# Meeting Notes 12/30/2022

**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 $\Gamma$, we can *build* a neural network $\mathcal{N}$ satisfying $\Gamma$.

**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.

## 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 $n$)
- $V : \text{propositions} \to \mathcal{P}(N)$ is an assignment of propositional variables to sets of neurons (the **valuation function**)

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 $\mathcal{N}$, $\mathsf{Set} = \mathcal{P}(N) = \{S | S \subseteq N\}$

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. 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. (LEITGEB) Let $\mathcal{N} \in \mathsf{Net}$. For all $S, S_1, S_2 \in \mathsf{Set}$, $\mathsf{Prop}$ 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 $S, S_1, S_2 \in \mathsf{Set}$, $n, m \in N$, $\mathsf{Reach}$ 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 $n \in N$, $\mathsf{Reach}^{-1}(n) = \bigcap_{n \notin \mathsf{Reach}(X)} X^{\complement}$

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

PROPOSITION 1.11. $\mathsf{Reach}^{-1}$ 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

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

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

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

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

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 $n \in N$.

# 2 Neighborhood Models

## 2.1 Basic Definitions

DEFINITION 2.1. A **neighborhood frame** is a pair $\mathcal{F} = \langle W, f \rangle$, where $W$ is a non-empty set of **worlds** and $f : W \to \mathcal{P}(\mathcal{P}(W))$ is a **neighborhood function**.

DEFINITION 2.2. A **multi-frame** is $\mathfrak{F} = \langle W, f, g \rangle$, where $f$ and $g$ are neighborhood functions.

DEFINITION 2.3. 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 2.4. 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 **acyclic** iff 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 $u_i = u_j$.
- $\mathcal{F}$ **guides** $\mathcal{G}$ iff for all $w \in W$, if $X \cup (\cap f(w))^{\mathsf{C}} \in g(w)$ then $X \in g(w)$.

DEFINITION 2.5. 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, acyclic, proper filter,
  - $\mathcal{G} = \langle W, g \rangle$ is reflexive, transitive, and $\mathcal{F}$ guides $\mathcal{G}$.

PROPOSITION 2.6. (PACUIT) If $\mathcal{F} = \langle W, f \rangle$ is a filter, and $W$ is finite, then $\mathcal{F}$ contains its core.

PROPOSITION 2.7. If $\mathcal{F} = \langle W, f \rangle$ is a proper filter, then for all $w \in W$, $Y^{\mathsf{C}} \in f(w)$ iff $Y \notin f(w)$.

## 2.2 Neighborhood Semantics

DEFINITION 2.8. 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 2.9. 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:

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

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

# 3 From Nets to Frames

**This is the easy ("soundness") direction!**

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

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

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

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

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

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

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

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

**Note.** This is the first big result: Given a neural network $\mathcal{N}$, we can build an equivalent "classical" model $\mathcal{M}^\bullet$. This $\mathcal{M}^\bullet$ is in fact a preferential multi-frame — each of the frame properties follows straightforwardly from the corresponding properties of Reach and Prop.

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

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

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

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

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

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

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

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

- $W(m_i, n) = p_i$, the $i$th 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.

**Note.** This is where we use the fact that $\mathcal{F}$ is **acyclic** and **contains the unit**.

LEMMA 4.5. $\mathsf{Reach}_{\mathcal{N}^\bullet}(S) = \{v | \exists u \in S \text{ such that } u \in \cap f(v)\}$.

**Note.** This is where we use the fact that $\mathcal{F}$ is **reflexive** and **transitive**.

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

**Note.** This is the lemma that really gave me a hard time. The proof is in the appendix — we should go over this one *very* carefully.

This is where we use the fact that $\mathcal{G}$ is **reflexive**, **transitive**, and $\mathcal{F}$ **guides** $\mathcal{G}$.

