From *Naive Physics* to *Connotation*: Modeling Commonsense in Frame Semantics

Yejin Choi

Paul G. Allen School of Computer Science & Engineering

UNIVERSITY *of* WASHINGTON

Intelligent Communication

➔ reading between the lines

Intelligent Communication

➔ Reading between the lines

Understanding

what is said

+

what is *not* said

"CHEESEBURGER STABBING"

- Someone stabbed a cheeseburger?
- A cheeseburger stabbed someone?
- A cheeseburger stabbed another cheeseburger?
- Someone stabbed someone else with a cheeseburger?
- Someone stabbed someone else over a cheeseburger?

Intelligent Communication

➔ Reading between the lines

Understanding

what is said

\+

what is *not* said

Knowledge about the world to fill in the gap!

"CHEESEBURGER STABBING"

- Someone stabbed a cheeseburger?
- A cheeseburger stabbed someone?
- A cheeseburger stabbed another cheeseburger?
- Someone stabbed someone else with a cheeseburger?
- Someone stabbed someone else over a cheeseburger?

Intelligent Communication

Blueberry Muffins

Ingredients
1 cup milk
1 egg
1/3 cup vegetable oil
2 cups all-purpose flour
2 teaspoons baking powder
1/2 cup white sugar
1/2 cup fresh blueberries

Procedure
1. Preheat ov… uffin tin with paper liners.
2. In a large bowl, … ether milk, egg, and oil. Add flour, baking powder, sugar, and blueberries; gently … ix the batter with only a few strokes. Spoon batter into cups.
3. Bake for 30 minutes. Serve hot.

http://allrecipes.com/Recipe/Blueberry-Muffins-I/

Knowledge about the world to fill in the gap!

Bake what? where?

Types of Knowledge

Information Extraction

Encyclopedic knowledge

- Who is the president of which country and born in what year…

Naïve Physics

Social Norms

Procedural (How-to) Knowledge

Commonsense knowledge

- It's not possible to stab someone using a cheeseburger
- Stabbing a cheeseburger is not newsworthy…
- Stabbing someone is generally immoral
- How to make a cheeseburger

Plan

1. Commonsense Frame Semantics

    Naive physics

2. Modeling the World, not just Language

Verb Physics

relative physical knowledge about actions and objects

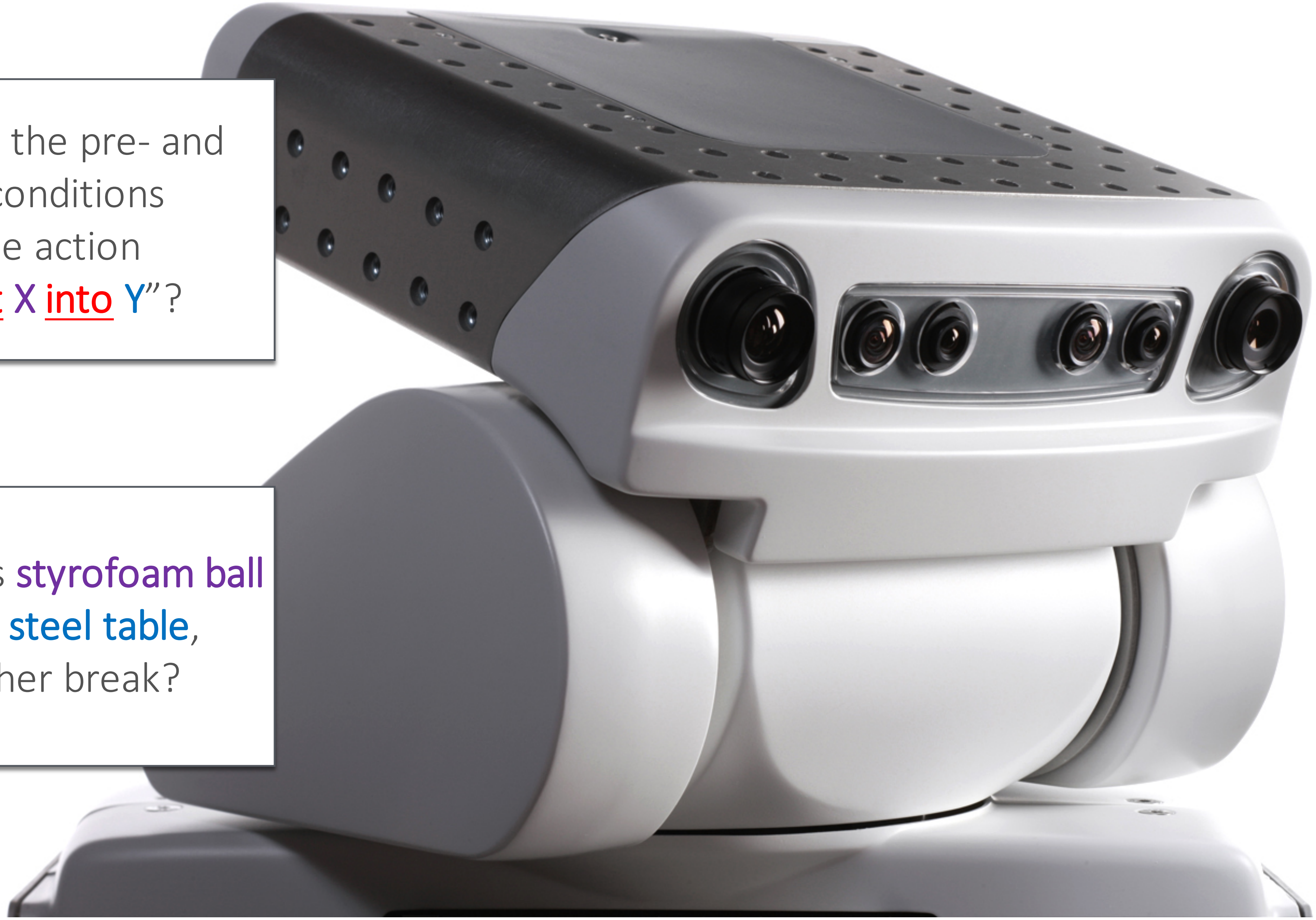
Maxwell Forbes et al. (ACL 2017)

Winograd Schema Challenge

- The trophy would not fit in the brown suitcase because it was too **big** (*small*).
  What was too **big** (*small*)?
  Answer 0: the trophy
  Answer 1: the suitcase

Physical Knowledge about the World

"Hey, Robot, Fetch me a container to put this pie."



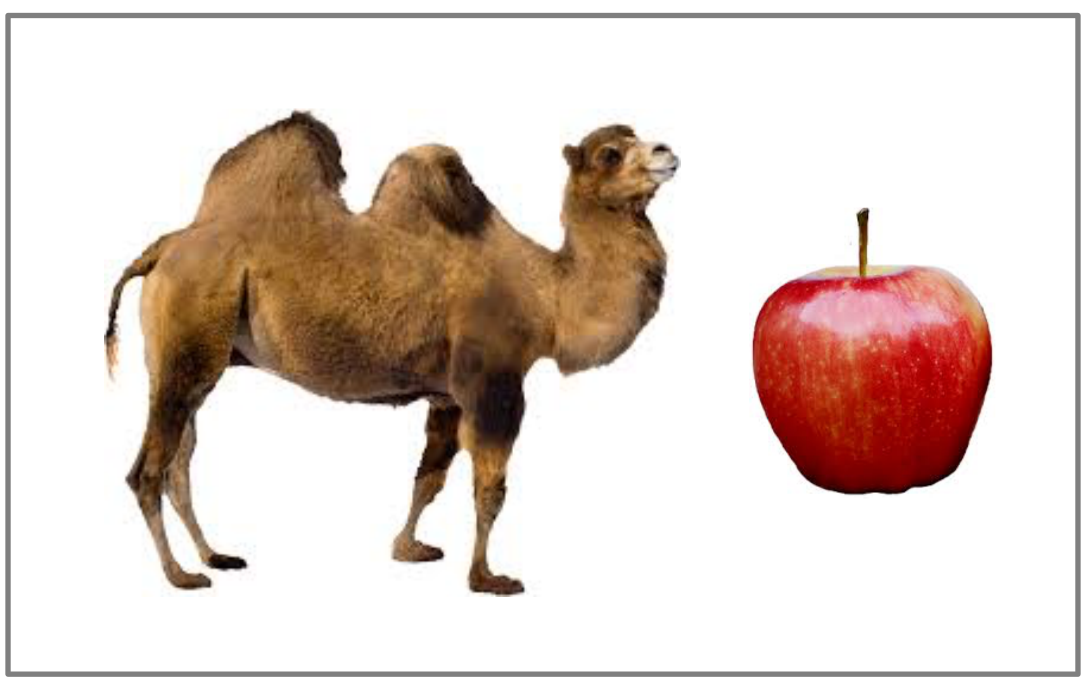
What are the pre- and
post-conditions
of the action
"to put X into Y"?

If I drop this styrofoam ball
into the steel table,
will either break?

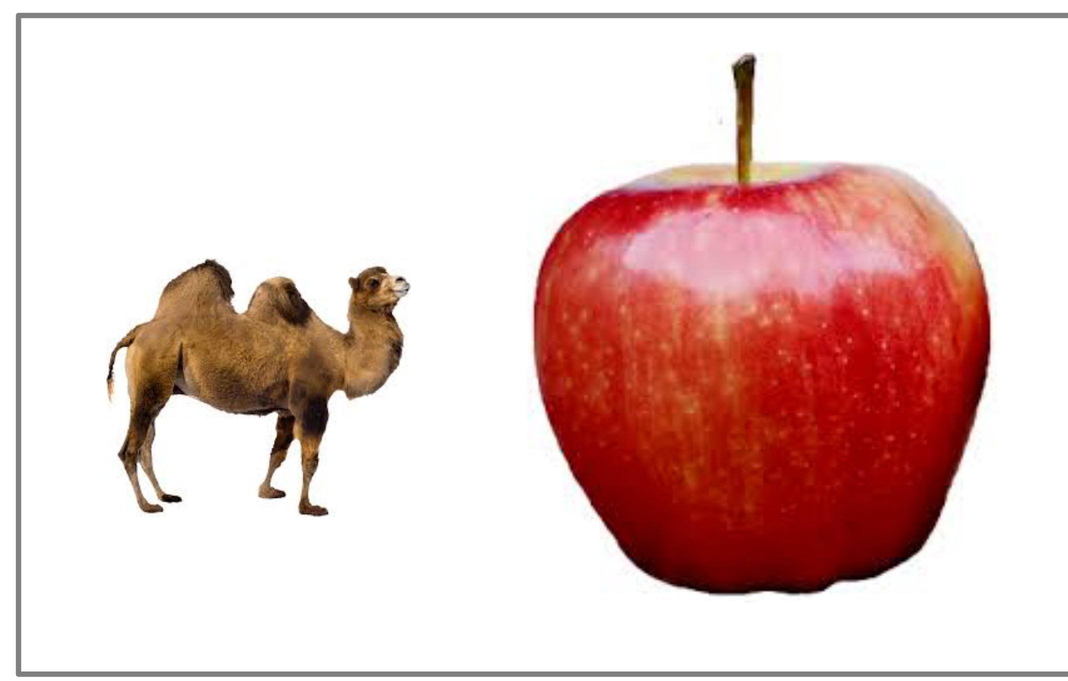
Support from Psychology Studies

- A familiar-size stroop effect. (Konkle, T., and Oliva, A. 2012.)
- Knowledge on size represented in a perceptual or analog format (Moyer, 1973; Paivio, 1975; Rubinsten & Henik, 2002).
- Objects have a canonical visual size (Konkle & Oliva, 2011; Linsen, Leyssen, Sammartino, & Palmer, 2011).

congruent



incongruent



Overcoming Reporting Bias

- People don't state the obvious
  - (Van Durme 2010, Sorower et al., 2011)

“I am larger than a chair”

~~*"I am larger than a chair"*~~

~~*“I am larger than a pen”*~~

~~*“I am larger than a stone”*~~

~~*“I am larger than a chair”*~~

~~*“I am larger than a ball”*~~

~~*“I am larger than a towel”*~~

"The horse was as small as a dog!"

$\Longrightarrow$ horse $=^{\text{size}}$ dog ?

Are Elephants Bigger than Butterflies?

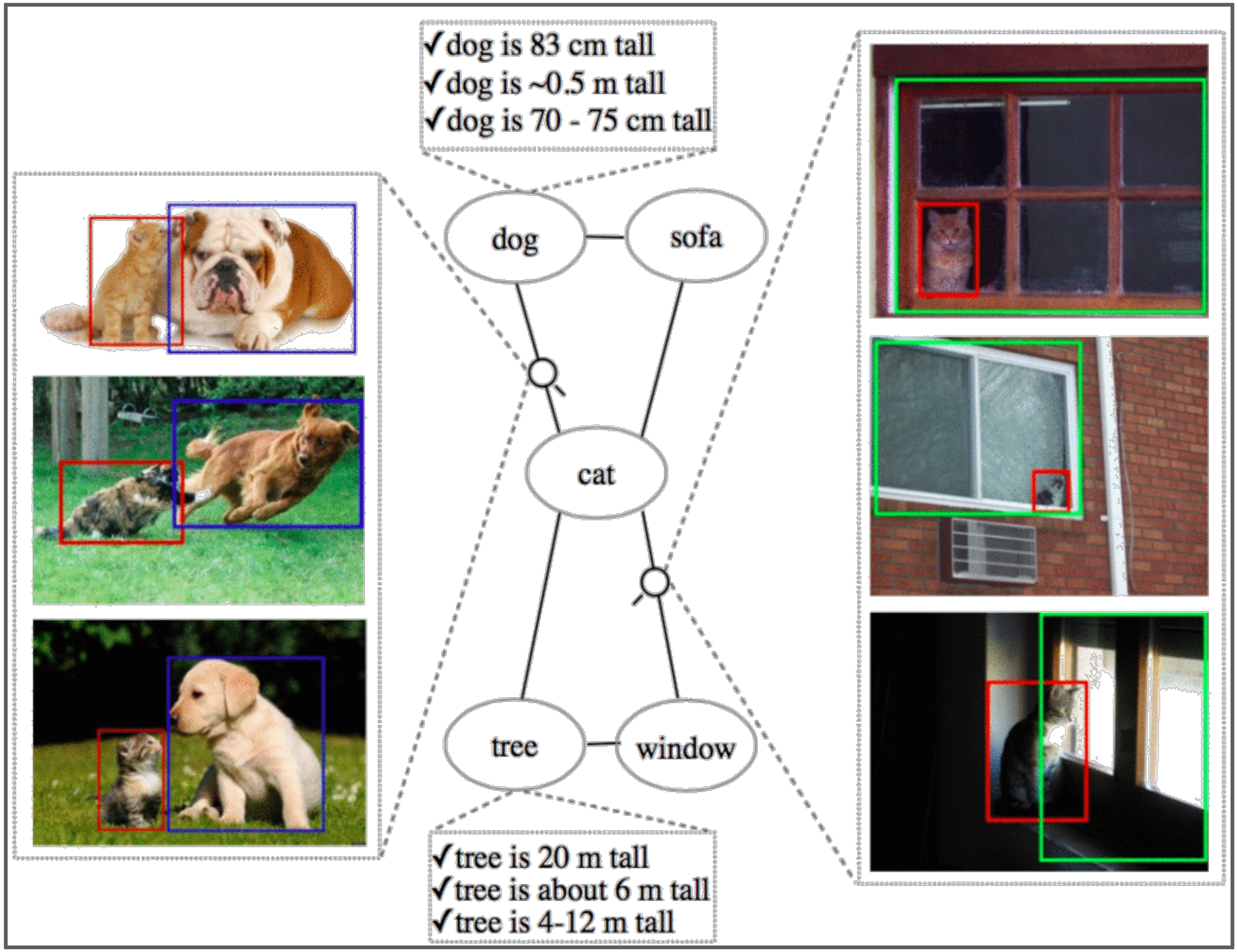
(Bagherinezhad et al. @ AAAI 2016)

- Language – absolute estimation
  - "car is * x * m"
  - "person is * m tall"

- Vision – relative estimation

$$\frac{size(O_i)}{size(O_j)} =$$

$$\frac{area(box_1)}{area(box_2)} \times \frac{depth(box_1)^2}{depth(box_2)^2}$$



Overcoming Reporting Bias

- People don't state the obvious
  - Elephants are bigger than butterflies
- Thoughts last year
  - Look beyond language, such as vision!
  - (AAAI 2016, ACL 2016, ICCV 2015)
  - Can't learn diverse physical knowledge (weight, strength, rigidness, strength…)
- Thoughts this year
  - Back to language, but with a different game plan!

