Jeroen's model

June 10, 2013

1 Notation

- Σ: partially-ordered list of symbols (types).

- T: number of symbols in grammar.

- $\mathcal{P}$: pose space.

- $\mathcal{B} = (b_1, \ldots, b_N)$: set of bricks.

- $\mathcal{P}(b) \in \mathcal{P}$: set of poses associated with brick b.

- $\mathcal{A} = (a_1, \ldots, a_M)$, $\mathcal{A} \subseteq \mathcal{B}$: set of active bricks.

- $\mathcal{A}_m = (a_1, \ldots, a_m)$, $m \leq M$ $\mathcal{A}_m \subseteq \mathcal{A}$: subset of active bricks.

- $t(b)$: type of brick b.

- $orph(\mathcal{A}_m)$: the set of bricks in $\mathcal{A}_m$ that are orphans (*ie*, do not have a parent).

- $\#(C)$: the cardinality (size) of some set C. *E.g.* $\#(\mathcal{A}) = M$.

- $\mathcal{R}$: set of production rules of the form $A \rightarrow B_1, \ldots, B_k$ with $r = A, B_1, \ldots, B_k \in \Sigma$. $\mathcal{R}$ is acyclic.

- $n(b)$: number of slots/children associated with bricks of type $t(b)$.

- $c(b)$: the set of production rules with $t(b)$ in the LHS.

- $\vec{\theta}_b \in \mathbb{R}^{\#(c(b))}$: distribution over rules with $t(b) \in \mathcal{B}$ in the LHS.

- $s_b \in \{0,1\}$: on/off state of brick b.

- $r_b \in \mathcal{R}(t(b))$: rule used by brick b.

- $\mathbf{g_b} \in (\mathcal{B} \cup \perp)^{n(b)}$: array of pointers, where each element can either point to a brick, or null (no child).

- $\vec{q}_b \in \{\mathcal{P}(b), \perp\}$: pose for brick b. $\vec{q}_b = \perp$ is a "nothing" pose that does not contribute to image evidence.

- $\mathbf{q}^{\mathcal{A}_m}$: poses for bricks $a \in \mathcal{A}_m$.

- $\mathbf{s}^{\mathcal{A}_m}$: on/off state for bricks $a \in \mathcal{A}_m$.

- $\mathbf{r}^{\mathcal{A}_m}$: rules for bricks $a \in \mathcal{A}_m$.

- $\mathbf{g}_b^{\mathcal{A}_m} \in (\mathcal{A}_m \cup \emptyset)^{n(b)}$: array of pointers, where each element can either point to a brick in the active set defined by $\mathcal{A}_m$, or blank (child not yet specified) denoted by $\emptyset$. Note that blank is **different** than null (no child) in that blank means "there could be a child in this slot, but we don't know if there is or which one yet" while null (no child) means "there definitely isn't a child in this slot".

- $comp(\mathbf{g}^{\mathcal{A}_m})$: the elements of $\mathbf{g}^{\mathcal{A}_m}$ for which $g_{b,k}^{\mathcal{A}_m} = \emptyset$. Note: we use "comp" as the identifier for this concept since these are the slots that may receive a child in the future; these are the slots to be "completed".

- $V(b)$: "before" set of bricks, which we define as $V(b) = \{b' : t(b') < t(b)\}$ where $<$ means "comes before in the partial ordering defined by Σ".

- $W(b)$: "after" set of bricks, which we define as $V(b) = \{b' : t(b') > t(b)\}$ where $>$ means "comes after in the partial ordering defined by Σ".

- $\vec{\phi}_{b,k,r} \in \mathbb{R}^{\#(W(b))+1}$: distribution over bricks to which brick b's kth slot points to if brick b is using rule r. Note that brick b may only point to bricks in its "after" set. Note that the $+1$ is present since a brick/slot may point to nothing, denoted by the symbol $\perp$. *edit*: this is a bit sloppy. Each slot has a particular type, and only bricks of that type may go in this

slot. The set of possibilities for each slot, in general, is a subset of the "after" set, and not the entire after set.

- $\vec{\pi} \in \mathbb{R}^T$: vector of mixing coefficients for the templates of each type of brick.

- $\mathbf{X} = \{\mathbf{s}, \mathbf{r}, \mathbf{g}, \mathbf{q}\}$: collection of hidden variables of model.

- $\mathbf{X}^{\mathcal{A}_m} = \{\mathbf{s}^{\mathcal{A}_m}, \mathbf{r}^{\mathcal{A}_m}, \mathbf{g}^{\mathcal{A}_m}, \mathbf{q}^{\mathcal{A}_m}\}$: collection of hidden variables for bricks $a \in \mathcal{A}_m$.