THEOREM 4.7. Let $\mathcal{M}$ be a model based on a preferential multi-frame $\mathfrak{F}$, and let $\mathcal{N}^\bullet$ be the corresponding simulation net. We have, for all $w \in W$,

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

COROLLARY 4.8. $\mathcal{M} \models \varphi$ iff $\mathcal{N}^\bullet \models \varphi$.

**Note.** Finally, we have the second big result: Given a "classical" model $\mathcal{M}$, we can build an equivalent feed-forward neural network $\mathcal{N}^\bullet$. The proof of this follows straightforwardly from the previous two lemmas (which I had to work for!).

This means that given a set of sentences $\Gamma$, if we can build a model $\mathcal{M} \models \Gamma$, then we can build $\mathcal{N}^\bullet \models \Gamma$ — i.e. we can build neural networks satisfying specific constraints!

P.S. this is where we use the fact that $\mathcal{F}$ is a **proper filter**, i.e. closed under subset and finite intersection. I should probably put this in a lemma instead. I don't actually use the fact that $\mathcal{F}$ doesn't contain the empty set, though…

## 5 What's Left To Do

**Completeness.** We just need to give axioms that characterize "preferential multi-frames." The sticking point is the property:

$$\text{for all } w \in W, \text{ if } X \cup (\cap f(w))^{\mathsf{C}} \in g(w) \text{ then } X \in g(w)$$

**[hebb]* Extension.** The whole point of this is was to make a language we could use to reduce [hebb]* I have a reduction of [hebb]* that I'd like to double-check. (i.e. characterizes what we learn in limit)

RESULTS ABOVE THIS LINE WOULD MAKE A GREAT PAPER

**Stable [hebb]*.** I should consider if it's possible to axiomatize stable Hebbian learning (i.e. can our language express convergence??? Important to know before backprop). I can run some computer simulations to see what these updates look like.

**Single-Step [hebb].** In the FLAIRS paper we used [hebb] instead of [hebb]*. Is there a way to get an axiomatization of [hebb] starting from [hebb]*? This would give us the full completeness result I wanted in the first place.

**Fuzzy Nets.** Show that we still get soundness & completeness when we upgrade to fuzzy nets. We need a lemma like:

For all $\varphi$ over this (note: non-fuzzy) language $\mathcal{L}$,

$$\mathcal{N}_{\text{bin}}, n \Vdash \varphi \quad \text{iff} \quad \mathcal{N}_{\text{fuzzy}}, n \Vdash \varphi$$

**Recurrent Nets.** See what properties Prop has in a recurrent net, and see how the logic needs to change in response. (What does relaxing the 'acyclic' condition do?) I should run some computer simulations to see what these propagations look like.

RESULTS ABOVE THIS LINE WOULD MAKE GREAT ADDITIONS TO A JOURNAL PAPER SYNTHESIZING ALL OF THIS

## Appendix A Proofs of Lemmas

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

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

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 $n\notin\mathsf{Reach}(X)$, $u\in X^{\complement}$. Consider in particular

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

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 $u\in\bigcap_{n\notin\mathsf{Reach}(X)}X^{\complement}=\mathsf{Reach}^{-1}(n)$. □

**Proof. (of Proposition 1.11)** Suppose $n_1\in\mathsf{Reach}^{-1}(n_2),\dots,n_{k-1}\in\mathsf{Reach}^{-1}(n_k),n_k\in\mathsf{Reach}^{-1}(n_1)$. 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 $n\in\bigcap_{n\notin\mathsf{Reach}(X)}X^{\complement}=\mathsf{Reach}^{-1}(n)$ (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 $m_1,\dots,m_k$ 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,

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

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 2.7)** ($\rightarrow$) 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 $\cap f(w) \not\subseteq Y$ and $\cap f(w) \not\subseteq Y^{\mathsf{C}}$. But this means $\cap f(w) \not\subseteq Y \cap Y^{\mathsf{C}} = \emptyset$, 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 Theorem 3.2)** By induction on $\varphi$. The propositional, $\neg\varphi$, and $\varphi \wedge \psi$ cases are trivial.