Key Insight

~~“I am larger than a pen”~~

~~“I am larger than a stone”~~

~~“I am larger than a chair”~~

~~“I am larger than a ball”~~

~~“I am larger than a towel”~~

"I threw the ______"

pen
stone
chair
ball
towel

x threw y

x is bigger than y
x weighs more than y
as a result, y will be moving faster than x

Let’s solve two related problems

Physical properties implied by predicates

“I **picked up** the <unk>.”

“I **dropped** a cherry **into** the <unk>.”

“The <unk> **shattered** when it hit the ground

Physical properties of objects

size

weight

strength

Along Five

$x >^{size} y$

$x >^{weight} y$

$x <^{rigidness} y$

$x >^{strength} y$

$x <^{speed} y$

Learning Knowledge from Language

Using IE Patterns
(*Gordon et al., 2010, Gordon and Schubert, 2012*, Narisawa et al. 2013, Tandon et al. 2014*)*

Narrative Schemas, Script knowledge
(*Chambers and Jurafsky. 2009, Pichotta and Mooney 2014, 2016)*

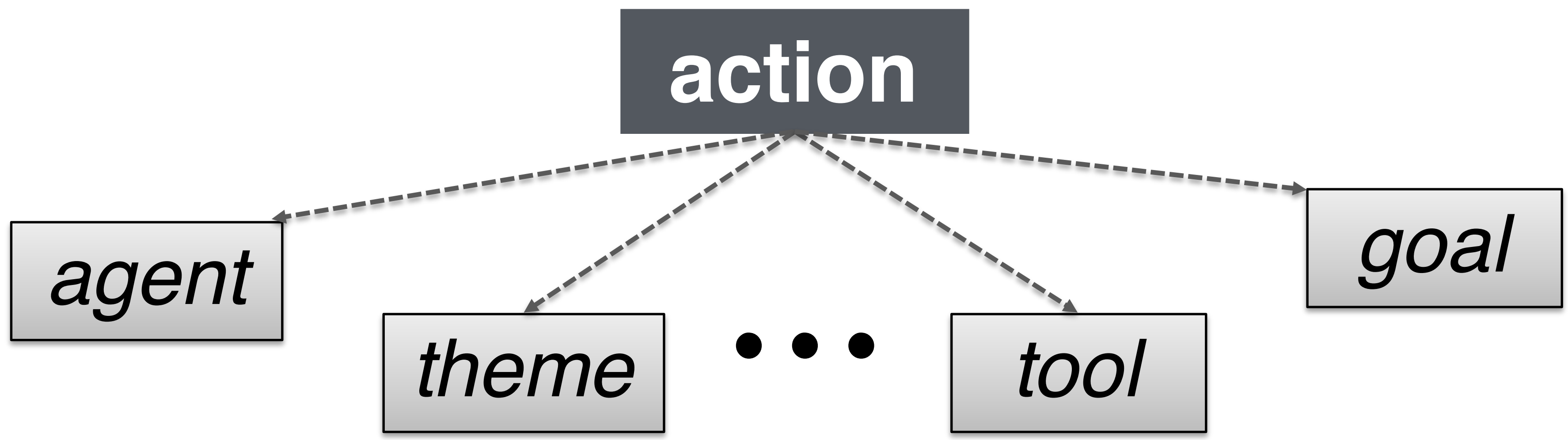
Inferring Knowledge through Reasoning

Inverting Grice's maxims
Mohammad S. Sorrower, et al., 2011

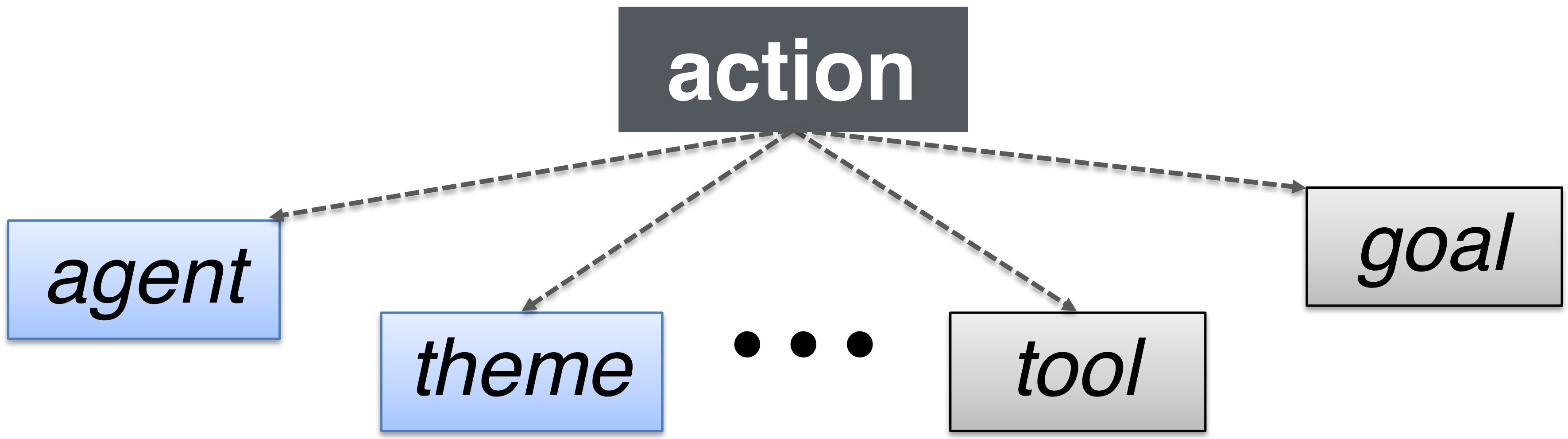
Natural logic, natural reasoning, and entailment
MacCartney and Manning, 2007, Angeli and Manning 2013, 2014, Bowman et al., 2015

Representation: **VerbPhysics**

Representation: VerbPhysics Frames



Representation: VerbPhysics Frames



$\mathcal{R}$(agent, theme)

"He threw the ball"

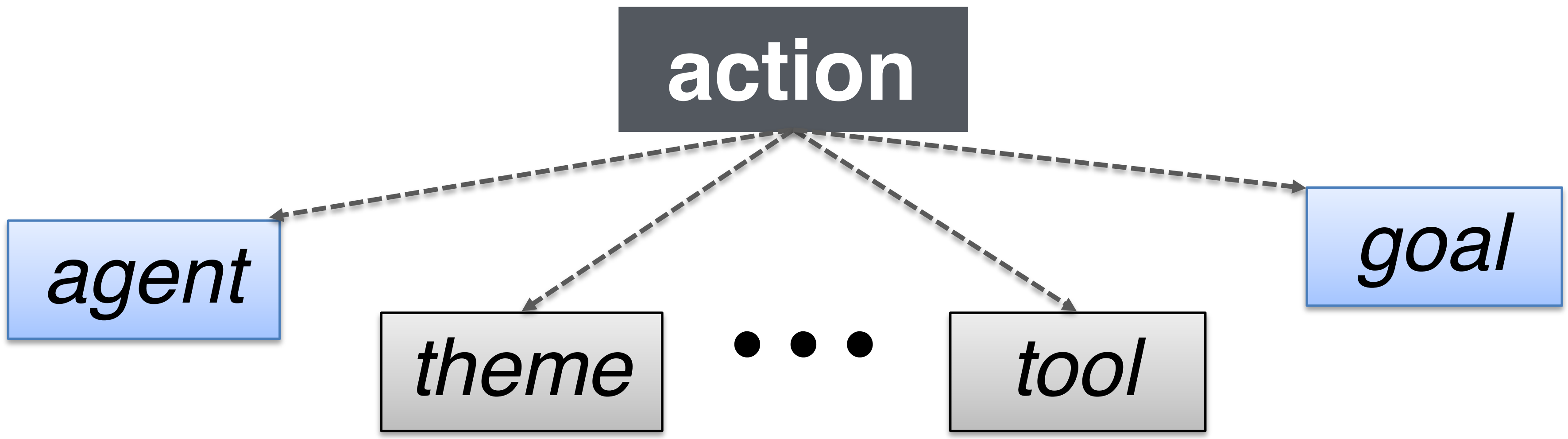
x threw y

⇒ x is **larger** than y

⇒ x is **heavier** than y

⇒ x is **slower** than y

Representation: VerbPhysics Frames



$\mathcal{R}$(agent, theme)

"He threw the ball"

x threw *y*

$\Rightarrow$ *x* is **larger** than *y*
$\Rightarrow$ *x* is **heavier** than *y*
$\Rightarrow$ *x* is **slower** than *y*

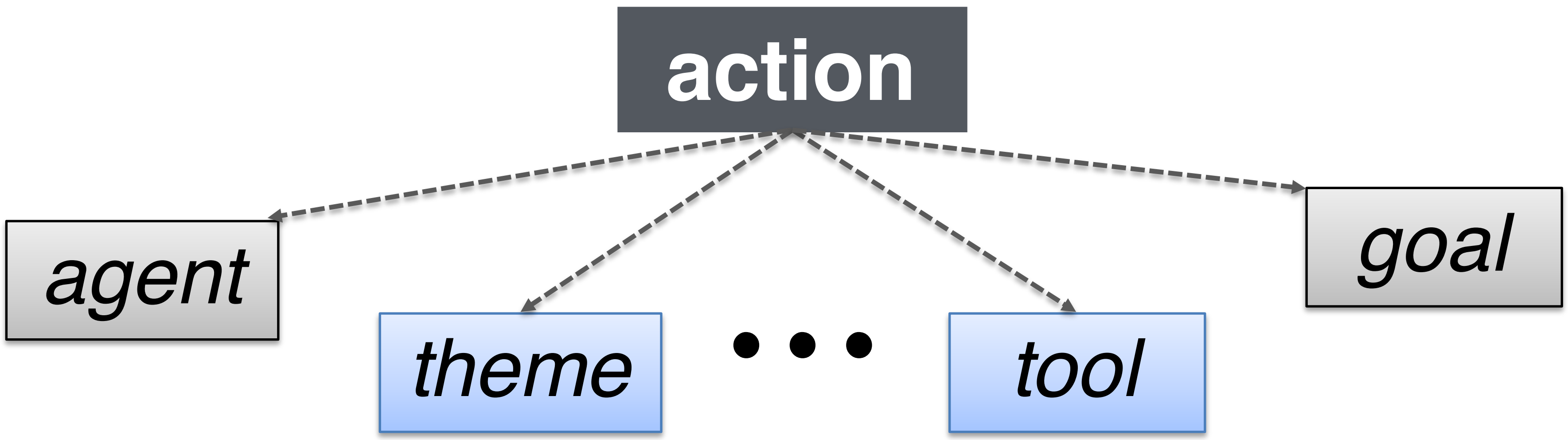
$\mathcal{R}$(agent, goal)

"We walked into the house"

x walked into *y*

$\Rightarrow$ *x* is **smaller** than *y*
$\Rightarrow$ *x* is **lighter** than *y*
$\Rightarrow$ *x* is **faster** than *y*

Representation: VerbPhysics Frames



$\mathcal{R}$(agent, theme)

"He threw the ball"

x threw y

⇒ x is **larger** than y
⇒ x is **heavier** than y
⇒ x is **slower** than y

$\mathcal{R}$(agent, goal)

"We walked int

x walked into

⇒ x is **smaller**
⇒ x is **lighter**
⇒ x is **faster** t

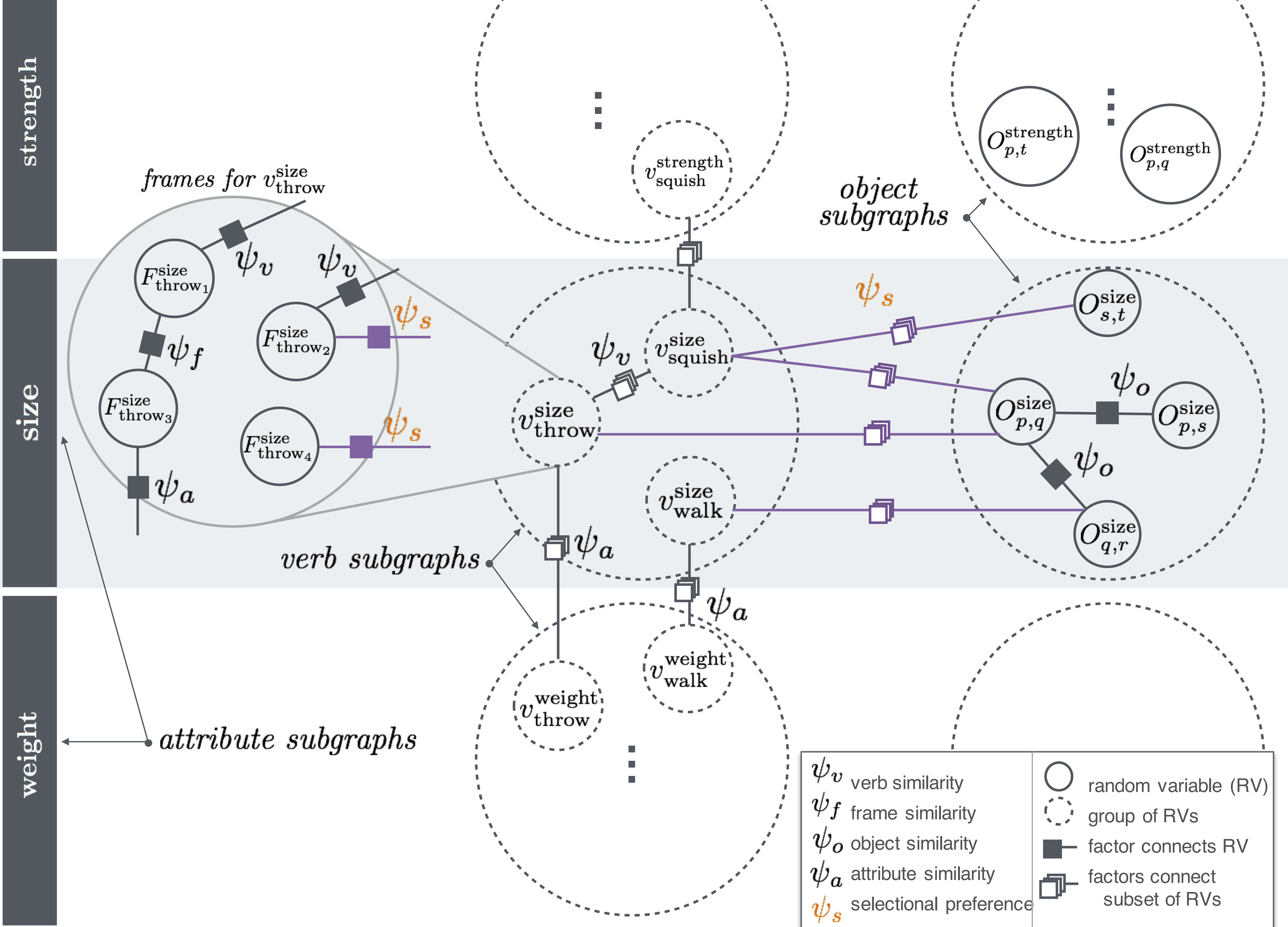
$\mathcal{R}$(theme, goal)

"I squashed the bug with my boot"

