Algorithm: Given a board state x and a value network $v(\cdot)$, the algorithm optimizes $\{I_{\text{and}}(S)\}_{S\subseteq N}$ and $\{I_{\text{or}}(S)\}_{S\subseteq N}$ for all subsets $S \subseteq N$. Then, we extract interaction primitives , select common coalitions T based on interaction primitives, and compute the coalition attributions $\varphi(T)$.

Input: a board state x, a value network $v(\cdot)$.

Output: Interactions $\{I_{\text{and}}(S)\}_{S\subseteq N}$ and $\{I_{\text{or}}(S)\}_{S\subseteq N}$ for all subsets $S \subseteq N$, the attribution $\varphi(T)$ of each coalition T.

1. Initialize parameters as $a_k = \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{T\subseteq N:\Delta n(T)=k}\log(\frac{p_{\text{white}}(\boldsymbol{x}_T)}{1-p_{\text{white}}(\boldsymbol{x}_T)}), k \in \{-\frac{n}{2}, -\frac{n}{2}+1, \ldots, \frac{n}{2}\}. \forall T \subseteq N, p_T = q_T = 0.$
2. **for** $S \subseteq N$ **do**

   compute initial $I_{\text{and}}(S)$ and $I_{\text{or}}(S)$ by setting $\forall T \subset N, v_{\text{and}}(\boldsymbol{x}_T) = \frac{1}{2}\cdot[u(\boldsymbol{x}_T) + q_T] + p_T$ and $v_{\text{or}}(\boldsymbol{x}_T) = \frac{1}{2}\cdot[u(\boldsymbol{x}_T) + q_T] - p_T$, subject to $u(\boldsymbol{x}_T) = v(\boldsymbol{x}_T) - a_k$, according to Equation (1) and Equation (3).

   end for
3. Iteratively update parameters $\boldsymbol{a} = \{a_{-\frac{n}{2}}, a_{-\frac{n}{2}+1}, \ldots, a_{\frac{n}{2}}\}, \{p_T\}_{T\subseteq N}, \{q_T\}_{T\subseteq N}$ via $\min_{\boldsymbol{a},\{p_T\}_{T\subseteq N},\{q_T:|q_T|<\tau\}_{T\subseteq N}} \|\boldsymbol{I}_{\text{and}}\|_1 + \|\boldsymbol{I}_{\text{or}}\|_1$
4. Determine a set of salient interaction primitives $\Omega_{\text{salient}} = \{S : |I(S)| > \xi\}$, where $\xi = 0.15 \cdot \max_S |I(S)|$.
5. Manually annotate 50 common coalitions T based on interaction primitives Ω_{salient}.
6. Compute coalition attributions $\varphi(T) = \sum_{S\supseteq T}\frac{|T|}{|S|}[I_{\text{and}}(S) + I_{\text{or}}(S)]$ for each annotated coalition T.