Some Notes on Completeness of the Logic of Hebbian Learning

Caleb Kisby

Navigation

1 Definitions

1.1 Nets and Forward Propagation

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

- N is a finite nonempty set (the set of **neurons**)
- $E \subseteq N \times N$ (the set of **excitatory neurons**)
- $W: N \times N \to \mathbb{R}$ (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)

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

- **Feed-forward**, i.e. E does not contain cycles of edges with all nonzero weights
- **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 $\chi_S \colon N \to \{0, 1\}$ be given by $\chi_S = 1$ iff $n \in S$

We write W_{ij} to mean $W(i, j)$ for $(i, j) \in E$. When m_i is drawn from a sequence $m_1, \ldots, m_k$, we write $\vec{W}(m_i, n)$ as shorthand for the sequence $\vec{W}(m_1, n), \ldots, \vec{W}(m_k, n)$. Similarly, we write $\vec{\chi}(m_i)$ as shorthand for the sequence $\vec{\chi}(m_1), \ldots \vec{\chi}(m_k)$.

Neurons in a state $S \in \mathsf{Set}$ can subsequently activate new neurons, which activate yet more neurons, until eventually the state of $\mathcal{N}$ stabilizes. We call this final state of affairs $\mathsf{Prop}(S)$, i.e. the **propagation** of S. We would also like to consider the set of all neurons that *could possibly* be in a propagation. We call this set $\mathsf{Reach}(S)$, the set of all nodes graph-reachable from S.

DEFINITION 1.5. Let $\mathsf{Prop} \colon \mathsf{Set} \to \mathsf{Set}$ 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

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

DEFINITION 1.6. Let $\mathsf{Reach} \colon \mathsf{Set} \to \mathsf{Set}$ 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 $(m, n) \in E$.

For notational convenience, we also define a function $\mathsf{Reach}^{-1} \colon N \to \mathsf{Set}$ that, given a neuron $n \in N$, recovers the smallest set containing *all* of those m that graph-reach n.

DEFINITION 1.7. For all $n \in N$, $\mathsf{Reach}^{-1}(n) = \bigcap_{n \notin \mathsf{Reach}(X)} X^{\complement}$

PROPOSITION 1.8. For all $n \in N$, $\mathsf{Reach}^{-1}(n) = \{m \mid \text{there is an } E\text{-path from } m \text{ to } n\}$

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

$$X = \{m \mid \text{there is an } E\text{-path from } m \text{ to } n\}^{\complement}$$

Notice that $n \notin \mathsf{Reach}(X)$. And so $u \in X^{\complement}$, 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. $u \in X^{\complement}$. So $u \in \bigcap_{n \notin \mathsf{Reach}(X)} X^{\complement} = \mathsf{Reach}^{-1}(n)$. □

1.2 Properties of Prop and Reach

PROPOSITION 1.9. (LEITGEB) 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 $i, j \in \{0, \ldots, n\}$

Proof. 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 $m_1, \dots, m_k$ such that $(m_i, n) \in E$,

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

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 $m_1, \dots, m_k$ such that $(m_i, n) \in E$,

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

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 $i \in \{0, \dots, n\}$ (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 $m_1, \dots, m_k$ such that $(m_i, n) \in E$,

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

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

Proposition 1.10. 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 $\mathsf{Reach}(S_1) \subseteq \mathsf{Reach}(S_2)$

(Pointwise-Antisymmetry). If $n \in \mathsf{Reach}^{-1}(m)$ and $m \in \mathsf{Reach}^{-1}(n)$ then $n = m$.

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

(Inverse-Antisymmetry). Suppose $n \in \mathsf{Reach}^{-1}(m)$ and $m \in \mathsf{Reach}^{-1}(n)$. By Proposition 1.8, there is an E-path from n to m and an E-path from m to n. But since E is acyclic, $m = n$. □

Intuitively, Reach is the fully monotonic extension of Prop. Prop is not monotonic because the weights of the net may be negative. Prop and Reach interact in the following way:

PROPOSITION 1.11. **(Minimal Cause)** For all $n \in N$, $n \in \mathsf{Prop}(S)$ iff $n \in \mathsf{Prop}(S \cap \mathsf{Reach}^{-1}(n))$

[Is there a property *expressible in our modal logic* that implies **(Minimal Cause)**?]

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

Base Step. $n \in S$. Our plan is to show $n \in \bigcap_{n \notin \mathsf{Reach}(X)} X^{\complement} = \mathsf{Reach}^{-1}(n)$ (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. $n \in X^{\complement}$. 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 $(m_i, n) \in E$,

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

By inductive hypothesis,

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

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

($\leftarrow$) Let $n \in \mathsf{Prop}(S \cap \mathsf{Reach}^{-1}(n))$. We again proceed by induction.

Base Step. $n \in S \cap \mathsf{Reach}^{-1}(n)$. So in particular, $n \in S$. By the base case of Prop, $n \in \mathsf{Prop}(S)$.

Inductive Step. Similar to the inductive step for the ($\rightarrow$) direction. $\square$

1.3 Syntax and (Neural) Semantics

DEFINITION 1.12. 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. Material implication $\varphi \rightarrow \psi$ is defined as $\neg\varphi \vee \psi$. We define $\bot, \vee, \leftrightarrow, \Leftrightarrow$, and the dual operators $\langle\mathbf{K}\rangle, \langle\mathbf{T}\rangle$ in the usual way.

DEFINITION 1.13. A **neural network model** is $\langle\mathcal{N}, [\![\cdot]\!]\rangle$, where $\mathcal{N}$ is a BFNN and $[\![\cdot]\!] : \mathcal{L} \rightarrow \mathsf{Set}_{\mathcal{N}}$ is an interpretation function.

DEFINITION 1.14. Let $\mathcal{N} \in \mathsf{Net}$. The semantics for $\mathcal{L}$ are defined recursively as follows:

$[\![p]\!]$		$\in$ Set is fixed, nonempty
$[\![\neg\varphi]\!]$	$=$	$[\![\varphi]\!]^{\complement}$
$[\![\varphi \wedge \psi]\!]$	$=$	$[\![\varphi]\!] \cap [\![\psi]\!]$
$[\![\langle\mathbf{K}\rangle\varphi]\!]$	$=$	$\mathsf{Reach}([\![\varphi]\!])$
$[\![\langle\mathbf{T}\rangle\varphi]\!]$	$=$	$\mathsf{Prop}([\![\varphi]\!])$

DEFINITION 1.15. $\mathcal{N} \models \varphi$ iff $[\![\varphi]\!]_{\mathcal{N}} = N$

1.4 Hypergraphs

DEFINITION 1.16. A **hypergraph** is a tuple $\mathcal{H} = \langle V, H\rangle$, where V is the set of vertices and $H : \mathcal{P}(V) \times V$ is the set of **hyperarcs**.