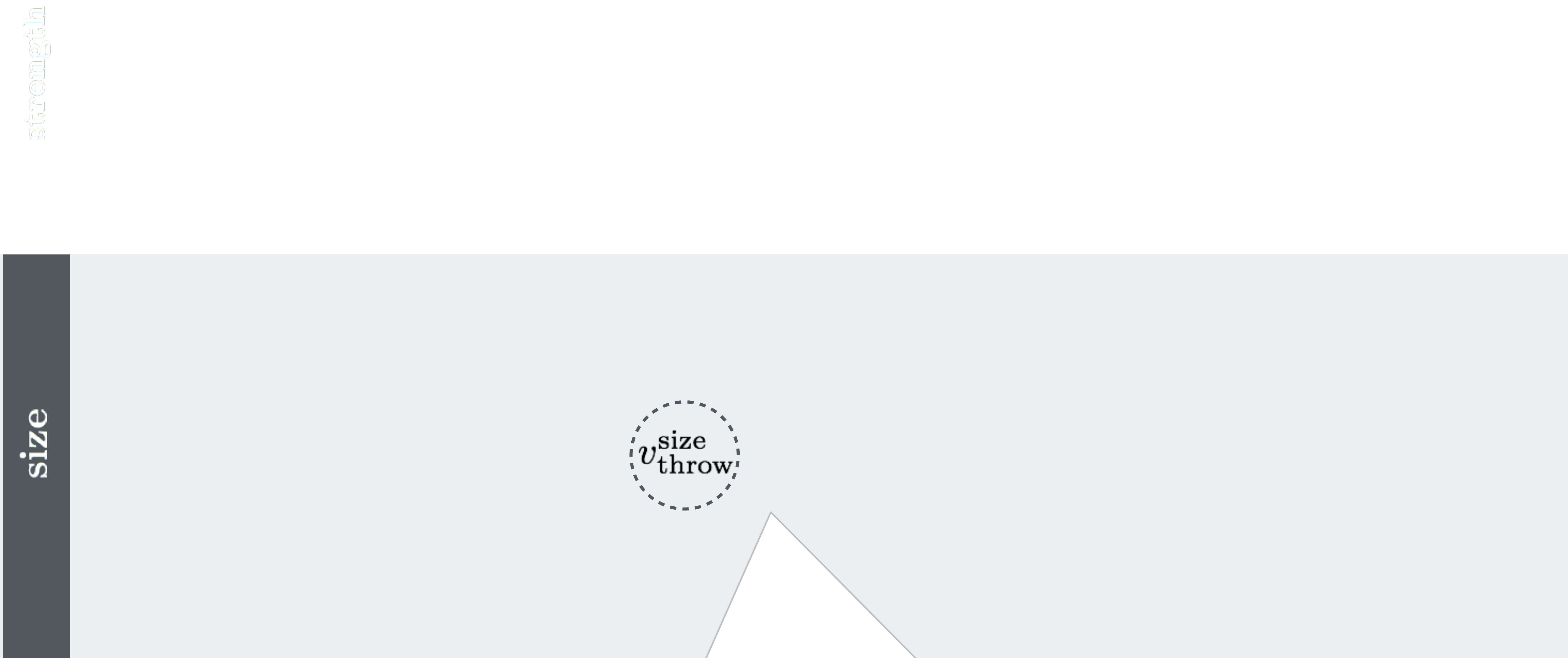
squashed x with y

⇒ x is **smaller** than y
⇒ x is **lighter** than y
⇒ x is **slower** than y
⇒ x is **less rigid** than y
⇒ x is **weaker** than y

Model



strength
size
weight
frames for $v^{\text{size}}_{\text{throw}}$
$F^{\text{size}}_{\text{throw}_1}$
$F^{\text{size}}_{\text{throw}_2}$
$F^{\text{size}}_{\text{throw}_3}$
$F^{\text{size}}_{\text{throw}_4}$
ψ_v
ψ_f
ψ_a
ψ_s
$v^{\text{strength}}_{\text{squish}}$
$v^{\text{size}}_{\text{squish}}$
$v^{\text{size}}_{\text{throw}}$
$v^{\text{size}}_{\text{walk}}$
$v^{\text{weight}}_{\text{throw}}$
$v^{\text{weight}}_{\text{walk}}$
object subgraphs
$O^{\text{strength}}_{p,t}$
$O^{\text{strength}}_{p,q}$
$O^{\text{size}}_{s,t}$
$O^{\text{size}}_{p,q}$
$O^{\text{size}}_{p,s}$
$O^{\text{size}}_{q,r}$
ψ_o
verb subgraphs
attribute subgraphs
ψ_v verb similarity
ψ_f frame similarity
ψ_o object similarity
ψ_a attribute similarity
ψ_s selectional preference
random variable (RV)
group of RVs
factor connects RV
factors connect subset of RVs



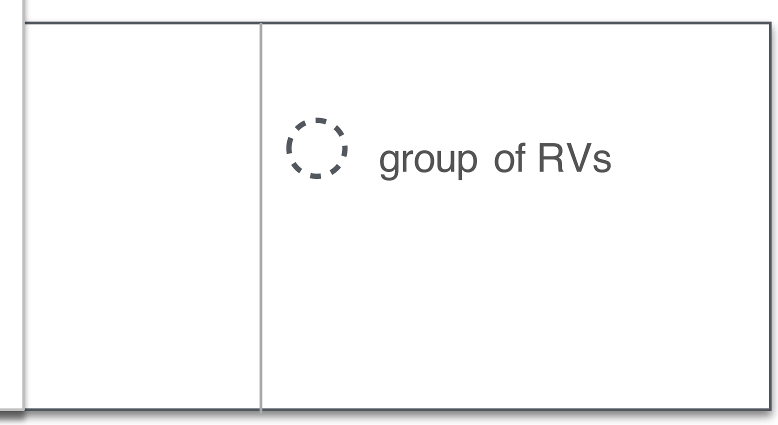
Each variable represents an entailment knowledge

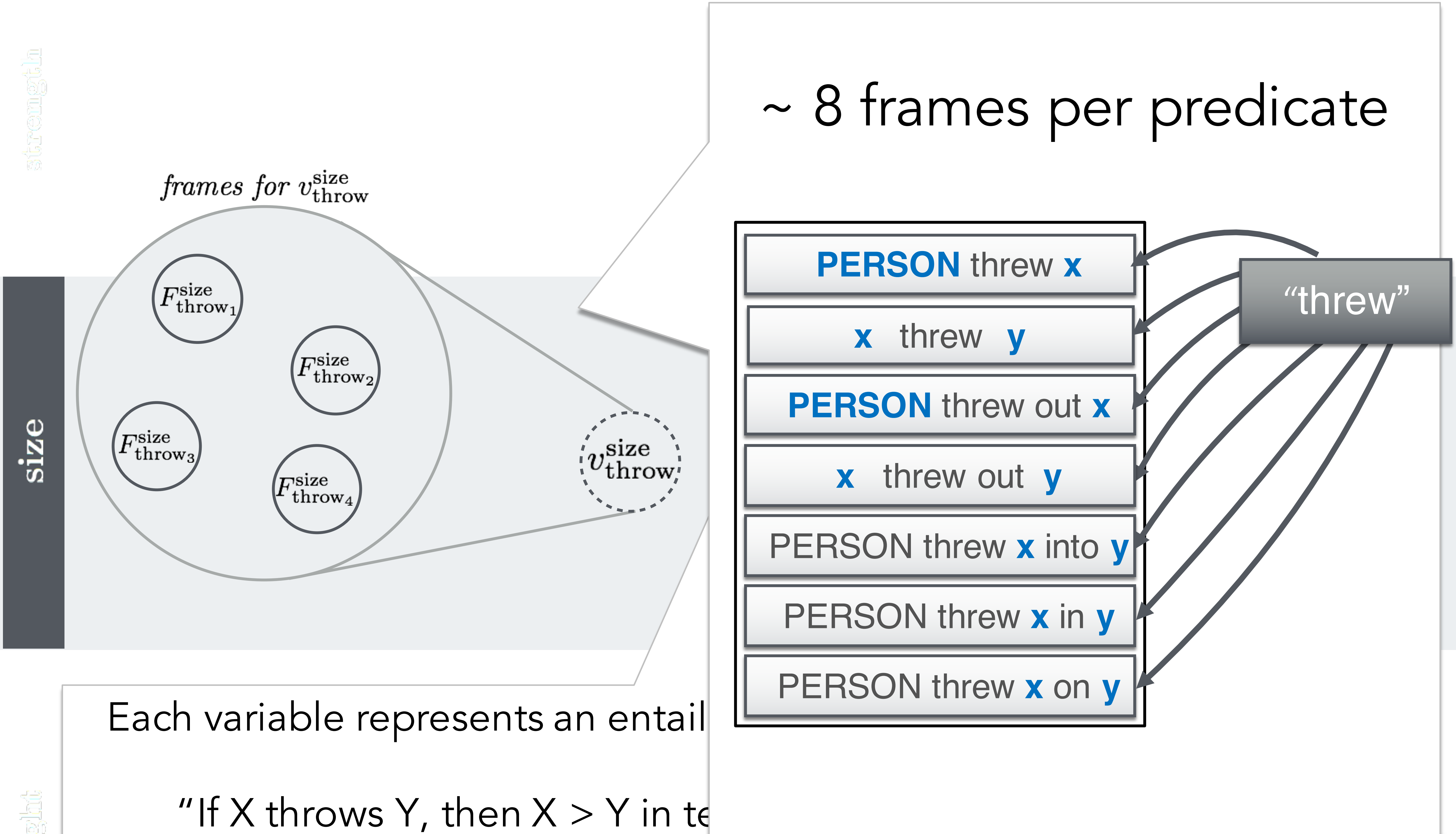
"If X throws Y, then X > Y in terms of size"

Or

"If X throws Y, then Z in size" where Z = {>, <, =}

(Using Levin action verbs (Levin, 1993))





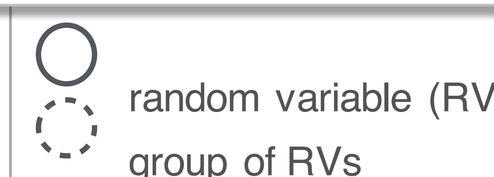
Each variable represents an entail

"If X throws Y, then X > Y in te

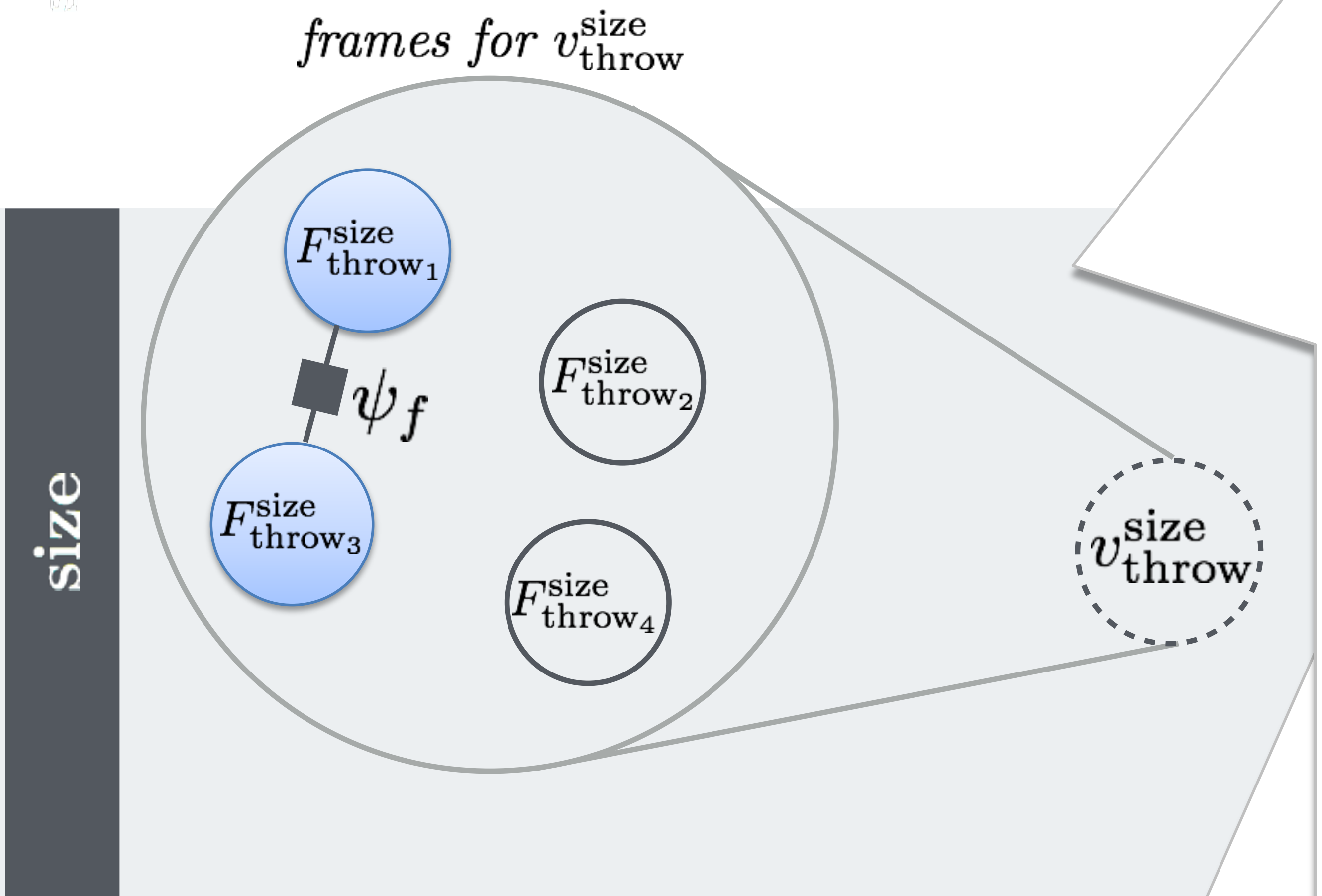
Or

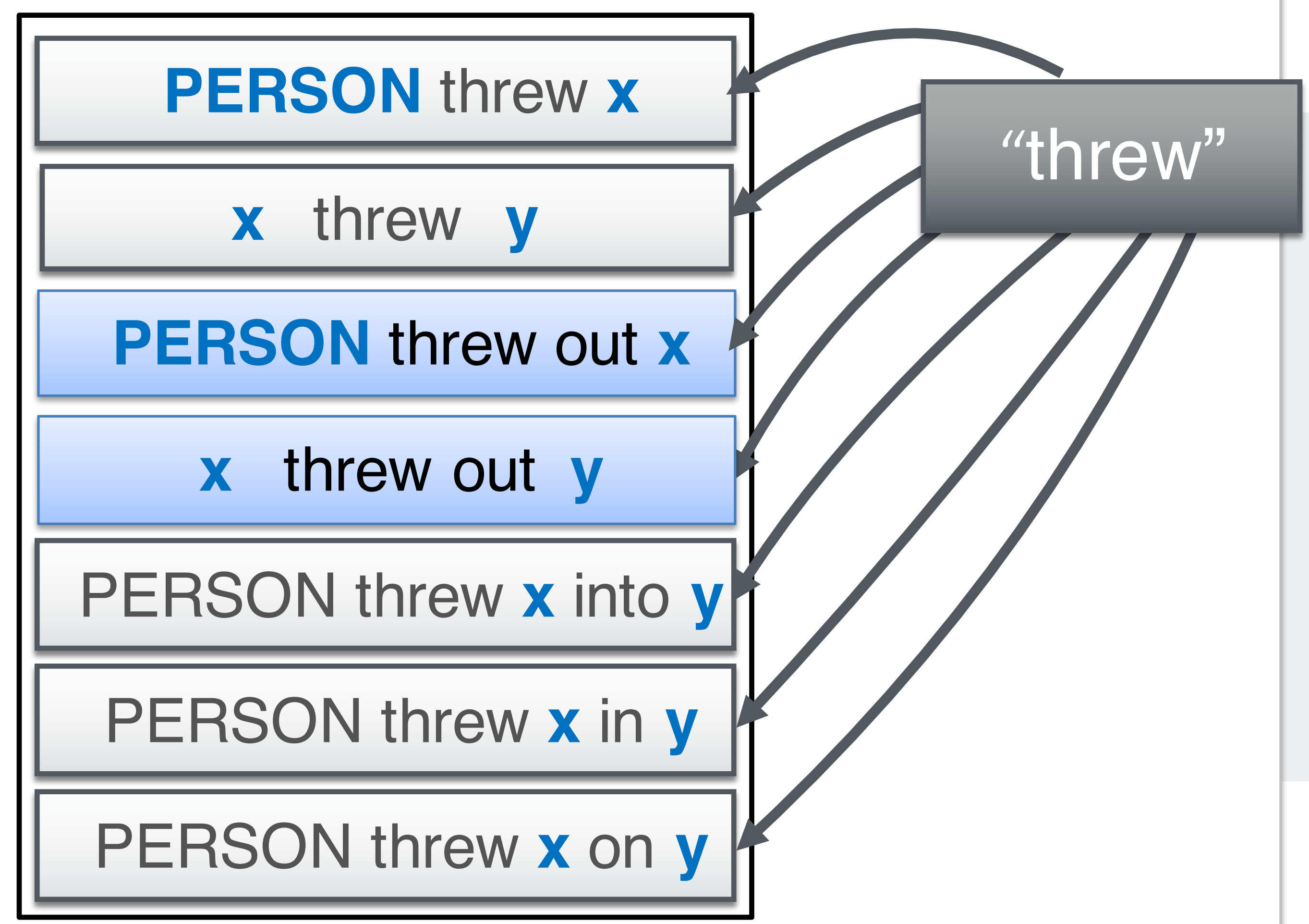
"If X throws Y, then Z in size" where Z = {>, <, =}

(Using Levin action verbs (Levin, 1993))