- $\mathbf{X}_{-r}^{\mathcal{A}_m} = \{\mathbf{s}^{\mathcal{A}_m}, \mathbf{r}^{\mathcal{A}_m}, \mathbf{g}^{\mathcal{A}_m}, \mathbf{q}^{\mathcal{A}_m}\}$: collection of hidden variables for bricks $a \in \mathcal{A}_m$ except the hidden variables representing rules used by the bricks.
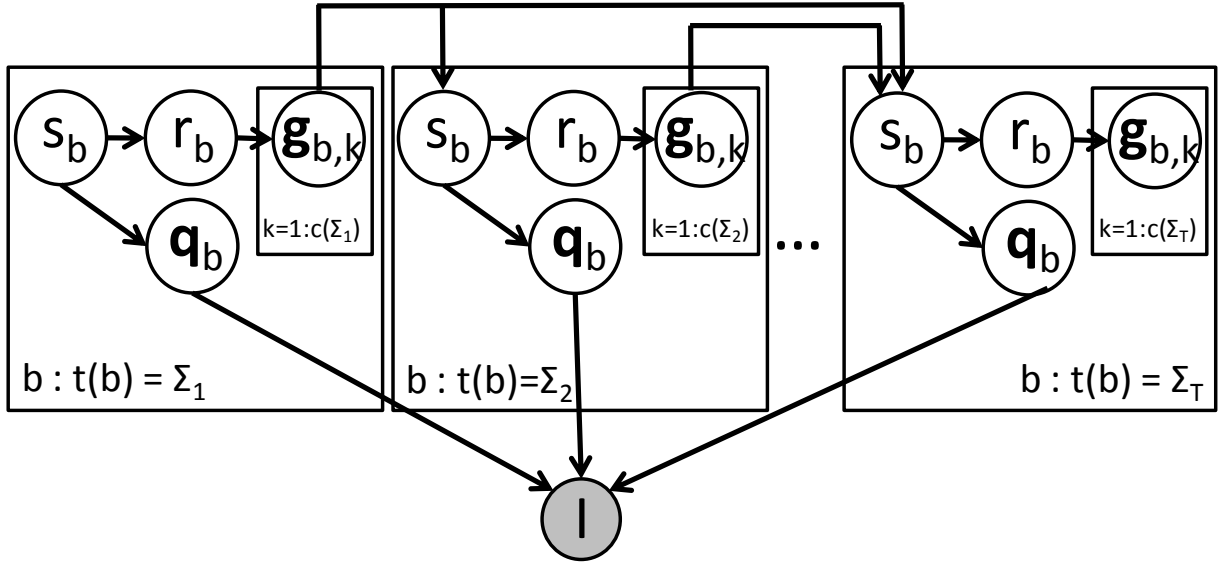
- I: the image.

2   Model

Figure 1: Graphical model

The decomposition of the joint probability is given by the graphical model in Fig. 1 and by:

$$P(\mathbf{X}, I) \quad = \quad \Big(\prod_{b \in \mathcal{B}} P(s_b \mid \mathbf{g}_{V(b)}) \Big) \tag{1}$$

$$\times \Big(\prod_{b \in \mathcal{B}} P(r_b \mid s_b) \prod_{k=1}^{n(b)} P(g_{b,k} \mid r_b) \Big)$$

$$\times \Big(\prod_{b \in \mathcal{B}} P(\vec{q}_b \mid s_b) \Big)$$

$$\times \big(P(I \mid \mathbf{q}) \big)$$

where we define:

$$P(s_b = 1 \mid \mathbf{g}_{V(b)}) = \begin{cases} 1, & \text{if } \exists \ \{a,k\} \text{ s.t. } g_{a,k} = b \ , \ g_{a,k} \in g_{V(b)} \tag{2a} \\[2ex] \epsilon, & \text{if } \nexists \ \{a,k\} \text{ s.t. } g_{a,k} = b \ , \ g_{a,k} \in g_{V(b)} \tag{2b} \end{cases}$$

$$P(r_b \mid s_b) \quad = \quad Discrete(\vec{\theta}_b) \tag{3}$$

$$P(g_{b,k} \mid r_b) \quad = \quad Discrete(\vec{\phi}_{b,k,r_b}) \tag{4}$$

$$P(\vec{q}_b = \vec{q} \mid s_b)) \begin{cases} \dfrac{1}{\#\mathcal{P}(b)}, & \text{if } s_b = 1, \ \vec{q} \in \mathcal{P}(b) \tag{5a} \\[2ex] 0, & \text{if } s_b = 1 \text{ and } \vec{q} = \perp \tag{5b} \\[2ex] 0, & \text{if } s_b = 0, \ \vec{q} \in \mathcal{P}(b) \tag{5c} \\[2ex] 1, & \text{if } s_b = 0 \text{ and } \vec{q} = \perp \tag{5d} \end{cases}$$