Unfortunately, there is no agreed-upon definition of a path in a hypergraph. There are several candidates to choose from, and this choice is made to fit the use-case [3]. The same goes for properties on $\mathcal{H}$ (e.g. reflexivity, transitivity, etc.). I've decided on the following definitions, which are highly non-standard but natural for our setting.

DEFINITION 1.17. Say we have a sequence $(S = S_1, \ldots S_l)$ such that for all i,

$$S_{i+1} = \{u \mid S_i H u\}$$

We call this the **sequence generated by S** (of length l).

PROPOSITION 1.18. The sequence of length l generated by S is unique.

Proof. By induction on l.

Base Step. $l = 1$, so both sequences are (S), and we are done.

Inductive Step. Say we have sequences $(S = A_1, A_2, \ldots, A_l)$ and $(S = B_1, B_2, \ldots, B_l)$ generated by S. By inductive hypothesis, $A_{l-1} = B_{l-1}$. And so

$$A_l = \{u \mid A_{l-1} H u\} = \{u \mid B_{l-1} H u\} = B_l$$

and we are done. □

DEFINITION 1.19. We say there is a **strong hyperpath** from source set S to node v iff for some l, the sequence $(S = S_1, S_2, S_3, \ldots S_l)$ generated by S is such that $v \in S_l$.

DEFINITION 1.20. We say there is a **strong hypercycle** iff for some S, l, the sequence generated by S begins and ends with S, i.e. is $(S = S_1, S_2, S_3, \ldots S_l = S)$.

DEFINITION 1.21. Let $\mathcal{H} = \langle V, H \rangle$ be a hypergraph, and $G = \langle V, E \rangle$ be a binary graph.

- $\mathcal{H}$ is **reflexive** iff for all $s \in S$, SHs.
- $\mathcal{H}$ is **transitive** iff $T = \{u \mid SHu\}$ and THv implies SHv.
- $\mathcal{H}$ is **acyclic** iff for all strong hypercycles $(S_1, S_2, S_3, \ldots S_l)$ in $\mathcal{H}$, $S_i = S_j$ for all $i, j \in \{1, \ldots, l\}$.
- $\mathcal{H}$ **extends** G whenever SHv iff $(S \cap \{u \mid uEv\})Hv$.

DEFINITION 1.22. Let $\mathcal{H} = \langle V, H \rangle$ be a hypergraph. Then

- $\mathcal{H}^* = \langle V, H^* \rangle$, the **reflexive-transitive closure** of $\mathcal{H}$, is that graph extending $\mathcal{H}$ with the minimum number of hyperedges such that it is reflexive and transitive.
- $\mathcal{H}^- = \langle V, H^- \rangle$, the **reflexive-transitive reduction** of $\mathcal{H}$, is the graph with the minimum number of hyperedges such that $(H^-)^* = H^*$.

PROPOSITION 1.23. Let $\mathcal{H}^*$ be the reflexive-transitive closure of $\mathcal{H}^*$. Then SH^*v iff there is a strong hyperpath from S to v in $\mathcal{H}$.

Proof. $(\rightarrow)$ If SH^*v, then SHv [Todo: This is a big mistake!!!] (since H^* is an extension of H). Let $S_2 = \{u \mid SHu\}$. Then (S, S_2) with $v \in S_2$ forms a strong hyperpath from S to v.

$(\leftarrow)$ We proceed by induction on the length of the strong hyperpath.

Base Step. The hyperpath is of length 1, i.e. (S), with $v \in S$. Since H^* is reflexive, SH^*v.

Inductive Step. Say we have strong hyperpath $(S = S_1, S_2, \ldots, S_l)$ from S to v (in H). By definition of strong hyperpath, $S_l = \{u \mid S_{l-1} H u\}$. In particular, $S_{l-1}Hv$.

Claim. $S_{l-1} = \{u | SH^*u\}$.

Proof. $(\rightarrow)$ Let $u \in S_{l-1}$. We have strong hyperpath $(S = S_1, S_2, \ldots, S_{l-1})$ from above, with $u \in S_{l-1}$. By inductive hypothesis, SH^*u.

$(\leftarrow)$ Suppose SH^*u. By inductive hypothesis, there is a strong hyperpath from S to u in H. But the strong sequence generated by S is unique, so we know the hyperpath is $(S = S_1, S_2, \ldots, S_{l-1})$ with $u \in S_{l-1}$. □

From here, we have $S_{l-1} = \{u | SH^*u\}$ and $S_{l-1}Hv$ (so $S_{l-1}H^*v$). By transitivity, we have SH^*v. □

Proposition 1.24. Let $\mathcal{H} = \langle V, H \rangle$ be a reflexive, transitive, acyclic hypergraph relation. If V is finite, $\mathcal{H}$ has a unique reflexive-transitive reduction $\mathcal{H}^{-}$.

Proof.

Existence. First, since $\mathcal{H}$ is reflexive and transitive, $H^* = H$. So the set $\{H^- = \langle V, E^- \rangle | (H^-)^* = H\}$ is not empty. Since V is finite, this set is also finite. By the well-ordering principle, there is such an H^- in this set with the minimum number of hyperedges.

Uniqueness. [TODO: Should use acyclic property] □

Proposition 1.25. The reflexive-transitive reduction $\mathcal{H}^{-}$ of $\mathcal{H}$ is irreflexive and anti-transitive.

Proof. [Todo: Use the fact that H^- is minimal.] □

Proposition 1.26. Let $u, v \in V$ be such that $u \neq v$ and there is no $x \neq v \in V$ such that uRx, xR^*v. Then

$$SR^*v \quad \text{iff} \quad SRv$$

In other words, R^* and R agree on every edge that does not result in reflexivity or transitivity.

Proof. [Todo] □

Proposition 1.27. Let $S \in \mathcal{P}(V), v \in V$ be such that $v \notin S$ and $\neg TH^*v$ for $T = \{u | SHu\}$. Then

$$SH^*v \quad \text{iff} \quad SHv$$

In other words, H^* and H agree on every hyperedge that does not result in reflexivity or transitivity.

Proof. $(\leftarrow)$ This direction follows by definition of H^*.

$(\rightarrow)$ Suppose not; SH^*v, but $\neg SHv$. Let $(S, S_2, \ldots, S_l)$ be a strong hyperpath with $v \in S_l$. Since $v \notin S$, $l \neq 0$. Since $\neg SHv$, $l \neq 1$. So we must have $l \geqslant 2$. This means that for some $2 \leqslant i < l$, $S_i \neq S$ and $S_i \neq S_l$. Without loss of generality, say $i = 2$.

By definition, $S_2 = \{u | SHu\}$. But S_2H^*v (via the hyperpath $(S_2, \ldots, S_l)$), which contradicts our assumption that $\neg TH^*v$ for $T = \{u | SHu\}$. □

1.5 Neighborhood Semantics

Modal logics are traditionally interpreted via a relational possible-worlds semantics. But relational frames do not allow for non-normal operators like **T**. Instead, we need to use neighborhood semantics (see [2]).