~ 8 frames per predicate





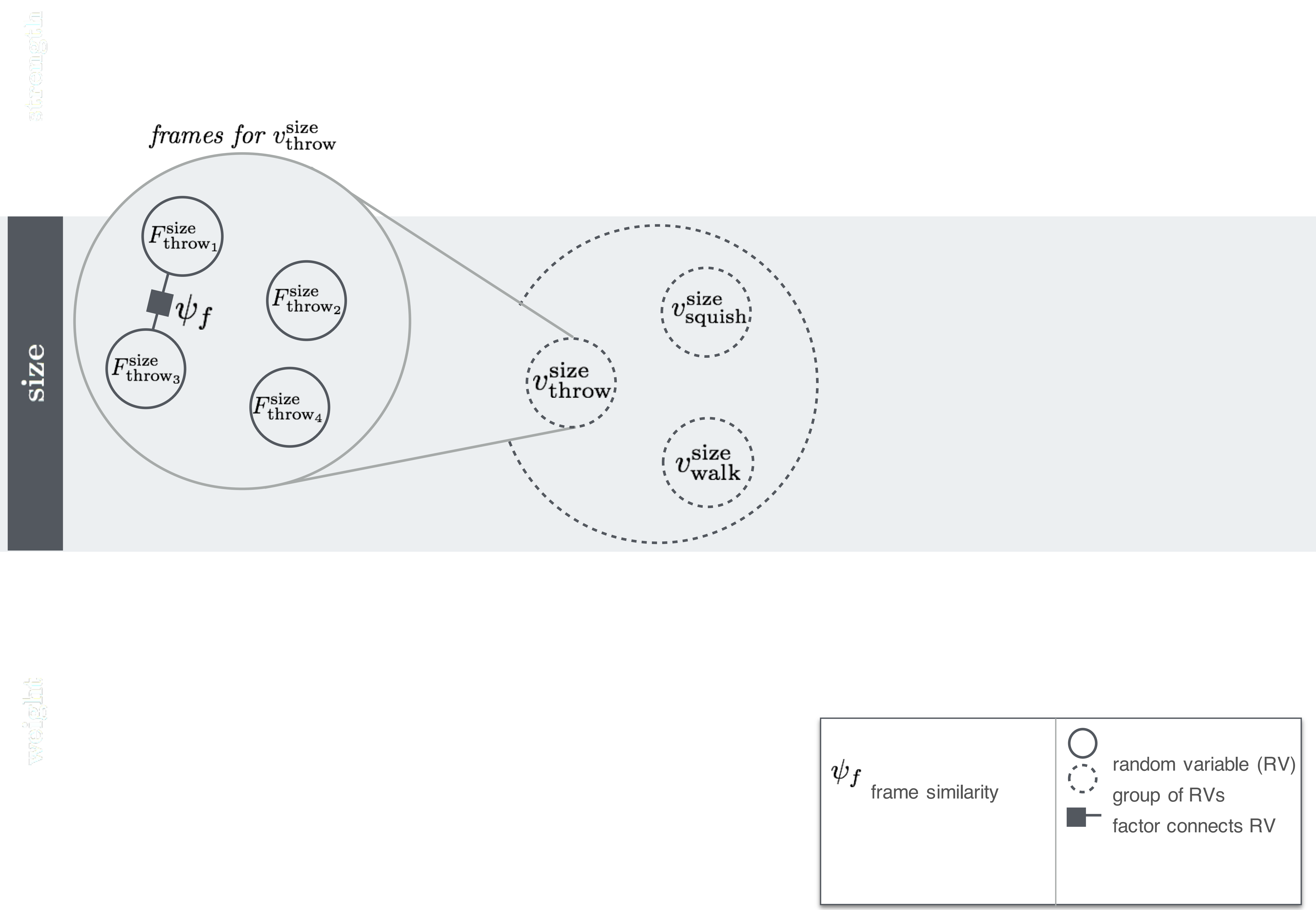
Each variable represents an entail

"If X throws Y, then X > Y in te
Or
"If X throws Y, then Z in size" where Z = {>, <, =}

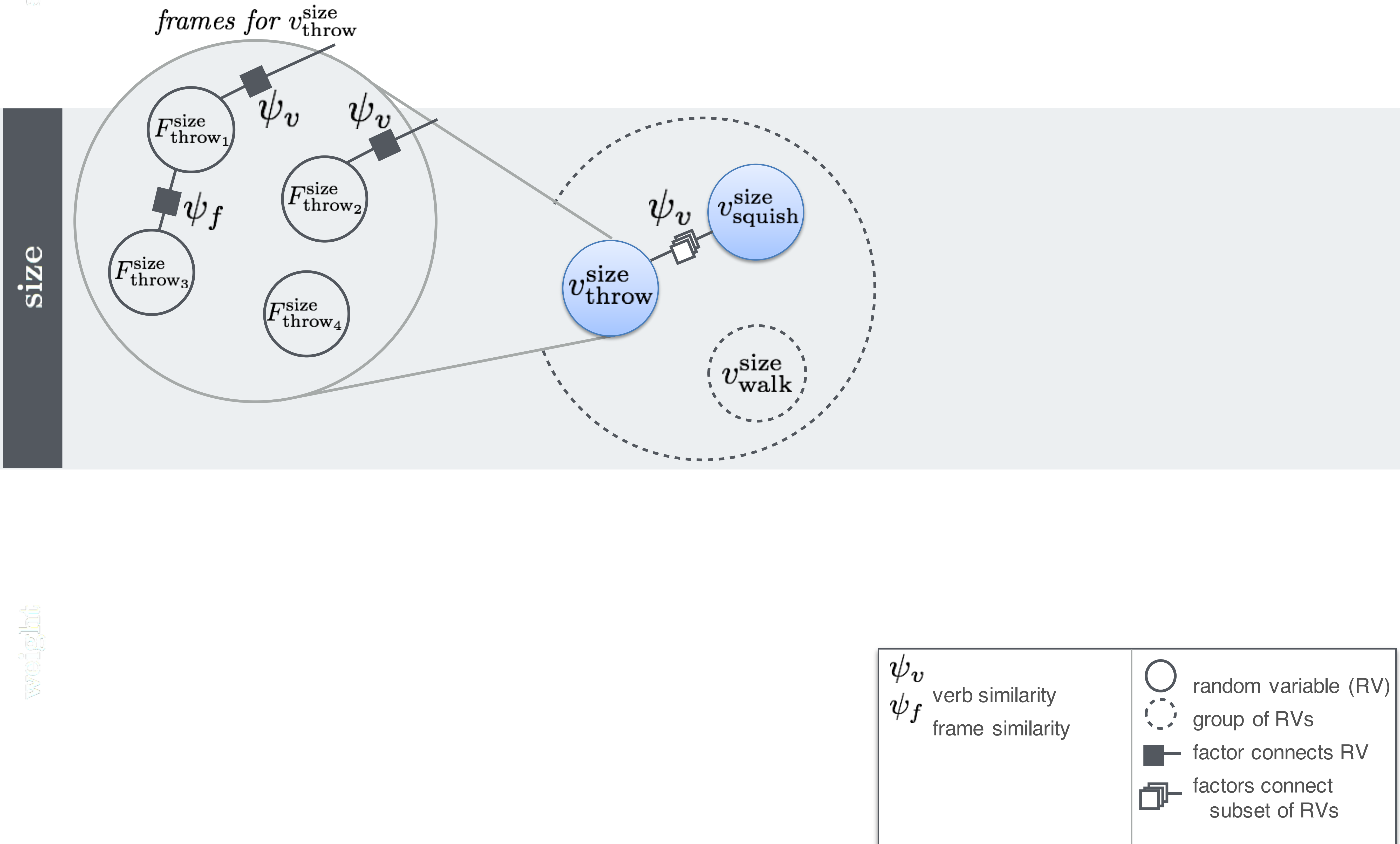
(Using Levin action verbs (Levin, 1993))

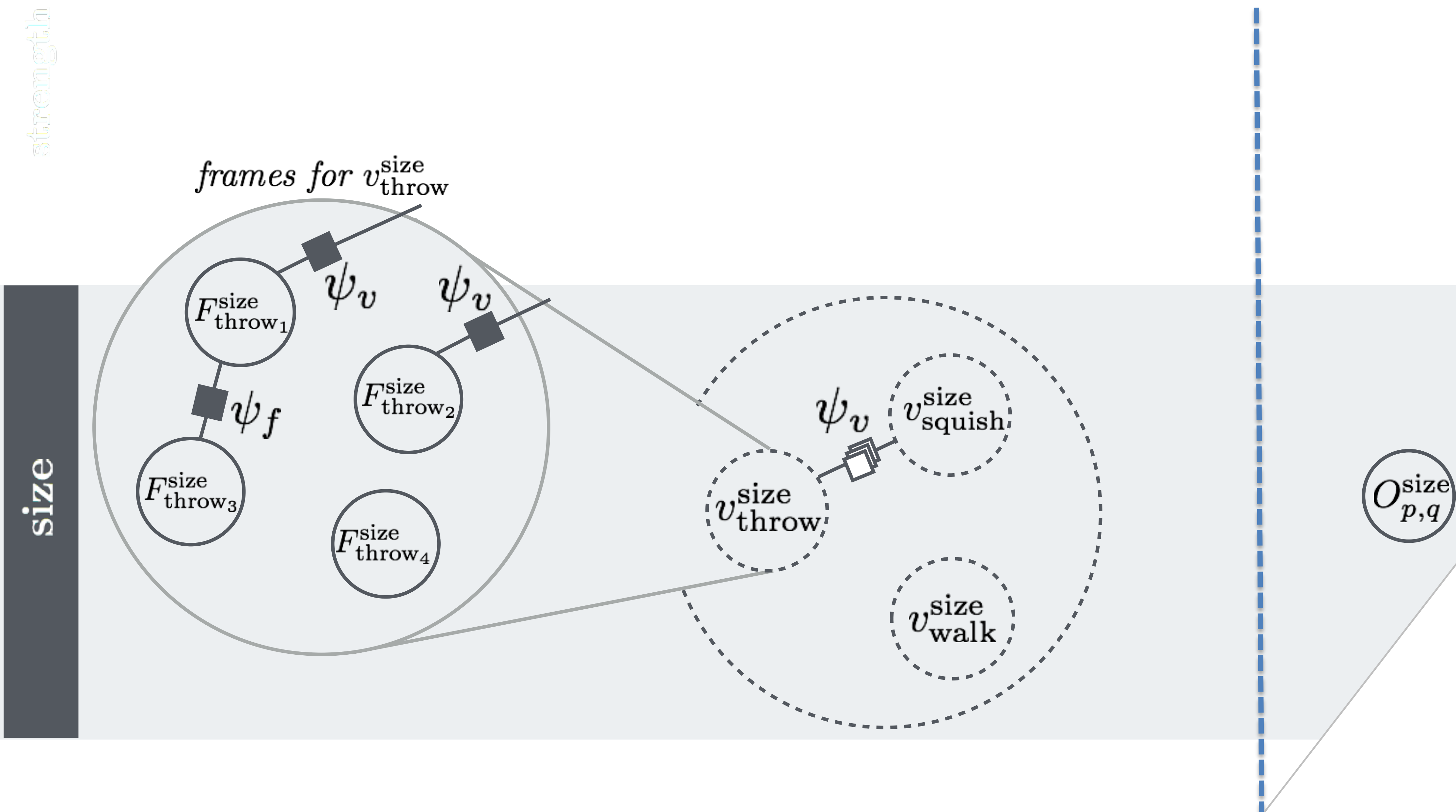
milarity

random variable (RV)
group of RVs
factor connects RV



strength
size
weight
frames for $v^{\text{size}}_{\text{throw}}$
$F^{\text{size}}_{\text{throw}_1}$
$F^{\text{size}}_{\text{throw}_2}$
$F^{\text{size}}_{\text{throw}_3}$
$F^{\text{size}}_{\text{throw}_4}$
ψ_f
$v^{\text{size}}_{\text{throw}}$
$v^{\text{size}}_{\text{squish}}$
$v^{\text{size}}_{\text{walk}}$
ψ_f frame similarity
random variable (RV)
group of RVs
factor connects RV


size
frames for $v^{\text{size}}_{\text{throw}}$
$F^{\text{size}}_{\text{throw}_1}$
$F^{\text{size}}_{\text{throw}_2}$
$F^{\text{size}}_{\text{throw}_3}$
$F^{\text{size}}_{\text{throw}_4}$
ψ_v
ψ_v
ψ_f
ψ_v
$v^{\text{size}}_{\text{squish}}$
$v^{\text{size}}_{\text{throw}}$
$v^{\text{size}}_{\text{walk}}$
ψ_v verb similarity
ψ_f frame similarity
random variable (RV)
group of RVs
factor connects RV
factors connect subset of RVs

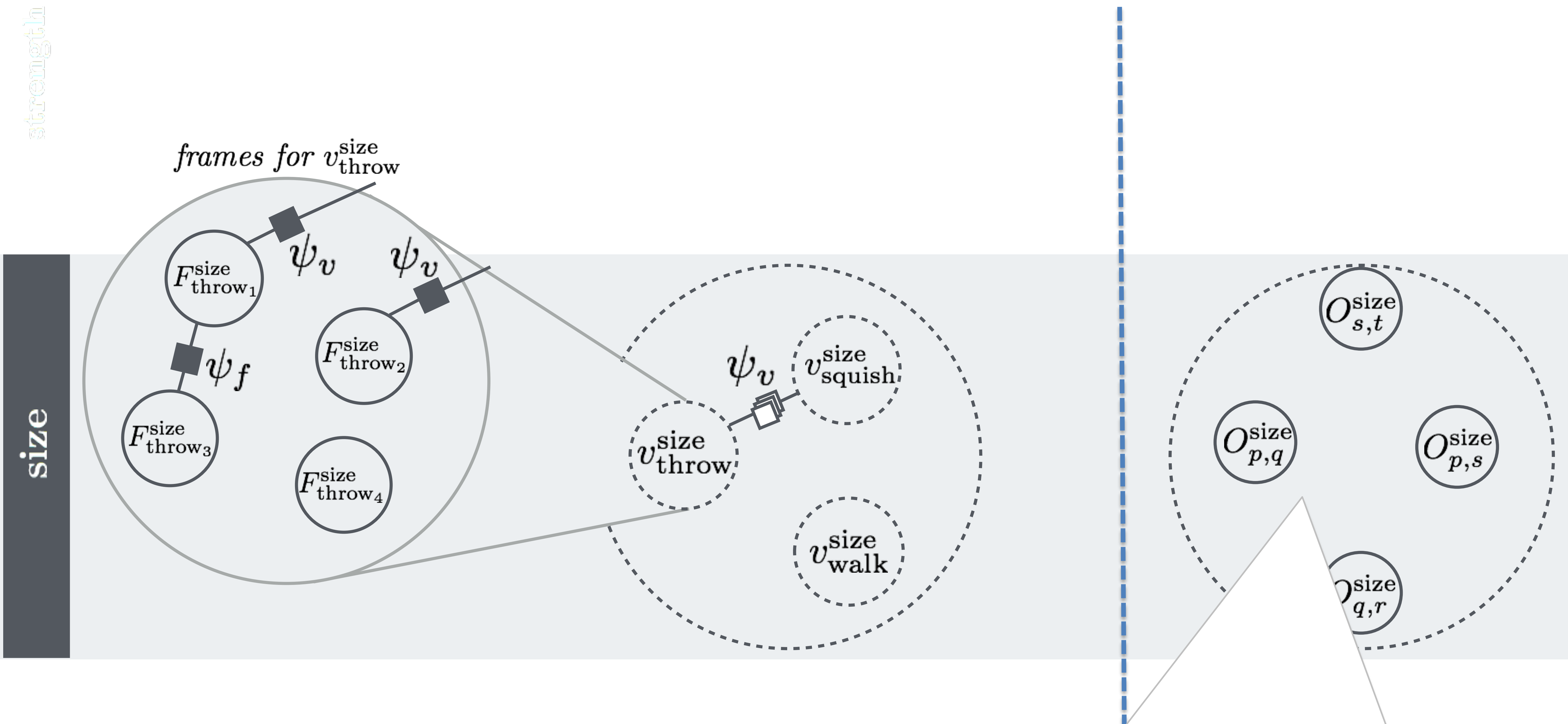


Each variable represents relative object knowledge

"p is generally bigger than q in terms of size"

Or

"p Z q in size" where Z = {>, <, =}



Each variable represents relative object knowledge

"p is generally bigger than q in terms of size"