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

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

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

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

□

**Proof. (of Theorem 3.4)** We show each in turn:

- **$\mathcal{F}$ is closed under finite intersection:** Suppose $X_1, \ldots, X_n \in f(w)$. By definition of $f$, $w \notin \bigcup_i \mathsf{Reach}(X_i^{\mathsf{C}})$ 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 $\bigcap_i X_i \in f(w)$.
- **$\mathcal{F}$ 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.
- **$\mathcal{F}$ contains the unit:** Note that for all $w \in W$, $w \notin \mathsf{Reach}(\emptyset) = \mathsf{Reach}(W^{\mathsf{C}})$. So $W \in f(w)$.
- **$\mathcal{F}$ is reflexive:** We want to show that $w \in \cap f(w)$. Well, suppose $X \in f(w)$, i.e. $w \notin \mathsf{Reach}(X^{\mathsf{C}})$ (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.
- **$\mathcal{F}$ is transitive:** Suppose $X \in f(w)$, i.e. $w \notin \mathsf{Reach}(X^{\mathsf{C}})$. Well,

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

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

- **$\mathcal{F}$ 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 $X^{\mathsf{C}}$ for $X$ we get $u_i \in \bigcap_{u_{i+1} \notin \mathsf{Reach}(X)} X^{\mathsf{C}}$. In other words, $u_1 \in \mathsf{Reach}^{-1}(u_2),\dots,u_{n-1} \in \mathsf{Reach}^{-1}(n), u_n \in \mathsf{Reach}^{-1}(u_1)$. By Proposition 1.11, each $u_i = u_j$.
- **$\mathcal{G}$ 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).
- **$\mathcal{F}$ guides $\mathcal{G}$:** 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 $Y^{\mathsf{C}}$ for $Y$, we get $w \notin \mathsf{Prop}(X^{\mathsf{C}} \cap (\bigcap_{w \notin \mathsf{Reach}(Y)} Y^{\mathsf{C}}))$. By definition of $\mathsf{Reach}^{-1}$, $w \notin \mathsf{Prop}(X^{\mathsf{C}} \cap \mathsf{Reach}^{-1}(w))$. From (Minimal Cause), we conclude that $w \notin \mathsf{Prop}(X^{\mathsf{C}})$, i.e. $X \in g(w)$.

□

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

**Proof. (of Proposition 4.4)** 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 $u_1,\dots,u_n \in N$ such that $u_1 E u_2,\dots,u_{n-1} E u_n, u_n E u_1$. Then we have $u_1 \in \cap f(u_2),\dots,u_{n-1} \in \cap f(u_{n-1}), u_n \in \cap f(u_1)$, which contradicts the fact that $\mathcal{F}$ 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 $\mathcal{F}$ contains the unit.

**$O^{(n)} \circ A^{(n)}$ is monotonically increasing.** 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 $O^{(v)}(A^{(v)}(\vec{x}, \vec{w_1})) = 1$, 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 $g(v)$. □

**Proof. (of Lemma 4.5)** For the $(\supseteq)$ direction, let $u \in S$ be such that $u \in \cap f(v)$. By definition of $E$, $(u, v) \in E$. And since $u \in S$, $u \in \mathsf{Reach}_{\mathcal{N}^{\bullet}}(S)$. By the constructor 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.** $v \in S$. Since $\mathcal{F}$ is reflexive, $v \in \cap f(v)$, and we are done.

**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, there is a $t \in S$ such that $t \in f(u)$. We are done if we can show that this $t \in S$ is also $t \in \cap f(v)$. So let $X \in f(v)$ — we want to show that $t \in X$.