Definition 1.28. 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**.

The intuition is that $f(w)$ selects those sets of worlds that are necessary (or known, or typical) at w.

Definition 1.29. A **multi-frame** is $\mathfrak{F} = \langle W, f, g\rangle$, where f and g are neighborhood functions.

Definition 1.30. 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 1.31. Let $\mathcal{F} = \langle W, f\rangle$ be a neighborhood frame, and let $w \in W$. We say that $X, Y \in f(w)$ **form a loop** in $\mathcal{F}$ iff there exist $X = X_1 \ldots, X_n = Y$ such that for all $1 \leq i \leq n$, $\{u|\, X_i \in f(u)\} \subseteq X_{i+1}$. (Including, of course, $\{u|\, X_n \in f(u)\} \subseteq X_1$).

Definition 1.32. Let $\mathcal{F} = \langle W, f\rangle, \mathcal{G} = \langle W, g\rangle$ be neighborhood frames with W nonempty.

- $\mathcal{F}$ is **closed under finite intersections** iff for all $w \in W$, if $X_1, \ldots, X_n \in f(w)$ then their intersection $\bigcap_{i=1}^{k} X_i \in f(w)$.
- $\mathcal{F}$ is **closed under supersets** iff for all $w \in W$, if $X \in f(w)$ and $X \subseteq Y \subseteq W$, then $Y \in f(w)$.
- $\mathcal{F}$ **contains the unit** iff $W \in f(w)$.
- $\mathcal{F}$ **contains the empty set** iff $\emptyset \in f(w)$.
- $\mathcal{F}$ is **reflexive** iff for all $w \in W$, $w \in \cap f(w)$
- $\mathcal{F}$ is **transitive** iff for all $w \in W$, if $X \in f(w)$ then $\{u \,|\, X \in f(u)\} \in f(w)$.
- $\mathcal{F}$ is **antisymmetric** iff for all $u, v \in W$, if $u \in \cap f(v)$ and $v \in \cap f(u)$ then $u = v$.
- $\mathcal{F}$ is **loop-cumulative** iff for all $w \in W$, if X and Y form a loop, then

$$\{u \,|\, X \in g(u)\} = \{u \,|\, Y \in g(u)\}$$

- $\mathcal{G}$ **subsumes** $\mathcal{F}$ iff for all $w \in W$, $X \cup (\cap f(w))^{\complement} \in g(w)$ iff $X \in g(w)$.

Definition 1.33. Let $\mathcal{F} = \langle W, f\rangle$ be a frame, and $\mathfrak{F} = \langle W, f, g\rangle$ be a multi-frame extending $\mathcal{F}$. We will focus on the following special classes of frames:

- $\mathcal{F}$ is a **proper filter** iff for all $w \in W$, $f(w)$ is closed under finite intersections, closed under supersets, contains the unit, and does not contain the empty set.
- $\mathcal{F}$ is a **loop-subfilter** iff for all $w \in W$, $f(w)$ contains the unit and is loop-cumulative.
- $\mathfrak{F}$ is a **preferential multi-frame** iff
  - $\mathcal{F} = \langle W, f\rangle$ forms a reflexive, transitive, proper filter,
  - $\mathcal{G} = \langle W, g\rangle$ forms a reflexive, transitive, loop-subfilter and
  - $\mathcal{G}$ subsumes $\mathcal{F}$

Proposition 1.34. (Pacuit) If $\mathcal{F} = \langle W, f\rangle$ is a filter, and W is finite, then $\mathcal{F}$ contains its core.

Proposition 1.35. If $\mathcal{F} = \langle W, f\rangle$ is a proper filter, then for all $w \in W$, $Y^{\complement} \in f(w)$ iff $Y \notin f(w)$.

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

($\leftarrow$) Suppose for contradiction that $Y \notin f(w)$, yet $Y^{\complement} \notin f(w)$. Since $\mathcal{F}$ is closed under intersection, $\cap f(w) \in f(w)$. Moreover, since $\mathcal{F}$ is closed under superset we must have $\cap f(w) \not\subseteq Y$ and $\cap f(w) \not\subseteq Y^{\complement}$. But this means $\cap f(w) \not\subseteq Y \cap Y^{\complement} = \emptyset$, i.e. there is some $x \in \cap f(w)$ such that $x \in \emptyset$. This contradicts the definition of the empty set. $\square$

Definition 1.36. Let $\mathcal{F} = \langle W, f \rangle$, $\mathcal{G} = \langle W, g \rangle$ be a neighborhood frame. A **neighborhood model** based on $\mathcal{F}$ and $\mathcal{G}$ is $\mathcal{M} = \langle W, f, g, V \rangle$, where $V : \mathcal{L} \to \mathcal{P}(W)$ is a valuation function.

Definition 1.37. Let $\mathcal{M} = \langle W, f, g, V \rangle$ be a model based on two frames $\mathcal{F} = \langle W, f \rangle, \mathcal{G} = \langle W, g \rangle$. The (neighborhood) semantics for $\mathcal{L}$ are defined recursively as follows:

$$\begin{array}{lll}
\mathcal{M}, w \vDash p & \text{iff} & w \in V(p) \\
\mathcal{M}, w \vDash \neg\varphi & \text{iff} & \mathcal{M}, w \nvDash \varphi \\
\mathcal{M}, w \vDash \varphi \wedge \psi & \text{iff} & \mathcal{M}, w \vDash \varphi \text{ and } \mathcal{M}, w \vDash \psi \\
\mathcal{M}, w \vDash \langle \mathbf{K} \rangle \varphi & \text{iff} & \{u \mid \mathcal{M}, u \nvDash \varphi\} \notin f(w) \\
\mathcal{M}, w \vDash \langle \mathbf{T} \rangle \varphi & \text{iff} & \{u \mid \mathcal{M}, u \nvDash \varphi\} \notin g(w)
\end{array}$$

1.6 Rules and Axioms

The proof system for our logic is as follows. We have $\vdash \varphi$ iff either φ is an axiom, or φ follows from previously obtained formulas by one of the inference rules (axioms and rules shown below). If $\Gamma \subseteq \mathcal{L}$ is a set of formulas, we consider $\Gamma \vdash \varphi$ to hold whenever there exist finitely many $\psi_1, \ldots, \psi_k \in \Gamma$ such that $\vdash \psi_1 \wedge \ldots \wedge \psi_k \to \varphi$.