Or

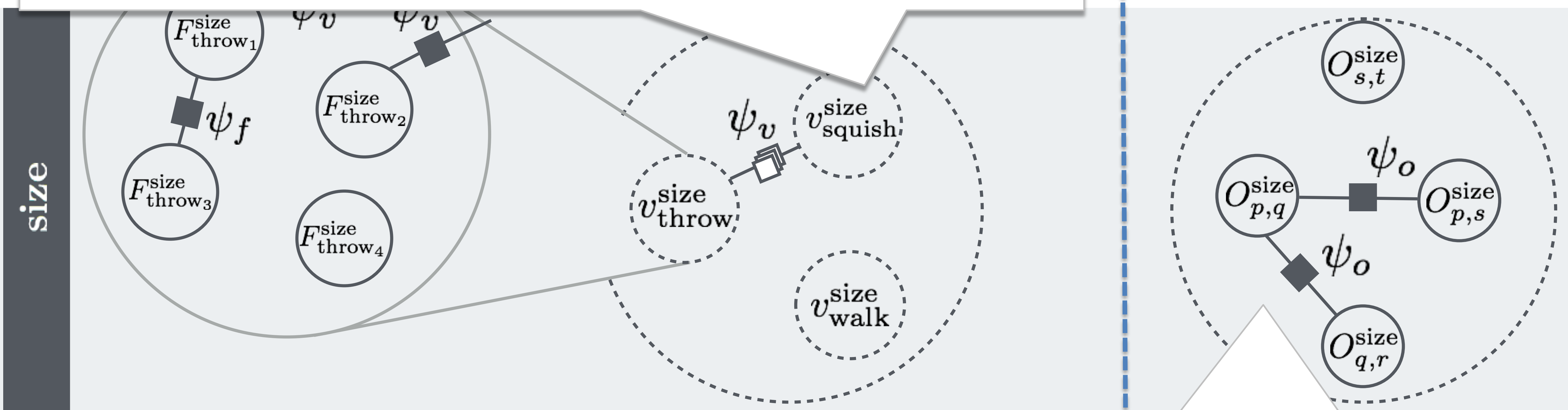
"p Z q in size" where Z = {>, <, =}

Each variable represents an entailment knowledge

"If X throws Y, then X > Y in terms of size"

Or

"If X throws Y, then Z in size" where Z = {>, <, =}



Each variable represents relative object knowledge

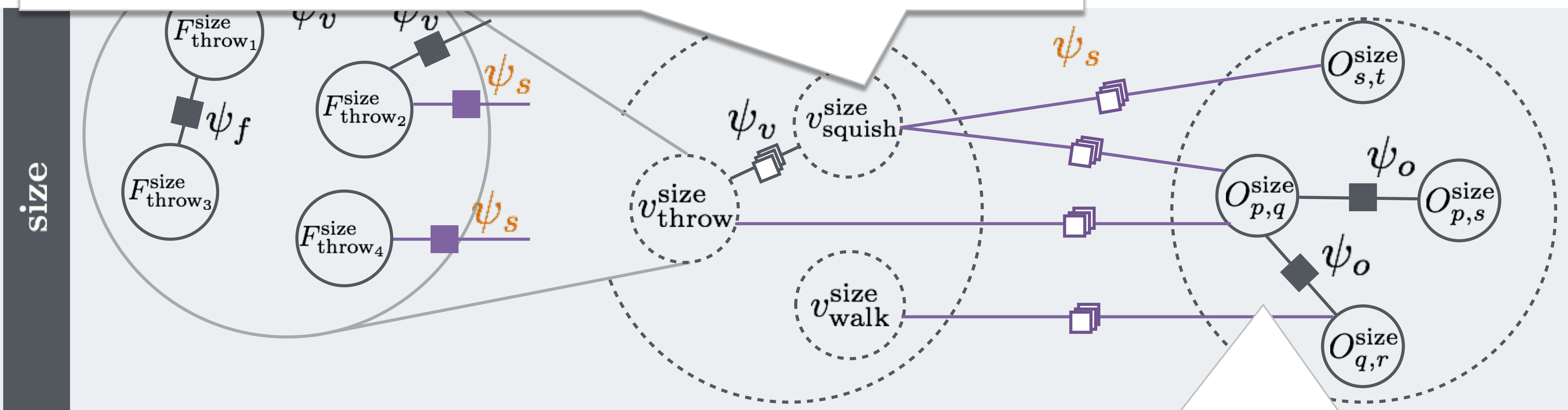
"p is generally bigger than q in terms of size"

Or

"p Z q in size" where Z = {>, <, =}

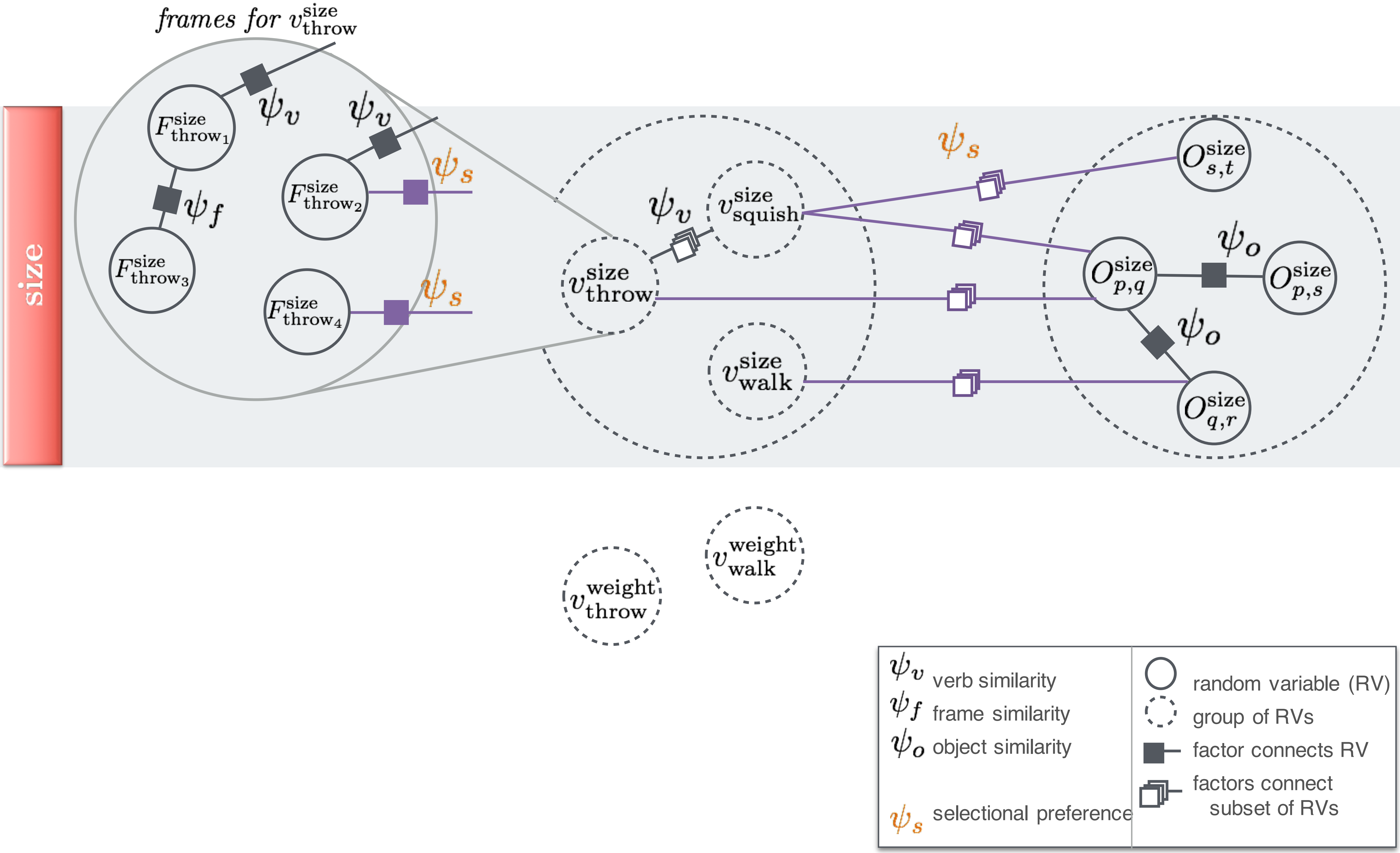
Each variable represents an entailment knowledge

"If X throws Y, then X > Y in terms of size"
Or
"If X throws Y, then Z in size" where Z = {>, <, =}

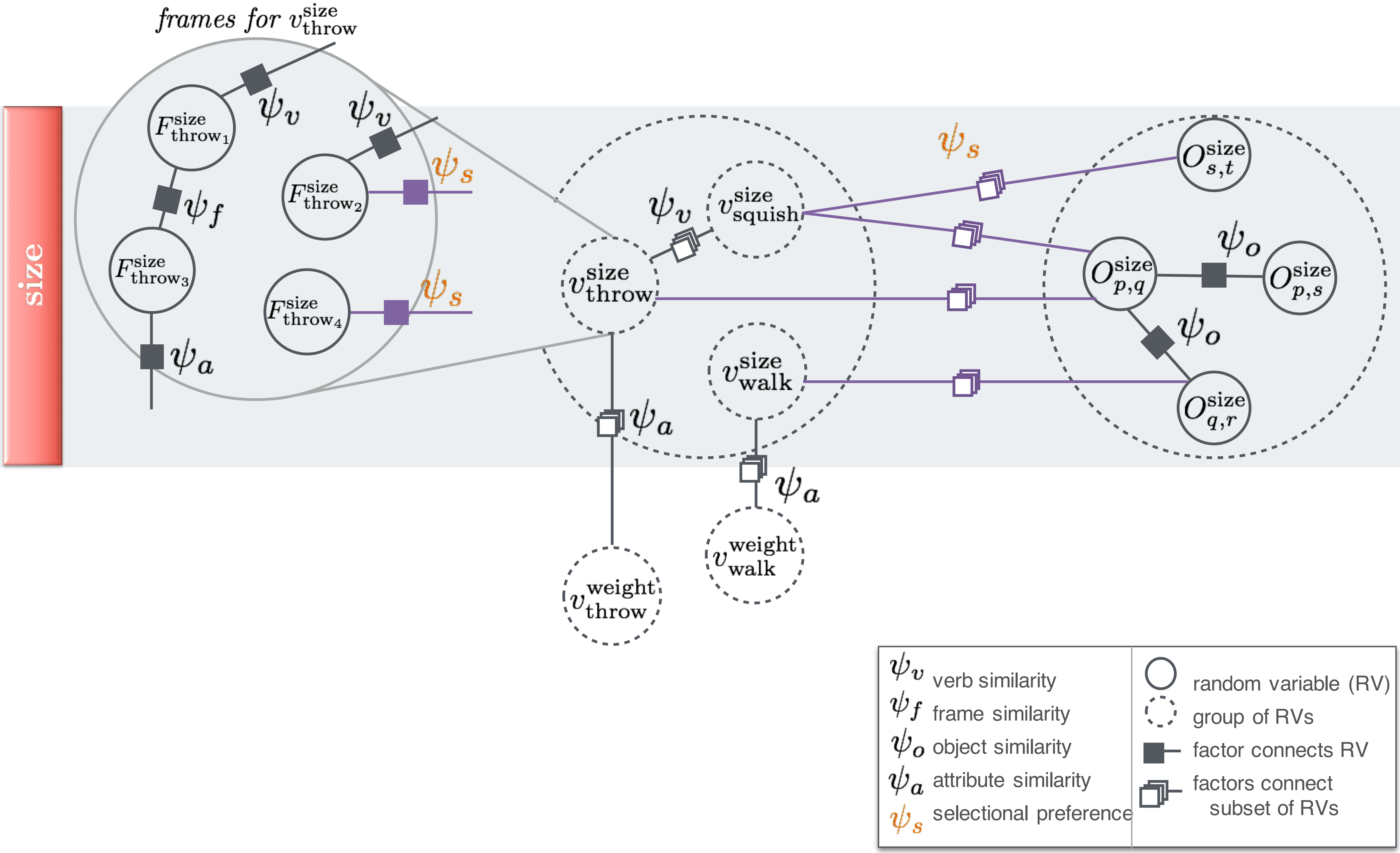


Each variable represents relative object knowledge

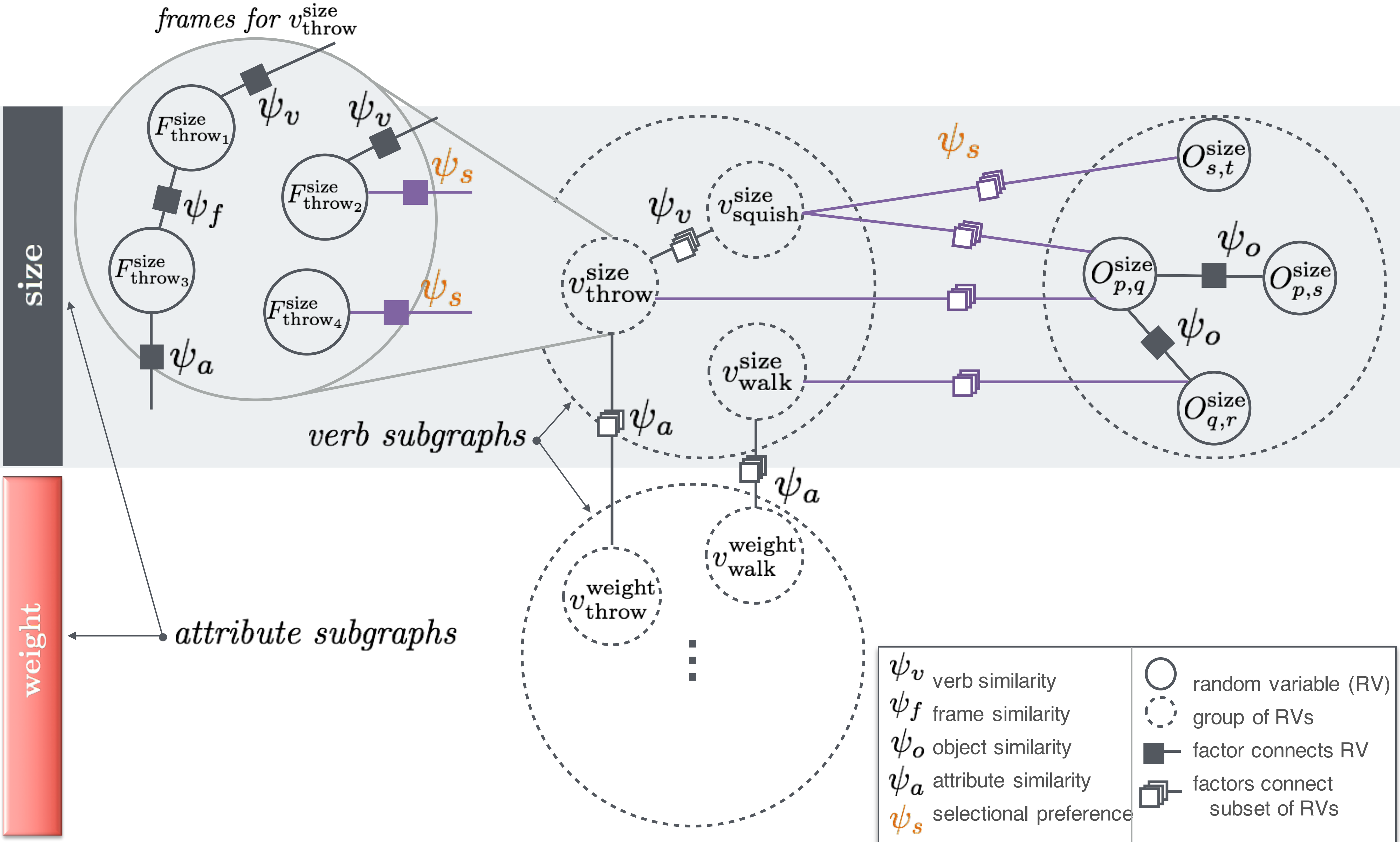
"p is generally bigger than q in terms of size"
Or
"p Z q in size" where Z = {>, <, =}