$$P(I \mid \mathbf{q}) \quad = \quad \prod_p P(I_p \mid \mathbf{q}) \tag{6}$$

$$P(I_p \mid \mathbf{q}) \quad \propto \quad \sum_b \pi_{t(b)} P(I_p \mid \vec{q}_b) \delta(\text{template overlaps with pixel } p) \tag{7}$$

$$P(I_p \mid \vec{q}_b) \quad = \quad \text{defn here, with Bernoulli distribution and stuff.} \tag{8}$$

where $\delta(\cdot)$ is an indicator variable.

Localization

We define localizations in terms of the components $\Big(\prod_{b \in \mathcal{B}} P(s_b \mid \mathbf{g}_{V(b)}) \Big)$, $\Big(\prod_{b \in \mathcal{B}} P(r_b \mid s_b) \prod_{k=1}^{n(b)} P(g_{b,k} \mid r_b) \Big)$, $\Big(\prod_{b \in \mathcal{B}} P(\vec{q}_b \mid s_b) \Big)$, and $P(I \mid \mathbf{q})$ separately.

$$\Big(\prod_{b \in \mathcal{B}} P(s_b \mid \mathbf{g}_{V(b)}) \Big) \quad \Rightarrow \quad \Big(\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) \Big) \tag{9}$$

where $\Rightarrow$ is used to mean "localiizes to", and we evaluate $P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m})$ in a similar fashion to $P(s_b \mid \mathbf{g}_{V(b)})$ outlined in Eqn 2. We note that $P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) \neq P(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m})$ since we do not perform the "correct" marginalizations.

Similarily,

$$\left(\prod_{b \in \mathcal{B}} P(r_b \mid s_b) \prod_{k=1}^{n(b)} P(g_{b,k} \mid r_b) \right) \; \Rightarrow \; \left(\prod_{a \in \mathcal{A}_m} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \right) \tag{10}$$

$$P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \;=\; \begin{cases} 1 - \sum_{a' \in \mathcal{A}_m} P(g_{a,k} = a' \mid r_a) \text{ if } g_{a,k} = \emptyset \\ P(g_{a,k} \mid r_a) \text{ , otherwise} \end{cases} \tag{11}$$

$$\left(\prod_{b \in \mathcal{B}} P(\vec{q_b} \mid s_b) \right) \Rightarrow \left(\prod_{a \in \mathcal{A}_m} P(\vec{q_a} \mid s_a) \right) \tag{12}$$

$$P(I \mid \mathbf{q}) \Rightarrow P^{\mathcal{A}_m}(I \mid \mathbf{q}^{\mathcal{A}_m}) \tag{13}$$

where we evaluate $P^{\mathcal{A}_m}(I \mid \mathbf{q}^{\mathcal{A}_m})$ in a similar fashion to $P(I \mid \mathbf{q})$ outlined in Eqn. 6. We note that $P^{\mathcal{A}_m}(I \mid \mathbf{q}^{\mathcal{A}_m}) \neq P(I \mid \mathbf{q}^{\mathcal{A}_m})$ since we do not perform the correct marginalizations.

We define the overall localized distribution, $P^{\mathcal{A}_m}(X^{\mathcal{A}_m}, I)$, in terms of the previously defined localization distributions:

$$\begin{aligned} P^{\mathcal{A}_m}(X^{\mathcal{A}_m}, I) \;=\; & \left(\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) \right) \\ & \times \left(\prod_{a \in \mathcal{A}_m} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \right) \\ & \times \prod_{a \in \mathcal{A}_m} P(\vec{q_a} \mid s_a) \\ & \times \left(P^{\mathcal{A}_m}(I \mid \mathbf{q}^{\mathcal{A}_m}) \right) \end{aligned} \tag{14}$$

State representation

Representing the configuration of an image for a **non-localized** model requires specifying s_b, g_b, r_b, $\vec{q_b}$, $\forall b \in \mathcal{B}$. For a **localized** model, the same random variables could be used as the state representation (but with $\mathbf{g}^{\mathcal{A}_m}$ defined over a different domain than $\mathbf{g}$). However, requiring