Basic Axioms and Inference Rules		K Axioms		T Axioms	
(PC)	All propositional tautologies	(K)	$\mathbf{K}(\varphi \to \psi) \to (\mathbf{K}\varphi \to \mathbf{K}\psi)$	(Loop)	$(\mathbf{T}\varphi_1 \to \varphi_2) \wedge \ldots \wedge (\mathbf{T}\varphi_n \to \varphi$ $\to (\mathbf{T}\varphi_i \leftrightarrow \mathbf{T}\varphi_j)$ (for each $i,$
(MP)	$\dfrac{\varphi \quad \varphi \to \psi}{\psi}$	(T)	$\mathbf{K}\varphi \to \varphi$	(T)	$\mathbf{T}\varphi \to \varphi$
($\text{Nec}_{\mathbf{K}}$)	$\dfrac{\varphi}{\mathbf{K}\varphi}$	(4)	$\mathbf{K}\varphi \to \mathbf{K}\mathbf{K}\varphi$	(4)	$\mathbf{T}\varphi \to \mathbf{T}\mathbf{T}\varphi$
($\text{Nec}_{\mathbf{T}}$)	$\dfrac{\varphi}{\mathbf{T}\varphi}$			($\mathbf{KT}$)	$\mathbf{K}\varphi \to \mathbf{T}\varphi$

2 Are the Semantics Flipped?

The semantics for $\wedge, \to$ are flipped, relative to what we had before. We could have alternatively given our semantics by:

$$\begin{array}{lll}
[\![p]\!]' & & \in \mathsf{Set} \text{ is fixed, nonempty} \\
[\![\neg\varphi]\!]' & = & [\![\varphi]\!]'^{\complement} \\
[\![\varphi \wedge \psi]\!]' & = & [\![\varphi]\!]' \cup [\![\psi]\!]' \\
[\![\mathbf{K}\varphi]\!]' & = & \mathsf{Reach}([\![\varphi]\!]') \\
[\![\mathbf{T}\varphi]\!]' & = & \mathsf{Prop}([\![\varphi]\!]')
\end{array}$$

and then defined

$$\mathcal{N} \vDash' \varphi \quad \text{iff} \quad [\![\varphi]\!]'_{\mathcal{N}} = \emptyset$$

As explained by Leitgeb in [1]: This choice reflects the intuition that neurons act as "elementary-feature-detectors". For example, say $[\![\varphi]\!]$ represents those neurons that are *necessary* for detecting an apple, and $[\![\psi]\!]$ represents those neurons that are *necessary* for detecting the color red. If the net observes a red apple ($\varphi \wedge \psi$), both the neurons detecting red-features $[\![\varphi]\!]$ and the neurons detecting apple-features $[\![\psi]\!]$ necessarily activate, i.e. $[\![\varphi]\!] \cup [\![\psi]\!]$ activates. As for implication, "every apple is red" ($\varphi \to \psi$) holds for a net iff whenever the neurons detecting apple-features $[\![\varphi]\!]$ necessarily activate, so do the neurons detecting red-features $[\![\psi]\!]$. But this is only true if $[\![\varphi]\!] \supseteq [\![\psi]\!]$.

However, our next result shows that these two choices are interchangeable — and so we stick with the choice of semantics that is easier to work with.

Proposition 2.1. For all $\mathcal{N} \in \mathsf{Net}$,

$$\mathcal{N} \vDash' \varphi \quad \text{iff} \quad \mathcal{N} \vDash \varphi$$

Proof. The key point is that for all φ, $[\![\varphi]\!]^{\complement} = [\![\varphi]\!]'$. To see this for $\mathbf{K}\varphi$ and $\mathbf{T}\varphi$ (assuming it holds inductively for φ), note that

$$[\![\mathbf{K}\varphi]\!]^{\complement} = [\![\neg\langle\mathbf{K}\rangle\neg\varphi]\!]^{\complement} = \mathsf{Reach}([\![\varphi]\!]^{\complement}) = \mathsf{Reach}([\![\varphi]\!]') = [\![\mathbf{K}\varphi]\!]'$$
$$[\![\mathbf{T}\varphi]\!]^{\complement} = [\![\neg\langle\mathbf{T}\rangle\neg\varphi]\!]^{\complement} = \mathsf{Prop}([\![\varphi]\!]^{\complement}) = \mathsf{Prop}([\![\varphi]\!]') = [\![\mathbf{T}\varphi]\!]'$$

And so

$$\begin{aligned} \mathcal{N} \models' \varphi \quad &\text{iff} \quad [\![\varphi]\!]' = \emptyset \\ &\text{iff} \quad [\![\varphi]\!] = N \\ &\text{iff} \quad \mathcal{N} \models \varphi \end{aligned}$$

□

3 Neural Semantics ⇝ Neighborhood Semantics

Definition 3.1. Given a BFNN $\mathcal{N}$, its **simulation frame** $\mathfrak{F}^{\bullet} = \langle W, f, g\rangle$ 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, given an interpretation function $[\![\cdot]\!]_{\mathcal{N}}$, the **simulation model** $\mathcal{M}^{\bullet} = \langle W, f, g, V\rangle$ based on $\mathfrak{F}^{\bullet}$ has:

- $V(p) = [\![p]\!]_{\mathcal{N}}$

Theorem 3.2. Let $\mathcal{N}$ be a BFNN, and let $\mathcal{M}^{\bullet}$ be the simulation model based on $\mathfrak{F}^{\bullet}$. Then

$$\mathcal{M}^{\bullet}, w \models \varphi \quad \text{iff} \quad w \in [\![\varphi]\!]_{\mathcal{N}}$$

Proof. By induction on φ.

p case:

$$\begin{aligned} \mathcal{M}^{\bullet}, w \models p \quad &\text{iff} \quad w \in V(p) \\ &\text{iff} \quad w \in [\![p]\!] \end{aligned}$$

$\neg\varphi$ case:

$$\begin{aligned} \mathcal{M}^{\bullet}, w \models \neg\varphi \quad &\text{iff} \quad \mathcal{M}^{\bullet}, w \not\models \varphi \\ &\text{iff} \quad w \notin [\![\varphi]\!] && \text{(IH)} \\ &\text{iff} \quad w \in [\![\neg\varphi]\!] \end{aligned}$$

$\varphi \wedge \psi$ case:

$$\begin{aligned} \mathcal{M}^{\bullet}, w \models \varphi \wedge \psi \quad &\text{iff} \quad \mathcal{M}^{\bullet}, w \models \varphi \text{ and } \mathcal{M}^{\bullet}, w \models \psi \\ &\text{iff} \quad w \in [\![\varphi]\!] \text{ and } w \in [\![\psi]\!] && \text{(IH)} \\ &\text{iff} \quad w \in [\![\varphi \wedge \psi]\!] \end{aligned}$$

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

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

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

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

□

Corollary 3.3. $\mathcal{M}^\bullet \vDash \varphi$ iff $\mathcal{N} \vDash \varphi$.

Proof.

$$\begin{array}{lll} \mathcal{M}^\bullet \vDash \varphi & \text{iff} & \mathcal{M}^\bullet, w \vDash \varphi \text{ for all } w \in W = N \\ & \text{iff} & w \in [\![\varphi]\!]_\mathcal{N} \text{ for all } w \in N \\ & \text{iff} & [\![\varphi]\!]_\mathcal{N} = N \\ & \text{iff} & \mathcal{N} \vDash \varphi \end{array}$$

□

Theorem 3.4. $\mathfrak{F}^\bullet$ is a preferential multi-frame.