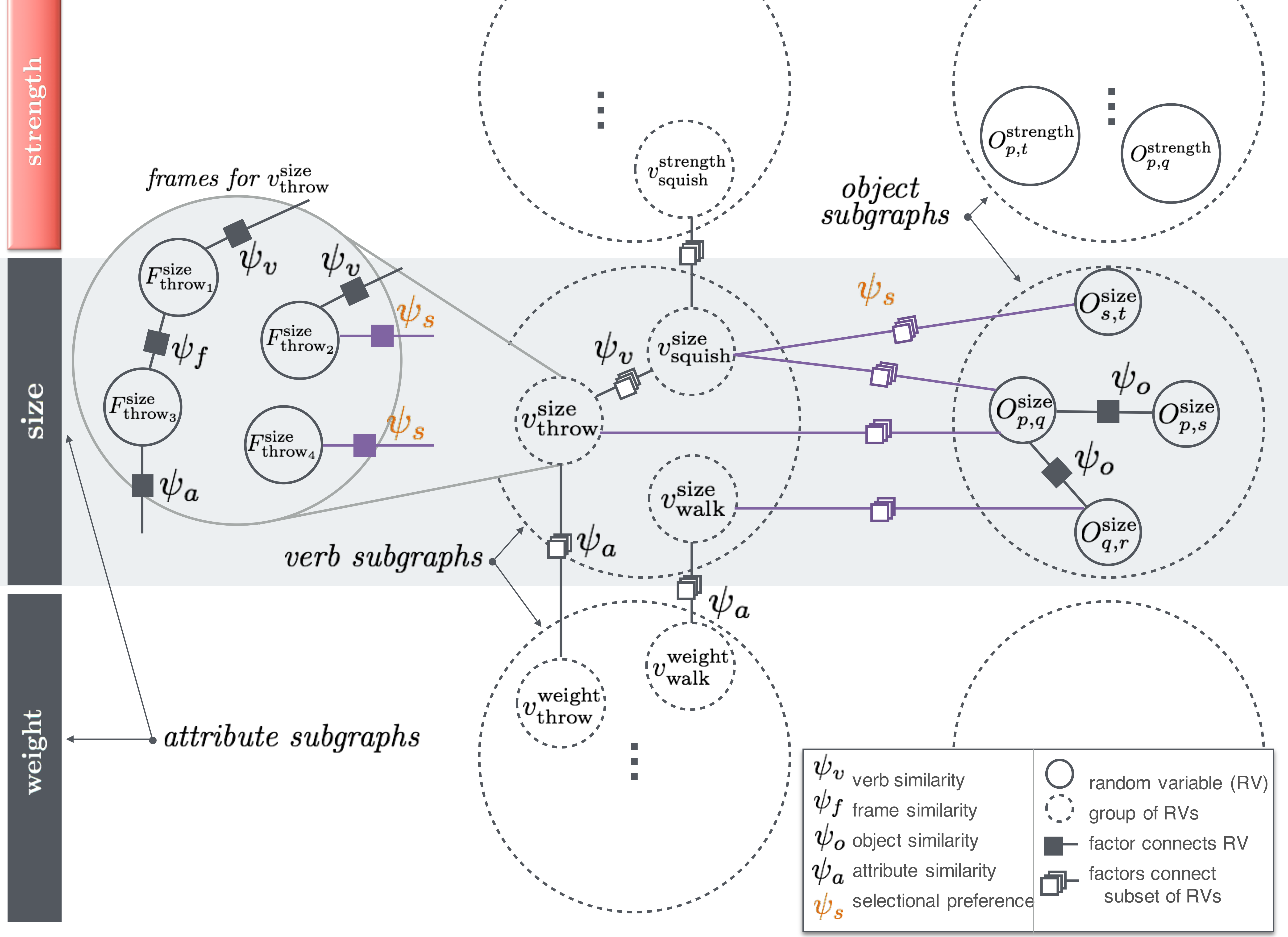


frames for $v^{\text{size}}_{\text{throw}}$
size
$F^{\text{size}}_{\text{throw}_1}$
$F^{\text{size}}_{\text{throw}_2}$
$F^{\text{size}}_{\text{throw}_3}$
$F^{\text{size}}_{\text{throw}_4}$
ψ_v
ψ_f
ψ_s
ψ_o
$v^{\text{size}}_{\text{throw}}$
$v^{\text{size}}_{\text{squish}}$
$v^{\text{size}}_{\text{walk}}$
$O^{\text{size}}_{s,t}$
$O^{\text{size}}_{p,q}$
$O^{\text{size}}_{p,s}$
$O^{\text{size}}_{q,r}$
$v^{\text{weight}}_{\text{throw}}$
$v^{\text{weight}}_{\text{walk}}$
ψ_v verb similarity
ψ_f frame similarity
ψ_o object similarity
ψ_s selectional preference
random variable (RV)
group of RVs
factor connects RV
factors connect subset of RVs



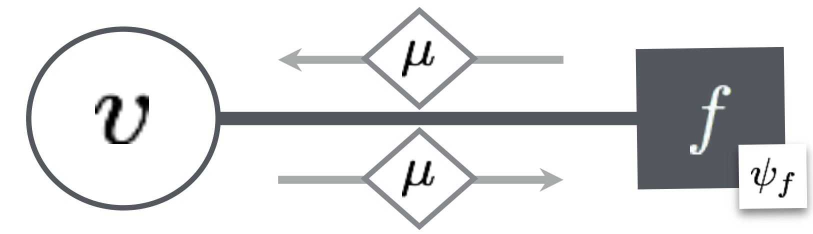
size
frames for $v^{\text{size}}_{\text{throw}}$
$F^{\text{size}}_{\text{throw}_1}$
$F^{\text{size}}_{\text{throw}_2}$
$F^{\text{size}}_{\text{throw}_3}$
$F^{\text{size}}_{\text{throw}_4}$
ψ_v
ψ_f
ψ_a
ψ_s
$v^{\text{size}}_{\text{throw}}$
$v^{\text{size}}_{\text{squish}}$
$v^{\text{size}}_{\text{walk}}$
$v^{\text{weight}}_{\text{throw}}$
$v^{\text{weight}}_{\text{walk}}$
$O^{\text{size}}_{s,t}$
$O^{\text{size}}_{p,q}$
$O^{\text{size}}_{p,s}$
$O^{\text{size}}_{q,r}$
ψ_o
ψ_v verb similarity
ψ_f frame similarity
ψ_o object similarity
ψ_a attribute similarity
ψ_s selectional preference
random variable (RV)
group of RVs
factor connects RV
factors connect subset of RVs


frames for $v^{\text{size}}_{\text{throw}}$
size
weight
$F^{\text{size}}_{\text{throw}_1}$
$F^{\text{size}}_{\text{throw}_2}$
$F^{\text{size}}_{\text{throw}_3}$
$F^{\text{size}}_{\text{throw}_4}$
ψ_v
ψ_f
ψ_s
ψ_a
ψ_o
$v^{\text{size}}_{\text{throw}}$
$v^{\text{size}}_{\text{squish}}$
$v^{\text{size}}_{\text{walk}}$
$v^{\text{weight}}_{\text{throw}}$
$v^{\text{weight}}_{\text{walk}}$
$O^{\text{size}}_{s,t}$
$O^{\text{size}}_{p,q}$
$O^{\text{size}}_{p,s}$
$O^{\text{size}}_{q,r}$
verb subgraphs
attribute subgraphs
ψ_v verb similarity
ψ_f frame similarity
ψ_o object similarity
ψ_a attribute similarity
ψ_s selectional preference
random variable (RV)
group of RVs
factor connects RV
factors connect subset of RVs



strength
size
weight
frames for $v^{\text{size}}_{\text{throw}}$
$F^{\text{size}}_{\text{throw}_1}$
$F^{\text{size}}_{\text{throw}_2}$
$F^{\text{size}}_{\text{throw}_3}$
$F^{\text{size}}_{\text{throw}_4}$
ψ_v
ψ_f
ψ_a
ψ_s
$v^{\text{strength}}_{\text{squish}}$
$v^{\text{size}}_{\text{squish}}$
$v^{\text{size}}_{\text{throw}}$
$v^{\text{size}}_{\text{walk}}$
$v^{\text{weight}}_{\text{throw}}$
$v^{\text{weight}}_{\text{walk}}$
$O^{\text{strength}}_{p,t}$
$O^{\text{strength}}_{p,q}$
$O^{\text{size}}_{s,t}$
$O^{\text{size}}_{p,q}$
$O^{\text{size}}_{p,s}$
$O^{\text{size}}_{q,r}$
ψ_o
object subgraphs
verb subgraphs
attribute subgraphs
ψ_v verb similarity
ψ_f frame similarity
ψ_o object similarity
ψ_a attribute similarity
ψ_s selectional preference
random variable (RV)
group of RVs
factor connects RV
factors connect subset of RVs

Loopy belief propagation



$$\mu_{v\to f}(x) \propto \prod_{f'\in N(v)\setminus\{f\}} \mu_{f'\to v}(x)$$

$$\mu_{f\to v}(x) \propto \sum_{\mathbf{x}'\in\mathbf{x}'_{\setminus x}} \psi(\mathbf{x}') \prod_{v'\in N(f)\setminus\{v\}} \mu_{v'\to f}(x'_{v'})$$

Log-linear factors trained using embeddings

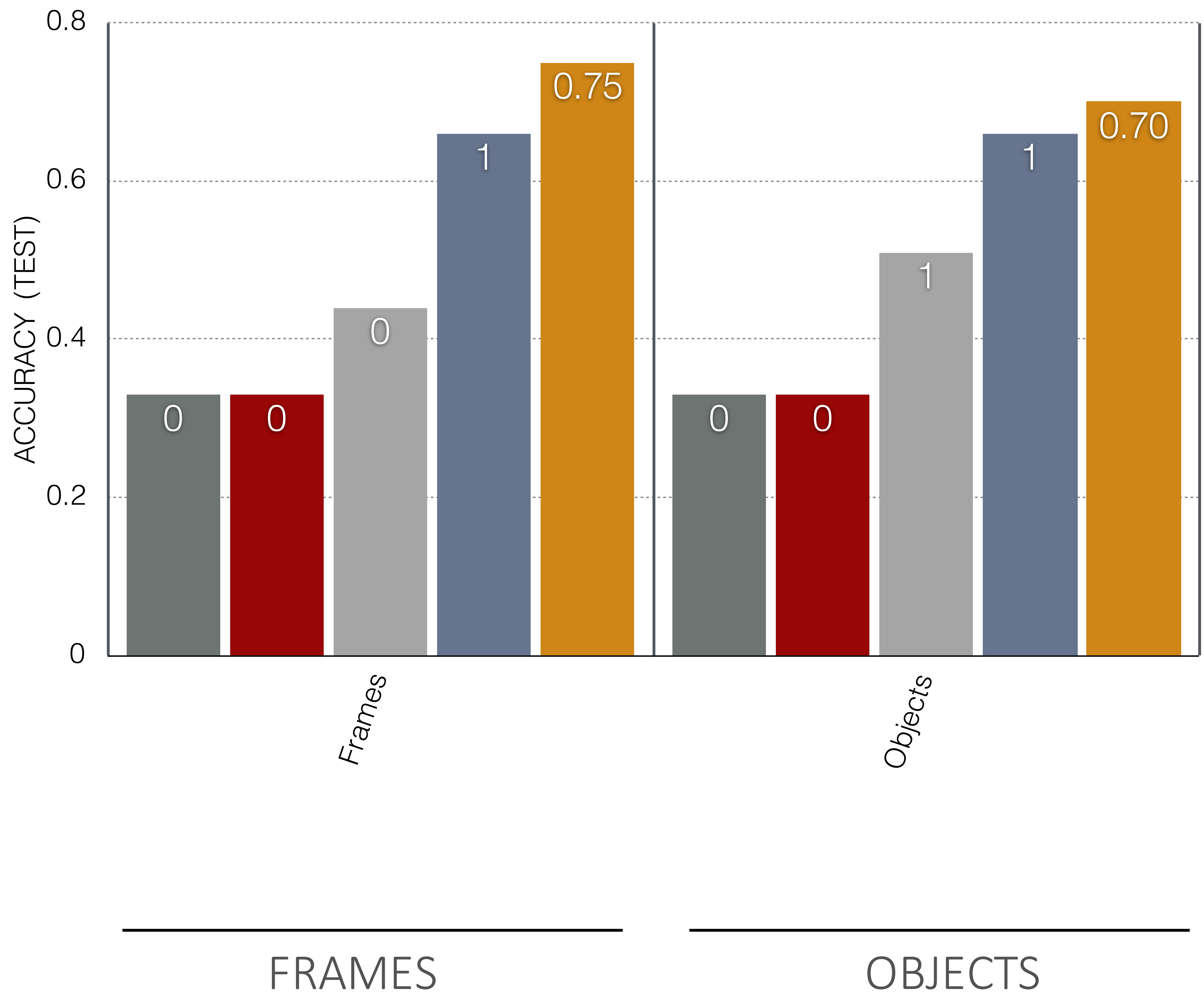
$$P(r|X^a) \propto e^{w_a\cdot f(X^a)}$$

$$f(O^a_{p,q}) := \langle g(p), g(q)\rangle$$

$$f(F^a_{v_t}) := \langle h(t), g(v), g(t)\rangle$$

$g(\cdot)$ GloVe

$h(\cdot)$ One-hot frame type


0.8
0.6
0.4
0.2
0
ACCURACY (TEST)
0
0
0
1
0.75
0
0
1
1
0.70
Frames
Objects
FRAMES
OBJECTS

To Conclude

- Reverse Engineering Commonsense Knowledge
- By solving both puzzles simultaneously:
  - Learn VerbPhysics frames
  - Learn object – object relations

- w.r.t. relative physical knowledge over 5 dimensions:
  - Size, weight, strength, flexibility, speed