5

specification of the r_b's is unnecessary and too restrictive; we may not be able to well predict the rule a brick has used by examining the set of active bricks. This is especially true if the active set only contains one brick! Instead, we represent the state representation of the localized model by the random variables $\mathbf{X}_{-r}^{\mathcal{A}_m} = \{s_a, g_a, \vec{q}_a\}$, $\forall a \in \mathcal{A}_m$, and we marginalize over r_a wherever it is used. In particular

$$\sum_{\mathbf{r}} P^{\mathcal{A}_m}(\mathbf{X}^{\mathcal{A}_m}, I) = P^{\mathcal{A}_m}(\mathbf{X}_{-r}^{\mathcal{A}_m}, I) \tag{15}$$

$$= \left(\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) \right) \tag{16}$$

$$\times \left(\prod_{a \in \mathcal{A}_m} \sum_{r_a} \left(P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \right) \right)$$

$$\times \prod_{a \in \mathcal{A}_m} P(\vec{q}_a \mid s_a)$$

$$\times \left(P^{\mathcal{A}_m}(I \mid \mathbf{q}^{\mathcal{A}_m}) \right).$$

Inference

Inference proceeds by iteratively adding one brick at a time to the active set. At any iteration, m, our goal is to maintain a representation of the localized posterior distribution $P^{\mathcal{A}_m}(\mathbf{X}_{-r}^{\mathcal{A}_m} \mid I)$. To maintain this representation, we employ a particle filtering approach. We treat samples from the previous localized posterior distribution $P^{\mathcal{A}_{m-1}}(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid I)$ as draws from a proposal distribution, and re-weight these particles accordingly (discuss re-weighting and evaluating marginals later). Note that $P^{\mathcal{A}_{m-1}}(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid I) \neq P^{\mathcal{A}_m}(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid I)$; the model defined by $P^{\mathcal{A}_m}$ is different than the one defined by $P^{\mathcal{A}_{m-1}}$.

Suppose we have sampled a previous state $\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}$, and suppose the brick we are adding to the active set is denoted b. To generate $\mathbf{X}_{-r}^{\mathcal{A}_m}$, we need to draw values for the set of random variables defined by $\mathbf{X}_{-r}^{\mathcal{A}_m} \setminus \mathbf{X}_{-r}^{\mathcal{A}_{m-1}} = \{s_b, \vec{q}_b, \mathbf{g}_{\mathbf{b}}^{\mathcal{A}_{\mathbf{m}}}, comp(g^{\mathcal{A}_{m-1}})\}$. That is, we need to "expand" the representation of $\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}$ to include brick b. As such, we need to draw a sample from $P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_{\mathbf{b}}^{\mathcal{A}_{\mathbf{m}}}, comp(g^{\mathcal{A}_{m-1}}) \mid \mathbf{X}_{-r}^{\mathcal{A}_{m-1}}, I)$. We note that for $g \in comp(\mathbf{g}^{\mathcal{A}_m})$, the only possible values that can be sampled for g are $g = \emptyset$, $g = b$. That is, each not-yet-assigned brick slot choose

either to point to the newly-added active brick b, or remain empty. Note that

$$P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_\mathbf{b}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}}) \mid \mathbf{X}_{-r}^{\mathcal{A}_{m-1}}, I) = \frac{P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_\mathbf{b}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}}), \mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid I)}{\sum_{s_b} \sum_{g_b} \sum_{\mathbf{g}_\mathbf{b}^{\mathcal{A}_\mathbf{m}}} \sum_{comp(g^{\mathcal{A}_{m-1}})} P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_\mathbf{b}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}}), \mathbf{X}_{-r}^{\mathcal{A}_{m-1}})}$$

$$P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_\mathbf{b}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}}), \mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid I) = P^{\mathcal{A}_m}(\mathbf{X}_{-r}^{\mathcal{A}_m} \mid I)$$

$$\propto P^{\mathcal{A}_m}(\mathbf{X}_{-r}^{\mathcal{A}_m}, I).$$

We first note that for any setting of the random variables outlined above, we may classify it as one of three events: 1) brick b is off, 2) brick b is on, and has no parents, and 3) brick b is on, and has at least one parent. We compute the mass associated with each of these three events, with approximations where needed. We will be able to evaluate the denominator in Eqn. 17 by summing the probability mass associated with these three events. We examine these events in more detail below.

