$ with an E-path from m to n. So in particular, $u\notin X$, i.e. $u\in X^{\mathsf{C}}$. So $u\in\bigcap_{n\notin\mathsf{Reach}(X)}X^{\mathsf{C}}=\mathsf{Reach}^{-1}(n)$. □

Proof. (of Proposition 1.11) Suppose $n_1\in\mathsf{Reach}^{-1}(n_2),\ldots,n_{k-1}\in\mathsf{Reach}^{-1}(n_k),n_k\in\mathsf{Reach}^{-1}(n_1)$. By Proposition 1.10, there is an E-path from each n_i to n_{i+1}, and an E-path from n_k to n_1. But since E is acyclic, each $n_i=n_j$. □

Proof. (of Proposition 1.12) Let $n\in\mathsf{Prop}(S)$. We proceed by induction on $\mathsf{Prop}(S)$.

Base Step. $n\in S$. Our plan is to show $n\in\bigcap_{n\notin\mathsf{Reach}(X)}X^{\mathsf{C}}=\mathsf{Reach}^{-1}(n)$ (so $n\in S\cap\mathsf{Reach}^{-1}(n)$), which will give us our conclusion by the base case of Prop. Let X be any set where $n\notin\mathsf{Reach}(X)$. So $n\notin X$ (since $X\subseteq\mathsf{Reach}(X)$), i.e. $n\in X^{\mathsf{C}}$. But this is what we needed to show.

Inductive Step. Suppose $n\in\mathsf{Prop}(S)$ via its constructor, i.e. for those $m_1,\ldots,m_k$ such that $(m_i,n)\in E$,

$$O^{(n)}(A^{(n)}(\overrightarrow{\chi}_{\mathsf{Prop}(S)}(m_i),\overrightarrow{W}(m_i,n)))=1$$

By inductive hypothesis,

$$\chi_{\mathsf{Prop}(S)}(m_i)=\chi_{\mathsf{Prop}(S\cap(\bigcap_{n\notin\mathsf{Reach}(X)}X^{\mathsf{C}}))}(m_i)$$

So we can substitute the latter for the former. By definition, $n\in\mathsf{Prop}(S\cap(\bigcap_{n\notin\mathsf{Reach}(X)}X^{\mathsf{C}}))$. □

Proof. (of Proposition 2.4) [Todo] □

Proof. (of Proposition 2.5) $(\rightarrow)$ 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. □

Proof. **(of Proposition 4.2)** To show that hash is injective, suppose $\mathsf{hash}(S_1) = \mathsf{hash}(S_2)$. So $\prod_{m_i \in S_1} p_i = \prod_{m_i \in S_2} p_i$, and since products of primes are unique, $\{p_i | m_i \in S_1\} = \{p_i | m_i \in S_2\}$. And so $S_1 = S_2$.

To show that hash is surjective, let $x \in P_k$. Now let $S = \{m_i | p_i \text{ divides } x\}$. Then $\mathsf{hash}(S) = \prod_{m_i \in S} p_i = \prod_{(p_i \text{ divides } x)} p_i = x$. □

Step 5. Step away (for a few days). Come back and check the proof *slowly* to make sure there aren't any missing edge cases or conditions.

- **If it's all good — congratulations, you got a free paper!**
- **Usually there will be some idiotic mistake in the proof. It may seem like *you're* the idiot for trying it — but in fact, it's now your job to figure out *what conditions will make this naive proof work*!**

Step 10. Move on to the write-up stage. But otherwise, step away from the problem — there are too many other interesting things to spend all of your time on this one. Trust that one day a different solution will come to you.