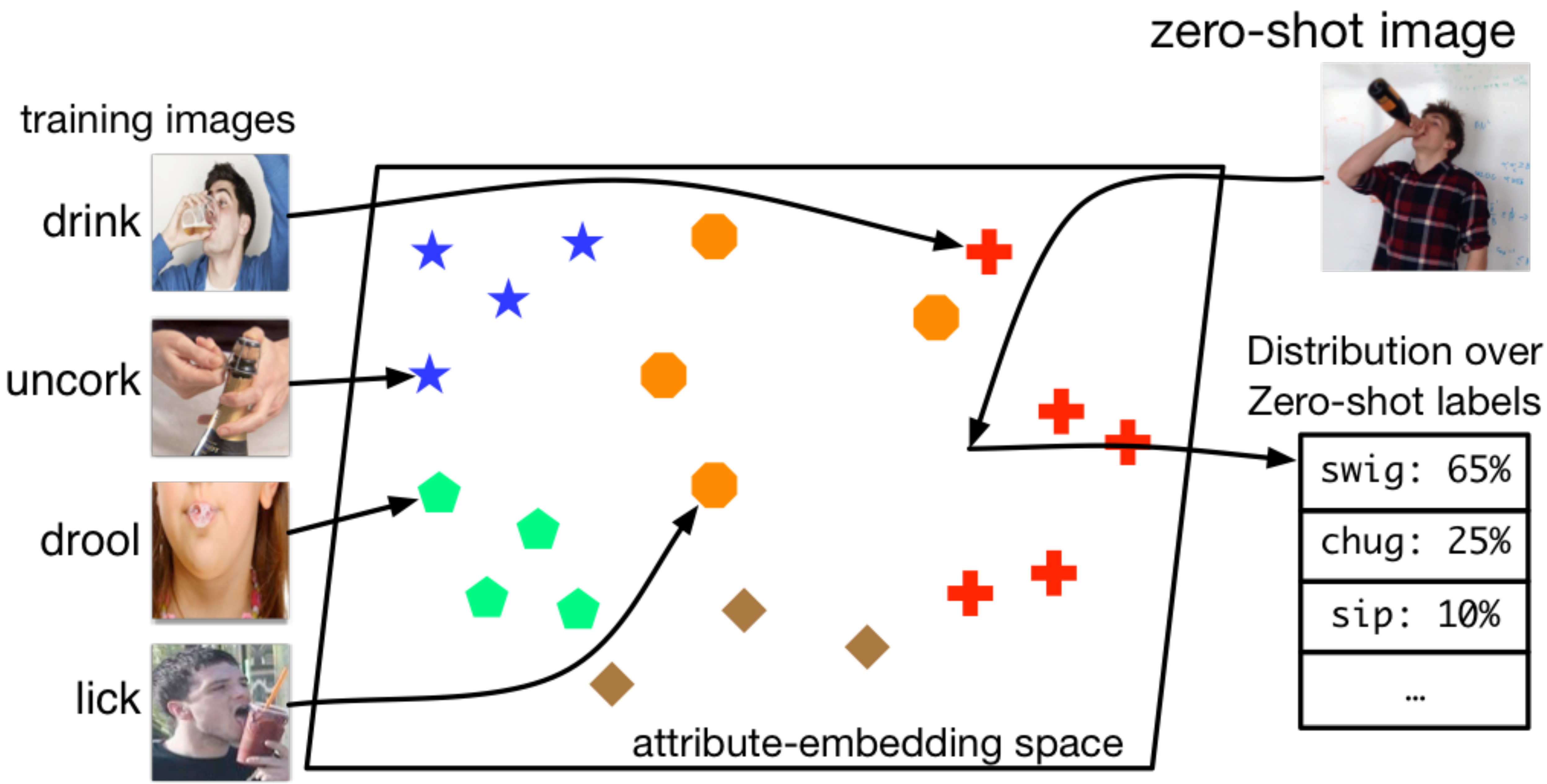
➔ Just from text, without embodiment

➔ Vision can't do weight, strength, flexibility …

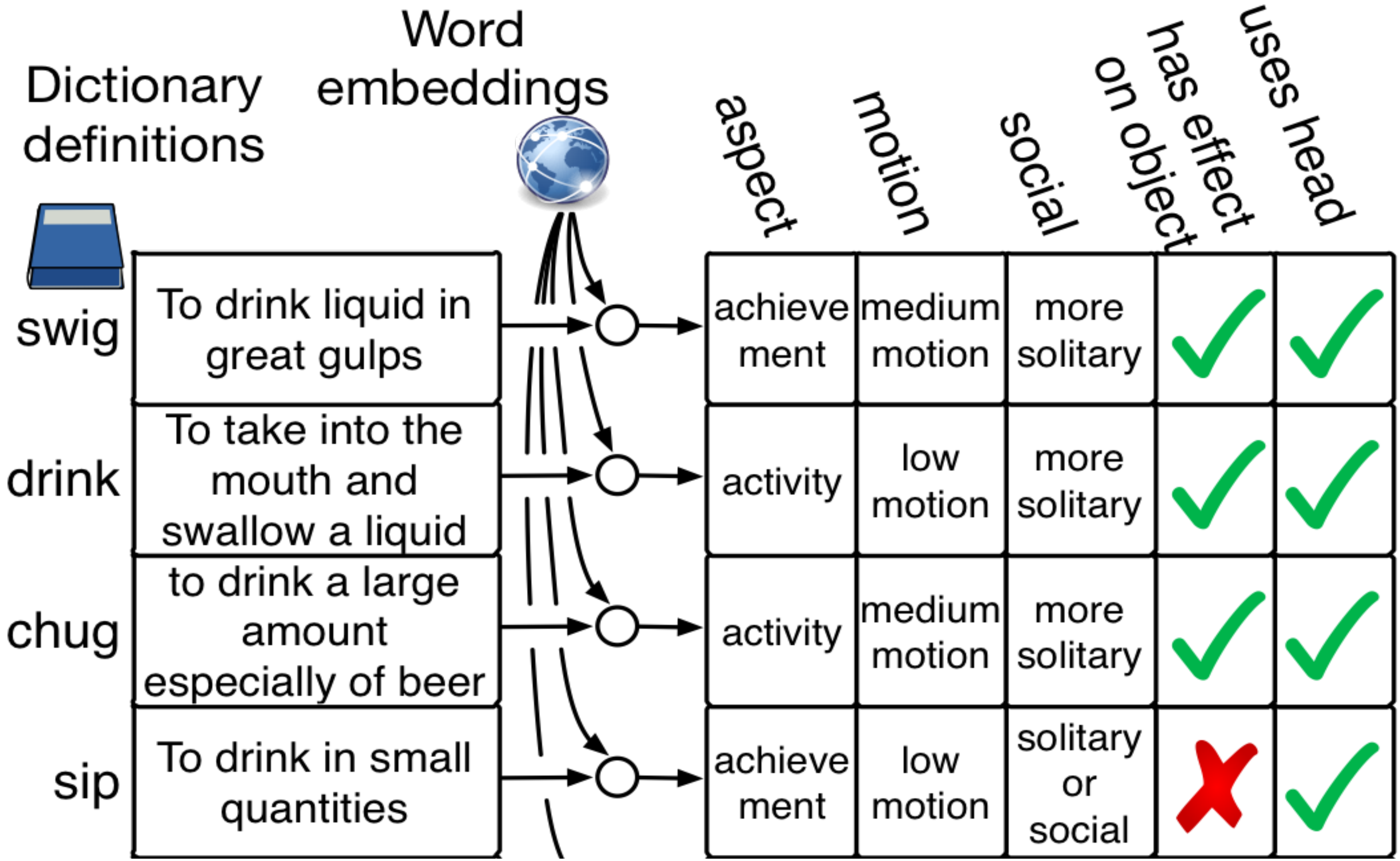
Zero-shot Activity Recognition with Verb Attribute Induction

Rowan Zellers et al. @ EMNLP 2017

Zero-shot Activity Recognition



Verb Attribute Induction



Plan

1. Commonsense Frame Semantics
   – Naive physics
   – Social commonsense

2. Modeling the World, not just Language

Connotation Frames

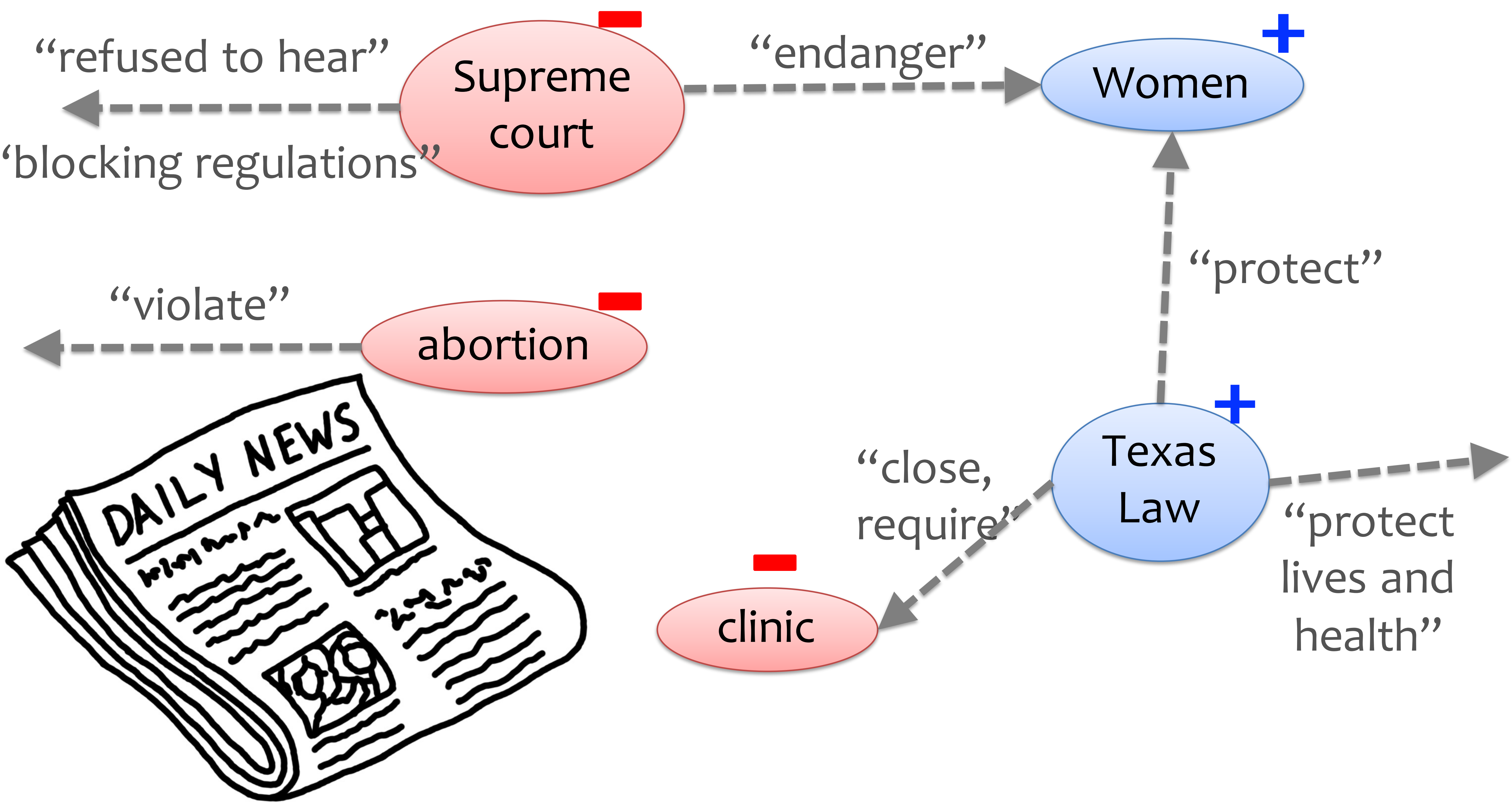
Hannah Rashkin et al. (ACL 2016)

How Language Describes the World

"All I know is what I read in the papers."
– Will Rogers

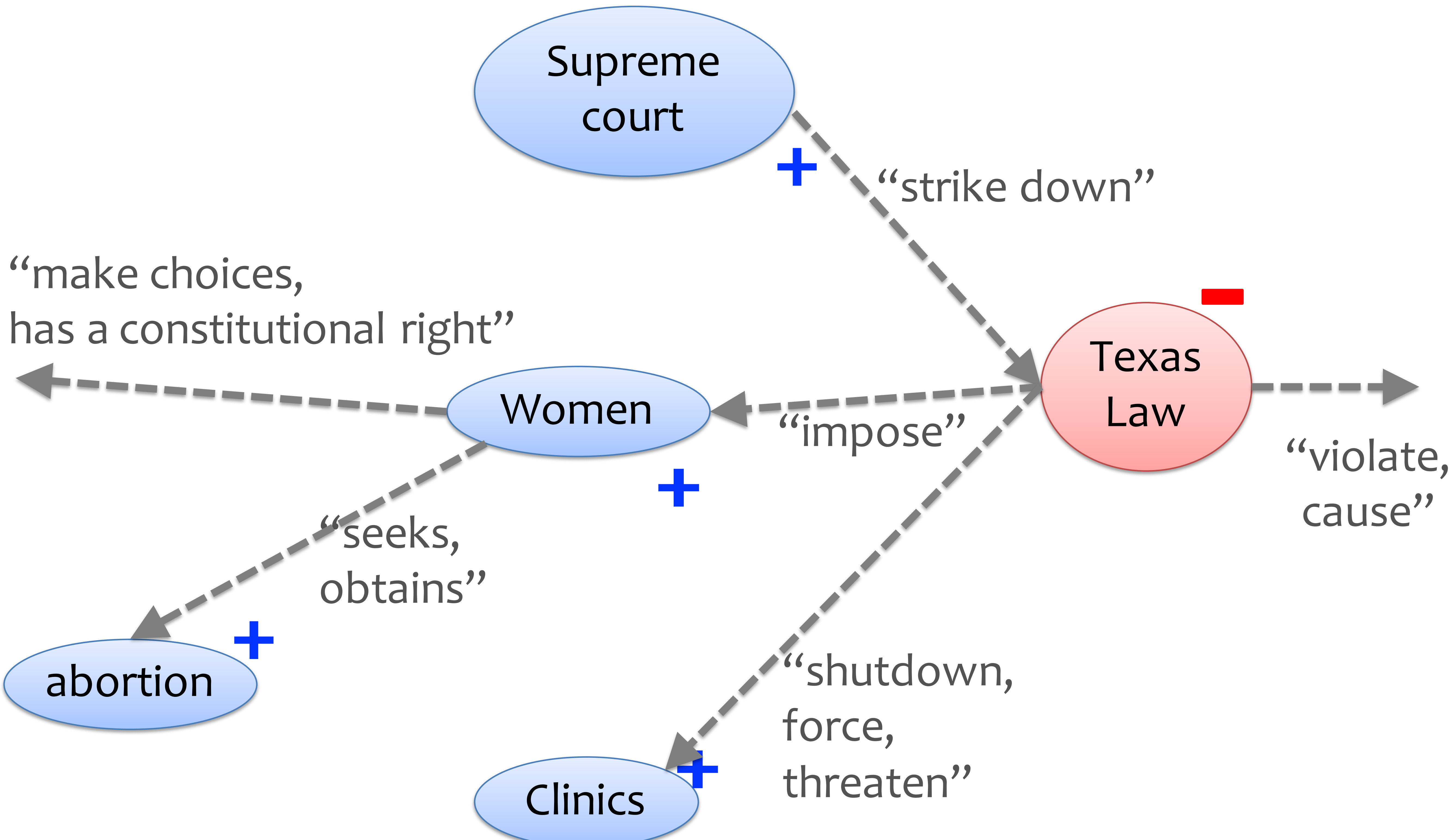
How Language Describes the World

(per pro-life)



An Alternative World View

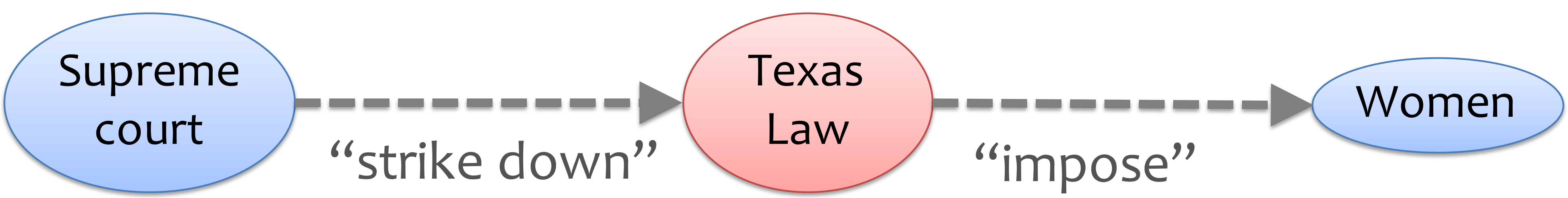
(per pro-choice)



Prior Literature

- A ton of work on sentiment analysis
  - Especially on product reviews and movie reviews
  - (Pang and Lee 2009 for survey, Socher et al 2013)

- Recent work on implied sentiment
  - (Greene and Resnik 2009 , Feng et al 2013; Choi and Wiebe 2014; Mohammad and Turney 2010; Deng and Wiebe 2014)



"criticize" frame

per *left*-leaning sources:

Obama *criticize* someone?

or

Obama *is critized by* someone?

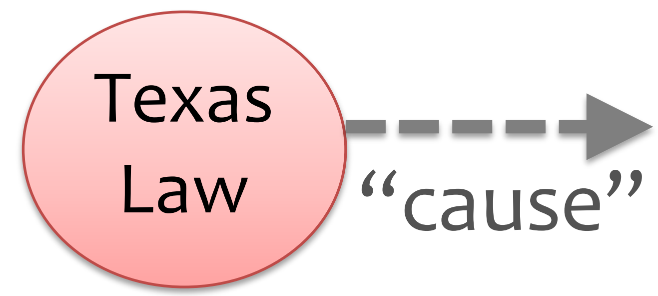
(Based on the Stream Corpus (2014) w.r.t. 30 news sources with known liberal / conservative leanings)

Connotative Meaning of Words

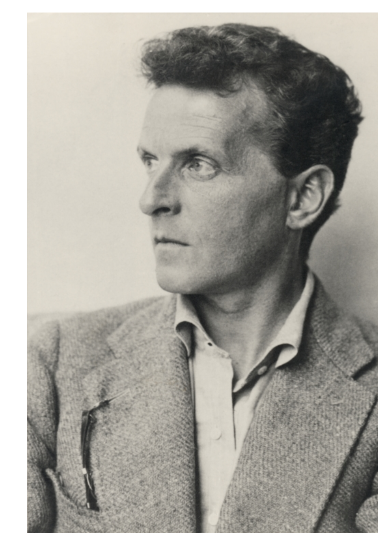
- Semantic prosody (Sinclair 1991, Louw 1993)

  "_______ *caused* _______"

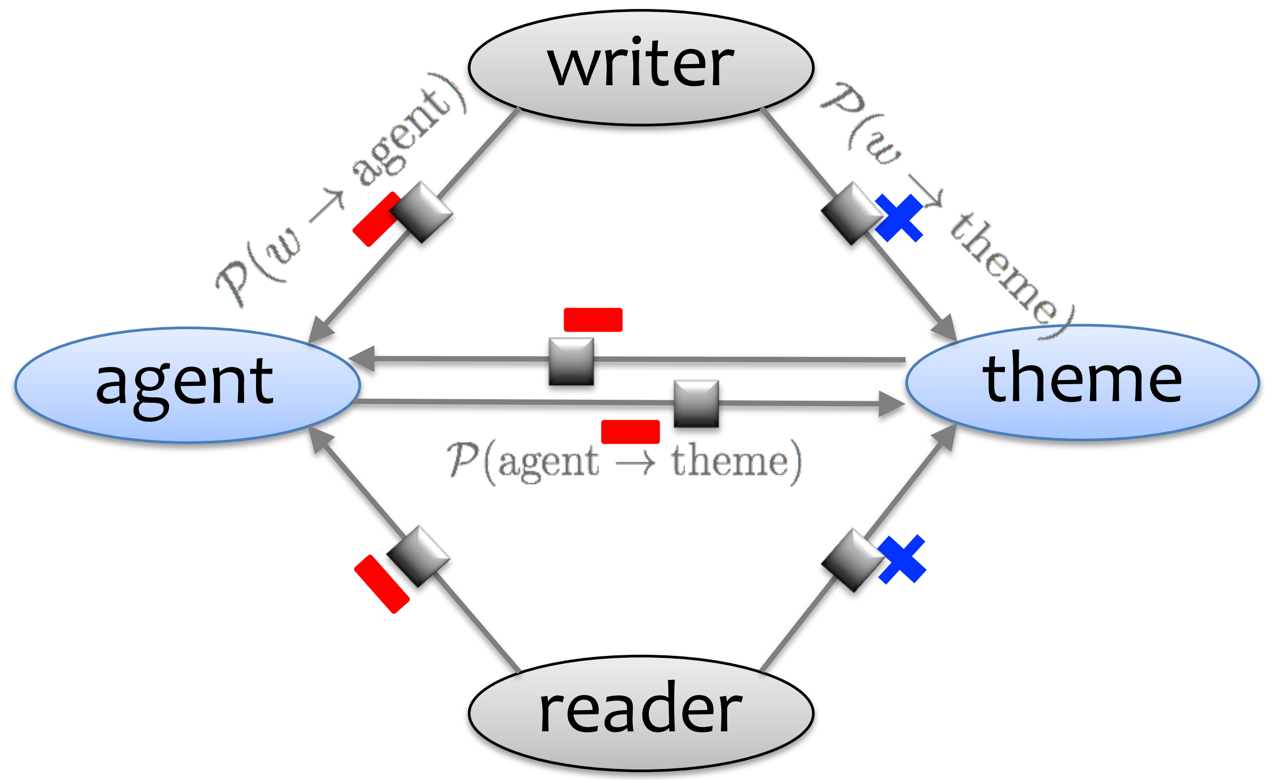
  headache, cancer, …
  appreciation, award?