Proof. We show each in turn:

- **$\mathcal{F}$ is closed under finite intersection:** Suppose $X_1, \ldots, X_n \in f(w)$. By definition of f, $w \notin \bigcup_i \mathsf{Reach}(X_i^\complement)$ for all i. Since Reach is monotonic, by Proposition [TODO] we have $\bigcup_i \mathsf{Reach}(X_i^\complement) = \mathsf{Reach}(\bigcup_i X_i^\complement) = \mathsf{Reach}((\bigcap_i X_i)^\complement)$. So $w \notin \mathsf{Reach}((\bigcap_i X_i)^\complement)$. But this means that $\bigcap_i X_i \in f(w)$.
- **$\mathcal{F}$ is closed under superset:** Suppose $X \in f(w), X \subseteq Y$. By definition of f, $w \notin \mathsf{Reach}(X^\complement)$. Note that $Y^\complement \subseteq X^\complement$, and so by monotonicity of Reach we have $w \notin \mathsf{Reach}(Y^\complement)$. But this means $Y \in f(w)$, so we are done.
- **$\mathcal{F}$ contains the unit:** Note that for all $w \in W$, $w \notin \mathsf{Reach}(\emptyset) = \mathsf{Reach}(W^\complement)$. So $W \in f(w)$.
- **$\mathcal{F}$ does not contain the empty set:** Similarly, for all $w \in W$, $w \in \mathsf{Reach}(W) = \mathsf{Reach}(W) = \mathsf{Reach}(\emptyset^\complement)$. So $\emptyset \notin f(w)$.
- **$\mathcal{F}$ is reflexive:** We want to show that $w \in \cap f(w)$. Well, suppose $X \in f(w)$, i.e. $w \notin \mathsf{Reach}(X^\complement)$ (by definition of f). Since for all S, $S \subseteq \mathsf{Reach}(S)$, we have $w \notin X^\complement$. But this means $w \in X$, and we are done.
- **$\mathcal{F}$ is transitive:** Suppose $X \in f(w)$, i.e. $w \notin \mathsf{Reach}(X^\complement)$. Well,

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

  So by definition of f, $\{u \mid X \in f(u)\} \in f(w)$.
- **$\mathcal{F}$ is antisymmetric:** Suppose $u \in \cap f(v)$ and $v \in \cap f(u)$, i.e. $u \in \bigcap_{X \in f(v)} X$ and $v \in \bigcap_{X \in f(u)} X$. By choice of f, $u \in \bigcap_{v \notin \mathsf{Reach}(X^\complement)} X$ and $v \in \bigcap_{u \notin \mathsf{Reach}(X^\complement)} X$. Substituting $X^\complement$ for X we get $u \in \bigcap_{v \notin \mathsf{Reach}(X)} X^\complement$ and $v \in \bigcap_{u \notin \mathsf{Reach}(X)} X^\complement$. In other words, $u \in \mathsf{Reach}^{-1}(v)$ and $v \in \mathsf{Reach}^{-1}(u)$. By Pointwise-Antisymmetry of Reach, $u = v$.
- **$\mathcal{G}$ contains the unit:** Similarly to g, for all $w \in W$, $w \notin \mathsf{Prop}(\emptyset)$, i.e. $W \in g(w)$.
- **$\mathcal{G}$ is loop-cumulative:**

$$\begin{array}{llcll} & \{u \mid X_1 \in g(u)\} \subseteq X_2 & ,\ldots, & \{u \mid X_n \in g(u)\} \subseteq X_1 & \\ (\rightarrow) & X_2^\complement \subseteq \{u \mid X_1 \notin g(u)\} & ,\ldots, & X_1^\complement \subseteq \{u \mid X_n \notin g(u)\} & \\ (\rightarrow) & X_2^\complement \subseteq \{u \mid u \in \mathsf{Prop}(X_1^\complement)\} & ,\ldots, & X_1^\complement \subseteq \{u \mid u \in \mathsf{Prop}(X_n^\complement)\} & \text{(by)} \\ (\rightarrow) & X_2^\complement \subseteq \mathsf{Prop}(X_1^\complement) & ,\ldots, & X_1^\complement \subseteq \mathsf{Prop}(X_n^\complement) & \\ (\rightarrow) & \mathsf{Prop}(X_i^\complement) & = & \mathsf{Prop}(X_j^\complement) & \text{for all } i, j \text{ (by Loop for Prop)} \\ (\rightarrow) & \{u \mid u \in \mathsf{Prop}(X_i^\complement)\} & = & \{u \mid u \in \mathsf{Prop}(X_j^\complement)\} & \\ (\rightarrow) & \{u \mid X_i \notin g(u)\} & = & \{u \mid X_j \notin g(u)\} & \text{(by definition of } g\text{)} \\ (\rightarrow) & \{u \mid X_i \in g(u)\} & = & \{u \mid X_j \in g(u)\} & \end{array}$$

So in particular, if X and Y form a loop, then $\{u|\, X \in g(u)\} = \{u|\, Y \in g(u)\}$.

- **$\mathcal{G}$ is reflexive:** Follows similarly, since $X \subseteq \mathsf{Prop}(X)$ by (Inclusion).
- **$\mathcal{G}$ is transitive:** Follows similarly, since $\mathsf{Prop}(X) = \mathsf{Prop}(\mathsf{Prop}(X))$ by (Idempotence).
- **$\mathcal{G}$ subsumes $\mathcal{F}$:** Suppose $X \cup (\cap f(w))^{\complement} \in g(w)$. By choice of g, $w \notin \mathsf{Prop}([X \cup (\cap f(w))^{\complement}]^{\complement})$. Distributing the outer complement, we have $w \notin \mathsf{Prop}(X^{\complement} \cap (\cap f(w)))$, i.e. $w \notin \mathsf{Prop}\big(X^{\complement} \cap \big(\bigcap_{Y \in f(w)} Y\big)\big)$. By choice of f, $w \notin \mathsf{Prop}\Big(X^{\complement} \cap \Big(\bigcap_{w \notin \mathsf{Reach}(Y^{\complement})} Y\Big)\Big)$. Substituting $Y^{\complement}$ for Y, we get $w \notin \mathsf{Prop}\big(X^{\complement} \cap \big(\bigcap_{w \notin \mathsf{Reach}(Y)} Y^{\complement}\big)\big)$. By definition of Reach^{-1}, $w \notin \mathsf{Prop}(X^{\complement} \cap \mathsf{Reach}^{-1}(w))$. From (Minimal Cause), we conclude that $w \notin \mathsf{Prop}(X^{\complement})$, i.e. $X \in g(w)$.