Since $\mathcal{F}$ is transitive, $\{y | X \in f(y)\} \in f(v)$. But by definition of core, $\cap f(v) \subseteq \{y | X \in f(y)\}$. So, since $u \in \cap f(v)$, $X \in f(u)$. But this means that $\cap f(u) \subseteq X$, and since $t \in f(u)$, we get $t \in X$. □

**Proof. (of Lemma 4.6)** For the $(\supseteq)$ direction, suppose $S^{\mathsf{C}} \notin g(v)$. Since $\mathcal{F}$ guides $\mathcal{G}$, 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 $S \cap (\cap f(v)) = \{u | u \in S \text{ and } (u, v) \in E\} = \mathsf{hash}^{-1}(\mathsf{hash}(\overrightarrow{\chi}_{\mathsf{Prop}_{\mathcal{N}^{\bullet}}(S)}(u), \vec{W}(u, v)))[0]$, 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. $T^{\mathsf{C}} \notin g(v)$. We would like to show that $S^{\mathsf{C}} \notin g(v)$. Suppose for contradiction that $S^{\mathsf{C}} \in g(v)$. Recall that $T = \{u_i | S^{\mathsf{C}} \notin g(u_i)\}$, i.e. $T^{\mathsf{C}} = \{u_i | S^{\mathsf{C}} \in g(u_i)\}$. Since $S^{\mathsf{C}} \in g(v)$ and $\mathcal{G}$ is transitive, $T^{\mathsf{C}} \in g(v)$, which contradicts $T^{\mathsf{C}} \notin g(v)$.

□

**Proof. (of Theorem 4.7)** By induction on $\varphi$. Again, the propositional, $\neg\varphi$, and $\varphi \wedge \psi$ cases are trivial.

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

| | | | |
|---|---|---|---|
| $\mathcal{M}, w \Vdash \langle \mathbf{K} \rangle \varphi$ | iff | $\{u \mid \mathcal{M}, w \not\Vdash \varphi\} \notin f(w)$ | (by definition) |
| | iff | $\{u \mid u \notin [\![\varphi]\!]_{\mathcal{N}^\bullet}\} \notin f(w)$ | (Inductive Hypothesis) |
| | iff | $[\![\varphi]\!]_{\mathcal{N}^\bullet} \in f(w)$ | (by Proposition 2.7) |
| | iff | $\exists u \in [\![\varphi]\!]_{\mathcal{N}^\bullet}$ such that $u \in \cap f(w)$ | (since $\mathcal{F}$ is a proper filter) |
| | iff | $w \in \mathsf{Reach}_{\mathcal{N}^\bullet}([\![\varphi]\!])$ | (by Lemma 4.5) |
| | iff | $w \in [\![\langle \mathbf{K} \rangle \varphi]\!]_{\mathcal{N}^\bullet}$ | (by definition) |
| | iff | $\mathcal{N}^\bullet, w \Vdash \langle \mathbf{K} \rangle \varphi$ | (by definition) |

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

| | | | |
|---|---|---|---|
| $\mathcal{M}, w \Vdash \langle \mathbf{T} \rangle \varphi$ | iff | $\{u \mid \mathcal{M}, u \not\Vdash \varphi\} \notin g(w)$ | (by definition) |
| | iff | $\{u \mid u \notin [\![\varphi]\!]_{\mathcal{N}^\bullet}\} \notin g(w)$ | (Inductive Hypothesis) |
| | iff | $[\![\varphi]\!]^{\mathsf{C}}_{\mathcal{N}^\bullet} \notin g(w)$ | |
| | iff | $w \in \mathsf{Prop}_{\mathcal{N}^\bullet}([\![\varphi]\!])$ | (by Lemma 4.6) |
| | iff | $w \in [\![\langle \mathbf{T} \rangle \varphi]\!]_{\mathcal{N}^\bullet}$ | (by definition) |
| | iff | $\mathcal{N}^\bullet, w \Vdash \langle \mathbf{T} \rangle \varphi$ | (by definition) |

□