Event 1: brick b is off

For this event, there is only one setting of the random variables $\{s_b, \vec{q}_b, \vec{g}_b^{\mathcal{A}_m}, comp(g^{\mathcal{A}_{m-1}}\}$ that has non-zero probability: $s_b = 0$, $\vec{q}_b = \perp$, $g_{b,k} = \perp \ \forall k$, $g = \emptyset \ \forall g \in comp(\mathbf{g}^{\mathcal{A}_{m-1}})$. Note that $g^{\mathcal{A}_m} \neq g^{\mathcal{A}_{m-1}}$ only because $g^{\mathcal{A}_{m-1}}$ specifies the (lack of) children for the brick b, which is not specified in $g^{\mathcal{A}_{m-1}}$. The other corresponding entries of $g^{\mathcal{A}_m}$ and $g^{\mathcal{A}_{m-1}}$ are indeed equal, however. It is trivial to evaluate Eqn. 16 given a particular sampled previous state, $\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}$.

Event 1: brick b is on and has no parents

There are many settings of the random variables $\{s_b, \vec{q}_b, \mathbf{g}_\mathbf{b}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}})\}$ that fall under this event. They can be characterized as: $s_b = 1$, $\vec{q}_b \neq \perp$, $g = \emptyset \ \forall g \in comp(\mathbf{g}^{\mathcal{A}_{m-1}})$, and $g_{b,k}^{\mathcal{A}_m} = \{\emptyset, a \text{ s.t. } a \in \mathcal{A}_m\}$, $\forall k$. So, the probability mass associated with this event is given by

$$\sum_{\vec{q}_b \in P(b)} \sum_{\vec{g}_b^{\mathcal{A}_m}} P^{\mathcal{A}_m}(s_b = 1, \vec{q}_b, \vec{g}_b^{\mathcal{A}_m} comp(\mathbf{g}^{\mathcal{A}_{m-1}}) = \emptyset, \mathbf{X}_{-r}^{\mathcal{A}_{m-1}}, I) \tag{20}$$

where we have used $comp(\mathbf{g}^{\mathcal{A}_{m-1}}) = \emptyset$ as a shorthand for $g = \emptyset \ \forall g \in comp(\mathbf{g}^{\mathcal{A}_{m-1}})$. Expanding, we get

$$= \sum_{\vec{g}_b^{\mathcal{A}_m}} \Big(\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) \Big) \Big(\prod_{a \in \mathcal{A}_m} \sum_{r_a} \big(P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \big) \Big) \qquad (21)$$

$$\Big(\prod_{a \in \mathcal{A}_{m-1}} P(\vec{q}_a \mid s_a) \Big) \Big(\sum_{\vec{q}_b \in P(b)} P(\vec{q}_b \mid s_b = 1) \Big) P^{\mathcal{A}_m}(I \mid \mathbf{q}^{\mathcal{A}_{m-1}}, \vec{q}_b).$$

The term $\big(\prod_{a \in \mathcal{A}_{m-1}} P(\vec{q}_a \mid s_a) \big) \big(\sum_{\vec{q}_b \in P(b)} P(\vec{q}_b \mid s_b = 1) \big) P^{\mathcal{A}_m}(I \mid \mathbf{q}^{\mathcal{A}_{m-1}}, \vec{q}_b)$ can be evaluated by enumerating all possible values for $\vec{q}_b \in P(b)$. In practice, whether this is practical or not is dependent on $\#(P(b))$. If $\#(P(b))$ is "small enough", then $\big(\prod_{a \in \mathcal{A}_{m-1}} P(\vec{q}_a \mid s_a) \big) \big(\sum_{\vec{q}_b \neq \perp} P(\vec{q}_b \mid s_b = 1) \big) P^{\mathcal{A}_m}(I \mid \mathbf{q}^{\mathcal{A}_{m-1}}, \vec{q}_b)$ can be computed. However, in general, it is not possible to efficiently compute the term $\big(\prod_{a \in \mathcal{A}_{m-1}} P(\vec{q}_a \mid s_a) \big) \big(\sum_{\vec{q}_b \neq \perp} P(\vec{q}_b \mid s_b = 1) \big) P^{\mathcal{A}_m}(I \mid \mathbf{q}^{\mathcal{A}_{m-1}}, \vec{q}_b)$ (explain why- each slot can have $\#(Am)$ possibilities, and needs to be enumerated for each of the possible rules b can take. So, this term is exponential in $\#(Am)$. This term can be decomposed into events stating how many unique children are pointed to, but this still requires lots of computation.), and so we come up with approximations for it.

Noting the definition given in Eqn. 2, we see that $\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) = \epsilon^{\#(oprh(\mathcal{A}_m))}$.

For coming child approximations, do both slot-centric view (Pedro's approximation) and child-centric view (Jeroen's approximation)