□

4 Neighborhood Semantics $\rightsquigarrow$ Neural Semantics

DEFINITION 4.1. Let $\mathfrak{F} = \langle W, f, g\rangle$ be a preferential multi-frame. We define the graph relation R_f^* and hypergraph relation H_g^* by:

- $u R_f^* v$ iff $u \in \cap f(v)$
- $S H_g^* v$ iff $S^{\complement} \notin g(v)$

PROPOSITION 4.2. R_f^* and H_g^* are both reflexive, transitive, and acyclic, and H_g^* extends R_f^*.

Proof. Let $\mathcal{F} = \langle W, f\rangle$, and $\mathcal{G} = \langle W, g\rangle$

- **R_f^* is reflexive:** Since $\mathcal{F}$ is a reflexive frame, for all $u \in W$ we have $u \in \cap f(u)$. But by definition this means $u\, R_f^*\, u$.
- **R_f^* is transitive:** Suppose $u\, R_f^*\, v$ and $v\, R_f^*\, w$. So $u \in \cap f(v)$ and $v \in \cap f(w)$. Our goal is to show that $u \in \cap f(w)$. So let $X \in f(w)$ — we want to show that $u \in X$.

  Since $X \in f(w)$, $\{y|\, X \in f(y)\} \in f(w)$ (since $\mathcal{F}$ is a transitive frame). But this means that $\cap f(w) \subseteq \{y|\, X \in f(y)\}$. Since $v \in \cap f(w)$, $X \in f(v)$. But this means that $\cap f(v) \subseteq X$. Since $u \in \cap f(v)$, $u \in X$.
- **R_f^* is acyclic:** Suppose $u_1 R_f^* u_2, \ldots, u_{n-1} R_f^* u_n, u_n R_f^* u_1$. We will show $u_1 = u_n$ (the other cases are similar). Since R_f^* is transitive, we have $u_1 R_f^* u_n$. $u_n R_f^* u_1$ gives us $u_n \in \cap f(u_1)$, and $u_1 R_f^* u_n$ gives us $u_1 \in \cap f(u_n)$. Since $\mathcal{F}$ is an antisymmetric frame, $u_1 = u_n$.
- **H_g^* is reflexive:** Suppose $s \in S$ and for contradiction $\neg S H_g^* s$, i.e. $S^{\complement} \in g(s)$. Since $s \in \cap g(s)$ (since $\mathcal{G}$ is a reflexive frame), and $\cap g(s) \subseteq S^{\complement}$ (by definition of core), $s \in S^{\complement}$, which contradicts $s \in S$.
- **H_g^* is transitive:** Suppose $T = \{v|\, S H_g^* v\}$ and $T H_g^* u$. We want to show that $S H_g^* u$. Suppose for contradiction that $\neg S H_g^* u$, i.e. $S^{\complement} \in g(u)$. Since $u \in U$, $T H_g^* u$, i.e. $T^{\complement} \notin g(u)$. Note also that $T = \{u|\, S^{\complement} \notin g(u)\}$, i.e. $T^{\complement} = \{u|\, S^{\complement} \in g(u)\}$. Since $\mathcal{G}$ is a transitive frame, $T^{\complement} \in g(u)$. But this contradicts $T^{\complement} \notin g(u)$.
- **H_g^* is acyclic:** Suppose we have a strong hypercycle $(S_1, \ldots, S_l, S_1)$ in H_g^*. We will show that $S_1 = S_l$ (the other cases are similar). By definition, for each $(i, i+1)$ (including $(l, 1)$), $S_{i+1} = \{u|\, S_i H_g^* u\}$. So $S_{i+1} = \{u|\, S_i^{\complement} \notin g(u)\}$, i.e.

$$S_{i+1}^{\complement} = \{u|\, S_i^{\complement} \in g(u)\}$$

  Since $\mathcal{G}$ is a loop-cumulative frame, $\{u|\, S_i^{\complement} \in g(u)\} = \{u|\, S_j^{\complement} \in g(u)\}$ for all $i, j \in \{1, \ldots, n\}$.

  From here, we can prove $S_1 = S_l$. We will show $S_1 \subseteq S_l$, but the other direction is similar (in fact, it's easier). Let $u \in S_1$. Well,

$$S_l^{\complement} = \big\{u|\, S_{l-1}^{\complement} \in g(u)\big\} = \{u|\, S_1^{\complement} \in g(u)\}$$

We claim that $S_1^{\complement} \notin g(u)$ — from which we can conclude $u \notin S_l^{\complement}$, i.e. $u \in S_l$ (which was our goal). Well, suppose not; say $S_1^{\complement} \in g(u)$. By definition of core we have $\cap g(u) \subseteq S_1^{\complement}$. Since $\mathcal{G}$ is a reflexive frame, $u \in \cap g(u)$, and so $u \in S_1^{\complement}$. This contradicts $u \in S_1$ from before.

- $\boldsymbol{H_g^*}$ **extends** $\boldsymbol{R_f^*}$**:** Suppose $S H_g^* v$. By definition, $S^{\complement} \notin g(v)$. Since $\mathcal{G}$ subsumes $\mathcal{F}$, $S^{\complement} \cup (\cap f(v))^{\complement} \notin g(v)$. But this means that $(S^{\complement} \cup (\cap f(v))^{\complement})^{\complement} H_g^* v$. Distributing the outer complement, we get $(S \cap (\cap f(v))) H_g^* v$. By definition of R_f^*, we conclude that $(S \cap \{u | \, u R_f^* v\}) H_g^* v$.

□

Proposition 4.3. The reflexive-transitive reductions R_f, H_g of R_f^* and H_g^* are well-defined.

Proof. Since R_f^* is just a binary graph relation, its reflexive-transitive reduction exists and is unique (since R_f^* is acyclic). Similarly, by Proposition 1.24 the transitive reduction of H_g^* exists and is unique. □

Proposition 4.4. (Pacuit) For all $w \in W$,

$$S \in f(w) \quad \text{iff} \quad \{v | \, v R_f^* w\} \subseteq S$$

Proof. ($\rightarrow$) Suppose $S \in f(w)$, and let v be such that $v R_f^* w$. So $v \in \cap f(w)$, and so in particular $v \in S$.

($\leftarrow$) Now suppose $\{v | \, v R_f^* w\} \subseteq S$. Note that by definition of R_f^*, $\cap f(w) \subseteq \{v | \, v R_f^* w\}$ (any v in the core will be $v R_f^* w$). Since $\mathcal{F}$ is closed under finite intersections and supersets, $S \in f(w)$. □

Proposition 4.5. For all $w \in W$,

$$S \in g(w) \quad \text{iff} \quad \neg S^{\complement} H_g^* w$$

Proof. By definition, $S \in g(w)$ iff $\neg S \notin g(w)$ iff $\neg S^{\complement} H_g^* w$. □

[TODO: Clean up and streamline the hash function & its properties, etc. Make sure that the (Loop) property doesn't secretly show up here.]

Definition 4.6. 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 $\mathsf{hash}\colon \mathcal{P}(\{m_1, \ldots, m_k\}) \times \mathbb{N}^k \rightarrow \mathbb{N}$ be defined by

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

Definition 4.7. Let $\mathfrak{F} = \langle W, f, g \rangle$ be a preferential multi-frame, and let R_f, H_g be the reflexive-transitive reductions of R_f^*, H_g^*. Its **simulation net** $\mathcal{N}^\bullet = \langle N, E, W, A, O \rangle$ is the BFNN given by:

- $N = W$
- $E = R_f$

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.
- $A^{(n)}(\vec{x}, \vec{w}) = \mathsf{hash}(\{m_i | \, x_i = 1\}, \vec{w})$
- $O^{(n)}(x) = 1$ iff $\mathsf{hash}^{-1}(x)[0] H n$ (hash^{-1} is well-defined in special circumstances, see the next proposition)

Moreover, given a valuation function V, the **simulation net model** $\langle \mathcal{N}^\bullet, [\![\cdot]\!] \rangle$ based on $\mathcal{N}$ has for all propositions p:

$$[\![p]\!]_{\mathcal{N}^\bullet} = V(p)$$

PROPOSITION 4.8. $\mathsf{hash}(S,\vec{W}(m_i,n))\colon \mathcal{P}(\{m_1,\ldots,m_k\})\rightarrow P_k$, 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}).

Proof. 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 $\mathsf{hash}(S)=\prod_{m_i\in S}p_i=\prod_{(p_i\,\text{divides}\,x)}p_i=x$. □

PROPOSITION 4.9. $O^{(n)}\circ A^{(n)}$ is zero at zero and monotonically increasing.

Proof. First, note that $O^{(n)}(A^{(n)}(\vec{0},\vec{w}))=0$, since we cannot have $\mathsf{hash}^{-1}(\mathsf{hash}(\emptyset))=\emptyset\,Hn$. Now let $\vec{w}_1,\vec{w}_2$ be such that O is well-defined (i.e. are vectors of prime numbers), and suppose $\vec{w}_1<\vec{w}_2$. If $O^{(n)}(A^{(n)}(\vec{x},\vec{w_1}))=1$, then $\mathsf{hash}^{-1}(\mathsf{hash}(\vec{x},\vec{w_1}))[0]$ Hn. But this just means $\{m_i|\,x_i=1\}Hn$. And so $\mathsf{hash}^{-1}(\mathsf{hash}(\vec{x},\vec{w_2}))[0]\,Hn$. 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. □

LEMMA 4.10. $\mathsf{Reach}_{\mathcal{N}^\bullet}(S)=\{v|\,\exists u\in S\,\text{such that}\,u R_f^* v\}$.

Proof. For the ($\supseteq$) direction, suppose there is a $u\in S$ such that $u R_f^* v$ and proceed by induction on the length of this path. If the path has length 0, $v\in S$, and so $v\in\mathsf{Reach}(S)$. Otherwise, let u immediately precede v on this path. By inductive hypothesis $u\in\mathsf{Reach}(S)$. Since $(u,v)\in E$, $v\in\mathsf{Reach}(S)$.

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

Base step. $v\in S$. Since R_f^* is reflexive, $v R_f^* v$, and we are done.

Inductive step. There is $u\in\mathsf{Reach}(S)$ such that $(u,v)\in E$ (and so $(u,v)\in R_f^*$). By inductive hypothesis, there is a $t\in S$ such that $t R_f^* u$. Since R_f^* is transitive, $t R_f^* v$ as well. □

LEMMA 4.11. $\mathsf{Prop}_{\mathcal{N}^\bullet}(S)=\{n|\,S H_g^* n\}$

Proof. For the ($\supseteq$) direction, suppose $S H_g^* n$. By Proposition 1.23 there is a strong hyperpath from S to n in H_g. We show that $n\in\mathsf{Prop}_{\mathcal{N}^\bullet}(S)$ by induction on the length of this hyperpath:

Base step. We have a strong hyperpath of length 1, (S) with $n\in S$. So $n\in\mathsf{Prop}_{\mathcal{N}^\bullet}(S)$.

Inductive step. Say the hyperpath is $(S=S_1,S_2,\ldots,S_l)$, with $n\in S_l$. By definition we have

$$S_l=\{u|\,S_{l-1}\,H_g u\}$$

In particular, $S_{l-1}H_g n$. So $S_{l-1}H_g^* n$, and by [Todo] we have $(S_{l-1}\cap\{m_i|\,m_i R_f^* n\})H_g^* n$.

CLAIM. $(S_{l-1}\cap\{m_i|\,m_i R_f^* n\})H_g n$

Proof. We make use of Proposition 1.27 to move from H_g^*. Let $T=\{u|\,(S_{l-1}\cap\{m_i|\,m_i R_f^* n\})H_g u\}$. We need to check that $n\notin S_{l-1}\cap\{m_i|\,m_i R_f^* n\}$ and $\neg T H_g^* n$.

Check 1: $n\notin S_{l-1}$. Suppose for contradiction that $n\in S_{l-1}\cap\{m_i|\,m_i R_f^* n\}$. Then in particular, $n\in S_{l-1}$. But $S_{l-1}H_g n$, which contradicts the fact that H_g is irreflexive.

Check 2: $\neg T H_g^* n$. Suppose for contradiction that $T H_g^* n$. [Todo – $n\in T$, and so it loops back on itself] □

CLAIM. $S_{l-1} \cap \{m_i | \, m_i R_f^* n\} = S_{l-1} \cap \{m_i | \, m_i R_f n\}$

Proof. $(\supseteq)$ This direction is easy.

$(\subseteq)$ Suppose $u \in S_{l-1}$ and $u R_f^* n$. This time, we make use of Proposition 1.26 to move from R_f^* to R_f. We need to check that $u \neq n$ and there is no $x \neq n \in N$ such that $u R_f x, x R_f^* n$.

Check 1: $u \neq n$. Suppose for contradiction that $u = n$. So $n \in S_{l-1}$ and $S_{l-1} H_g n$, which contradicts the fact that H_g is irreflexive.

Check 2: There is no such x. Suppose for contradiction that we have x such that $u R_f x, x R_f^* n$. [Todo – $u \in S_{l-1}$, and $n \in S_l$. Think about it. Try to prove that $x = n$, which contradicts $x \neq n$.] ☐

From here, we have

$$\begin{array}{lll} & (S_{l-1} \cap \{m_i | \, m_i R_f n\}) H_g n & \\ (\rightarrow) & \{m_i | \, m_i \in S_{l-1} \text{ and } (m_i, n) \in E\} H_g n & \text{(by choice of } E) \\ (\rightarrow) & \{m_i | \, m_i \in \mathsf{Prop}_{\mathcal{N}^\bullet}(S) \text{ and } (m_i, n) \in E\} H_g n & \text{(by Inductive Hypothesis)} \\ (\rightarrow) & \mathsf{hash}^{-1}(\mathsf{hash}(\vec{\chi}_{\mathsf{Prop}_{\mathcal{N}^\bullet}(S)}(m_i), \vec{W}(m_i, n)))[0] \, H_g \, n & \text{(by definition of } \mathsf{hash}) \\ (\rightarrow) & O^{(n)}(A^{(n)}(\vec{\chi}_{\mathsf{Prop}_{\mathcal{N}^\bullet}(S)}(m_i), \vec{W}(m_i, n))) = 1 & \text{(by choice of } O \text{ and } A) \end{array}$$

By definition, $n \in \mathsf{Prop}_{\mathcal{N}^\bullet}(S)$.

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

Base step. $n \in S$. Since H_g^* is reflexive, $S H_g^* n$.

Inductive step. Let $m_1, \ldots, m_k$ list those nodes such that $(m_i, n) \in E$. We have

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

Let $T = \{u | \, S H_g^* u\}$. By our inductive hypothesis,

$$O^{(n)}(A^{(n)}(\vec{\chi}_T(m_i), \vec{W}(m_i, n))) = 1$$

By choice of O and A,

$$\mathsf{hash}^{-1}(\mathsf{hash}(\vec{\chi}_T(m_i), \vec{W}(m_i, n)))[0] \, H_g \, n$$

i.e. $T H_g n$ (and so $T H_g^* n$). Since H_g^* is transitive, $S H_g^* n$.

☐

THEOREM 4.12. Let $\mathcal{M}$ be a model based on a preferential multi-frame $\mathfrak{F}$, and let $\langle \mathcal{N}^\bullet, [\![\cdot]\!] \rangle$ be the corresponding simulation net model. We have

$$\mathcal{M}, w \vDash \varphi \quad \text{iff} \quad w \in [\![\varphi]\!]_{\mathcal{N}^\bullet}$$

Proof. By induction on φ.

***p* case:**

$$\begin{array}{rll} \mathcal{M}, w \vDash p & \text{iff} & w \in V(p) \\ & \text{iff} & w \in [\![p]\!]_{\mathcal{N}^\bullet} \end{array}$$

$\neg\varphi$ case:

$$\begin{array}{rlll} \mathcal{M}, w \vDash \neg\varphi & \text{iff} & \mathcal{M}, w \nvDash \varphi & \\ & \text{iff} & w \notin [\![\varphi]\!]_{\mathcal{N}^\bullet} & \text{(IH)} \\ & \text{iff} & w \in [\![\neg\varphi]\!]_{\mathcal{N}^\bullet} & \end{array}$$

$\varphi \wedge \psi$ case:

$$\begin{array}{rlll} \mathcal{M}, w \vDash \varphi \wedge \psi & \text{iff} & \mathcal{M}, w \vDash \varphi \text{ and } \mathcal{M}, w \vDash \psi & \\ & \text{iff} & w \in [\![\varphi]\!]_{\mathcal{N}^\bullet} \text{ and } w \in [\![\psi]\!]_{\mathcal{N}^\bullet} & \text{(IH)} \\ & \text{iff} & w \in [\![\varphi \wedge \psi]\!]_{\mathcal{N}^\bullet} & \end{array}$$

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

$$\begin{array}{llll}
\mathcal{M}, w \vDash \langle \mathbf{K} \rangle \varphi & \text{iff} & \{u \mid \mathcal{M}, u \nvDash \varphi\} \notin f(w) & \text{(by definition)} \\
& \text{iff} & \{u \mid u \notin [\![\varphi]\!]_{\mathcal{N}^\bullet}\} \notin f(w) & \text{(IH)} \\
& \text{iff} & \{u \mid u R_f^* w\} \nsubseteq \{u \mid u \notin [\![\varphi]\!]_{\mathcal{N}^\bullet}\} & \text{(by Proposition 4.4)} \\
& \text{iff} & \exists u \text{ such that } u R_f^* w \text{ and } u \in [\![\varphi]\!]_{\mathcal{N}^\bullet} & \\
& \text{iff} & w \in \mathsf{Reach}_{\mathcal{N}^\bullet}([\![\varphi]\!]) & \text{(by Lemma 4.10)} \\
& \text{iff} & w \in [\![\langle \mathbf{K} \rangle \varphi]\!]_{\mathcal{N}^\bullet} & \text{(by definition)}
\end{array}$$

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

$$\begin{array}{llll}
\mathcal{M}, w \vDash \langle \mathbf{T} \rangle \varphi & \text{iff} & \{u \mid \mathcal{M}, u \nvDash \varphi\} \notin g(w) & \text{(by definition)} \\
& \text{iff} & \{u \mid u \notin [\![\varphi]\!]_{\mathcal{N}^\bullet}\} \notin g(w) & \text{(IH)} \\
& \text{iff} & [\![\varphi]\!]^{\complement}_{\mathcal{N}^\bullet} \notin g(w) & \\
& \text{iff} & [\![\varphi]\!]_{\mathcal{N}^\bullet} H_g^* w & \text{(by Proposition 4.5)} \\
& \text{iff} & w \in \mathsf{Prop}_{\mathcal{N}^\bullet}([\![\varphi]\!]) & \text{(by Lemma 4.11)} \\
& \text{iff} & w \in [\![\langle \mathbf{T} \rangle \varphi]\!]_{\mathcal{N}^\bullet} & \text{(by definition)}
\end{array}$$

□

Corollary 4.13. $\mathcal{M} \vDash \varphi$ iff $\mathcal{N}^\bullet \vDash \varphi$.

Proof.

$$\begin{array}{lll}
\mathcal{M} \vDash \varphi & \text{iff} & \mathcal{M}, w \vDash \varphi \text{ for all } w \in W = N \\
& \text{iff} & w \in [\![\varphi]\!]_{\mathcal{N}^\bullet} \text{for all } w \in N \\
& \text{iff} & [\![\varphi]\!]_{\mathcal{N}^\bullet} = N \\
& \text{iff} & \mathcal{N}^\bullet \vDash \varphi
\end{array}$$

□

5 Soundness and Completeness

[Soundness is straightforward, it's just a check that the properties hold for all models based on the frame.]

[Corollary: Soundness of neural semantics!]

[[The plan for completeness: High-level describe that we can do the canonical model construction (from, e.g., Pacuit, see page 65, second to last paragraph) & Lindenbaum Lemma stuff for our particular logic. [actually state canonical model definition] We then, as usual, prove the truth lemma [actually state truth lemma] The last thing to show is that our canonical model's frame is a preferential frame, i.e. satisfies all the right properties [then put it all together]]]

[Strong completeness follows straightforwardly from weak completeness]

[Corollary: Strong completeness for neural semantics!]

6 Dynamics of Neural Network Update

References

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

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

[3] Mayur Thakur and Rahul Tripathi. Linear connectivity problems in directed hypergraphs. *Theoretical Computer Science*, 410(27-29):2592–2618, 2009.