- Wittgenstein's view:
  - the meaning of a word is its use in the language
- Connotation arises from the (typical) context in which words are used
- Can we encode connotation in frame semantics?



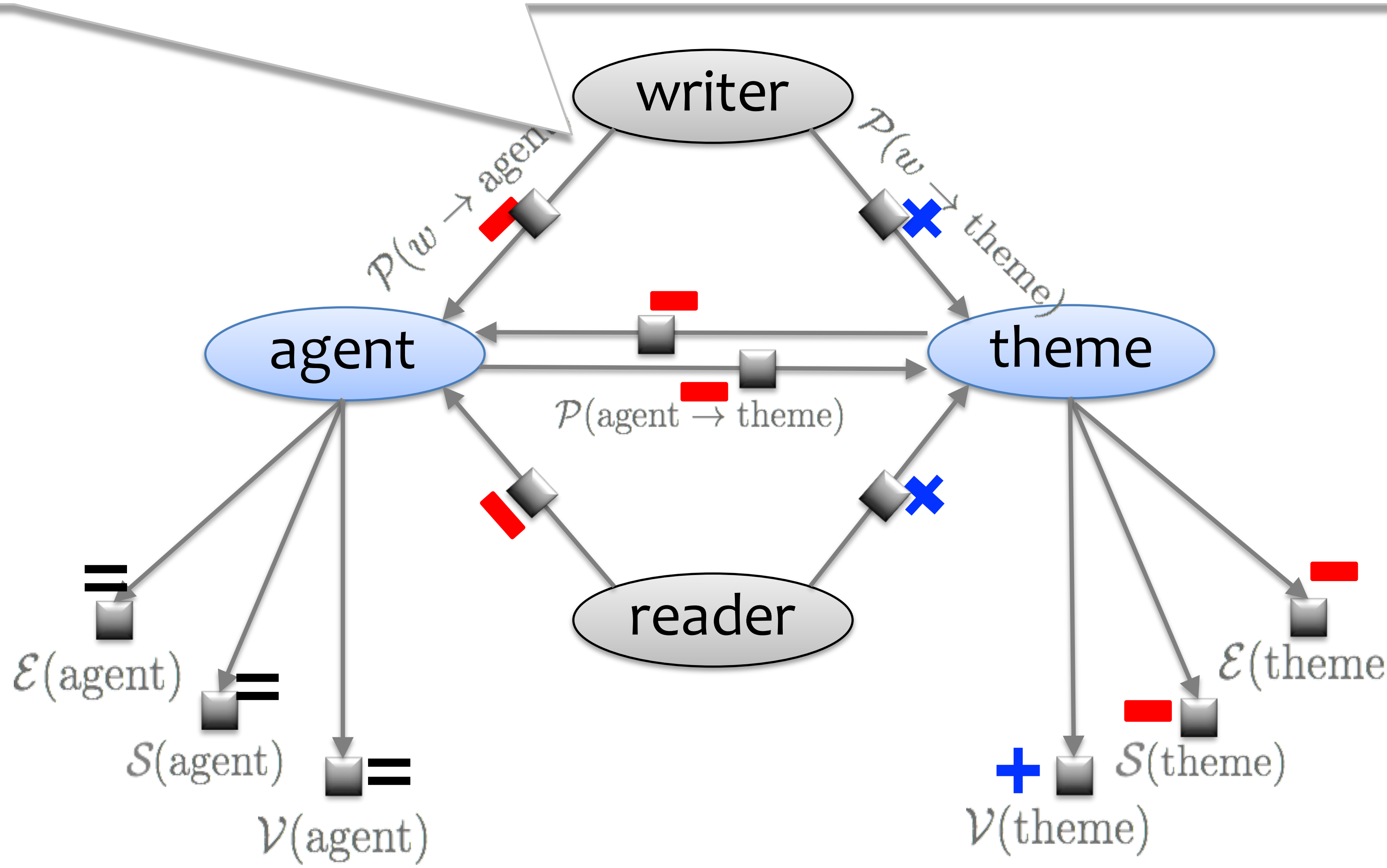
“violate” frame



$\mathcal{P}(w \to \text{agent})$

- **P**erspective (x->y) : directed sentiment

Connotation judgments are not always categorical, some are better to be modeled as distributions. (analogously to de Marneffe et al. 2012 on veridicality assessment)



- $\mathcal{P}(w \rightarrow \text{agent})$ **P**erspective (x->y) : directed sentiment
- $\mathcal{E}(\text{agent})$ **E**ffect (x) : whether the event is good for (or bad for) x
- $\mathcal{S}(\text{agent})$ **S**tate (x) : the private state of x as a result of the event
- $\mathcal{V}(\text{agent})$ **V**alue (x) : the intrinsic value of x

“criticize” frame

per *left*-leaning sources:

Obama *criticize* someone?

or

Obama *is critized by* someone?

(Based on the Stream Corpus (2014) w.r.t. 30 news sources with known liberal / conservative leanings)

Left/right leaning sources

(Based on the Stream Corpus (2014) w.r.t. 30 news sources with known liberal / conservative leanings)

Verb	Role	Left-leaning sources	Right-leaning sources
accuse	subject [-]	Putin, Progressives, Limbaugh, Gingrich	activist, U.S., protestor, Chavez
	object [+]	Official, rival, administration, leader	Romney, Iran, Gingrich, regime, opposition
attack	subject [-]	McCain, Trump, Limbaugh	Obama, campaign, Biden, Israel
	object [+]	Gingrich, Obama, policy	citizen, Zimmerman
criticize	subject [-]	Ugandans, rival, Romney, Tyson	Britain, passage, Obama, Maddow
	object [+]	sensation, Obama, Allen, Cameron, Congress	threat, Pelosi, Romey, GOP, Republicans
suffer	subject [+]	woman, victim, officer, father	Christie, Rangers, people, man
	object [-]	attack, blow, burn, seizure, casualty	injury, fool, loss, concussion

Connotation Frames

1. Analysis on large-scale news corpora
2. Crowdsourcing experiments

Writer: The judge upheld the verdict.
Questions: How the judge feels about the verdict:

Negative	Negative or Neutral	Neutral	Positive or Neutral	Positive

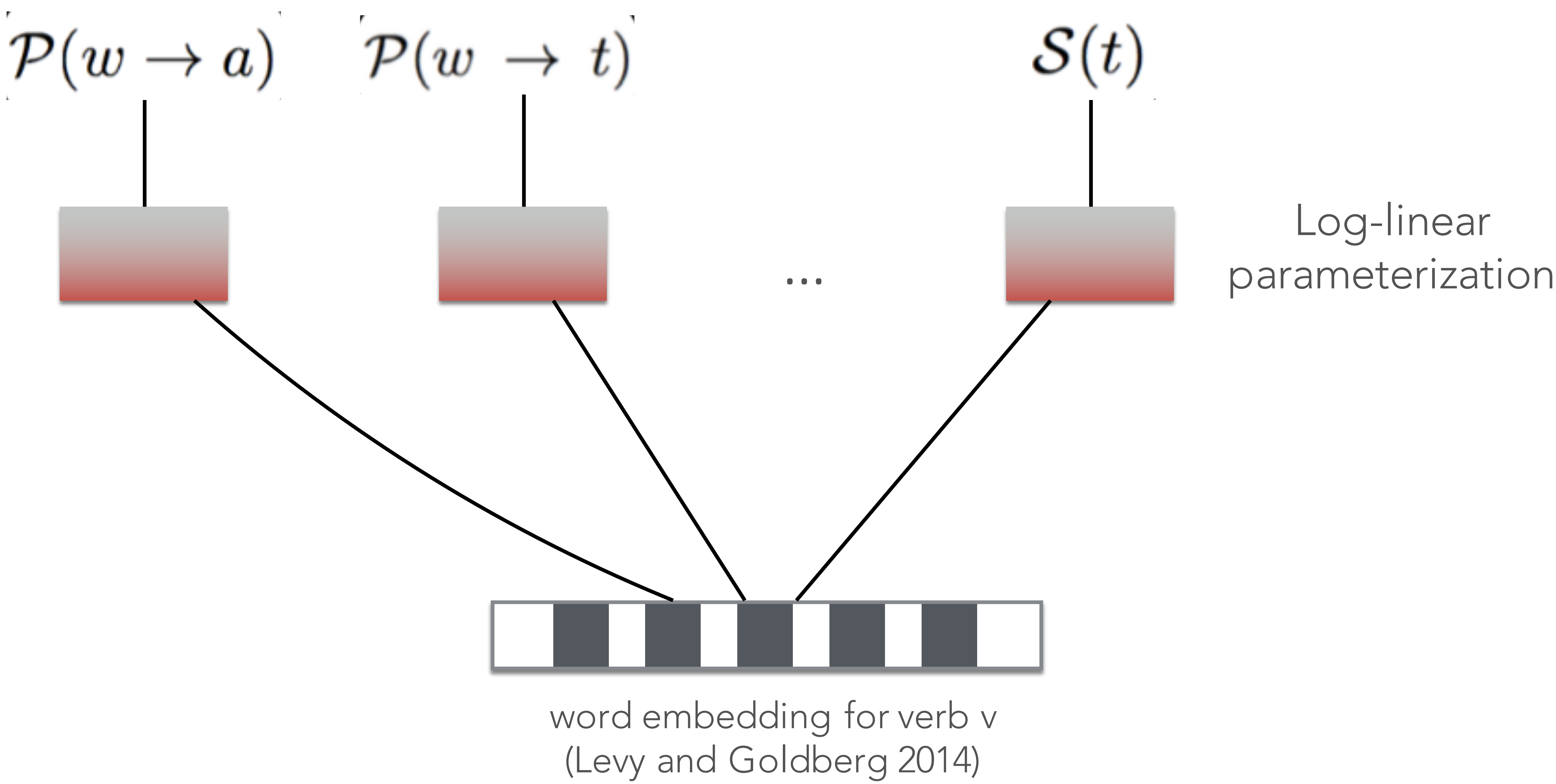
15 annotators answered 16 questions for each of the top 1000 verbs from the NY times corpus.
Average Krippendorff's alpha is **0.25** and 2-way soft agreement is **95%**.

3. Computational models

Computational Models

Base Model:
Embedding to Connotation Frames

- Frame lexicon induction
- Separate frame prediction for each predicate

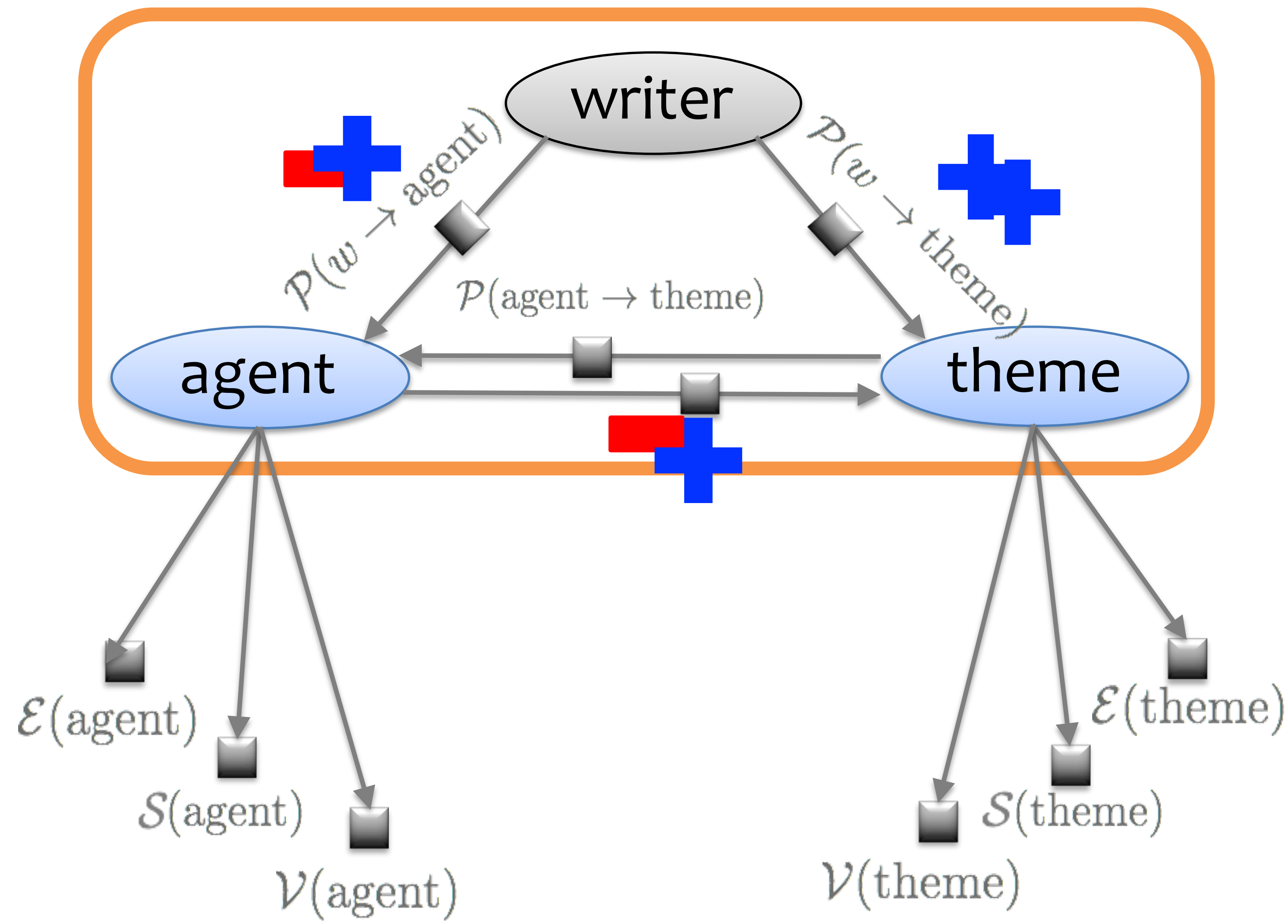


Connotation Dynamics

1. **Perspective Triads:**

$$\mathcal{P}_{w \to x_i} = \neg \left(\mathcal{P}_{w \to x_j} \oplus \mathcal{P}_{x_i \to x_j} \right)$$

~ social balance theory (Cartwright 1956)



Connotation Dynamics

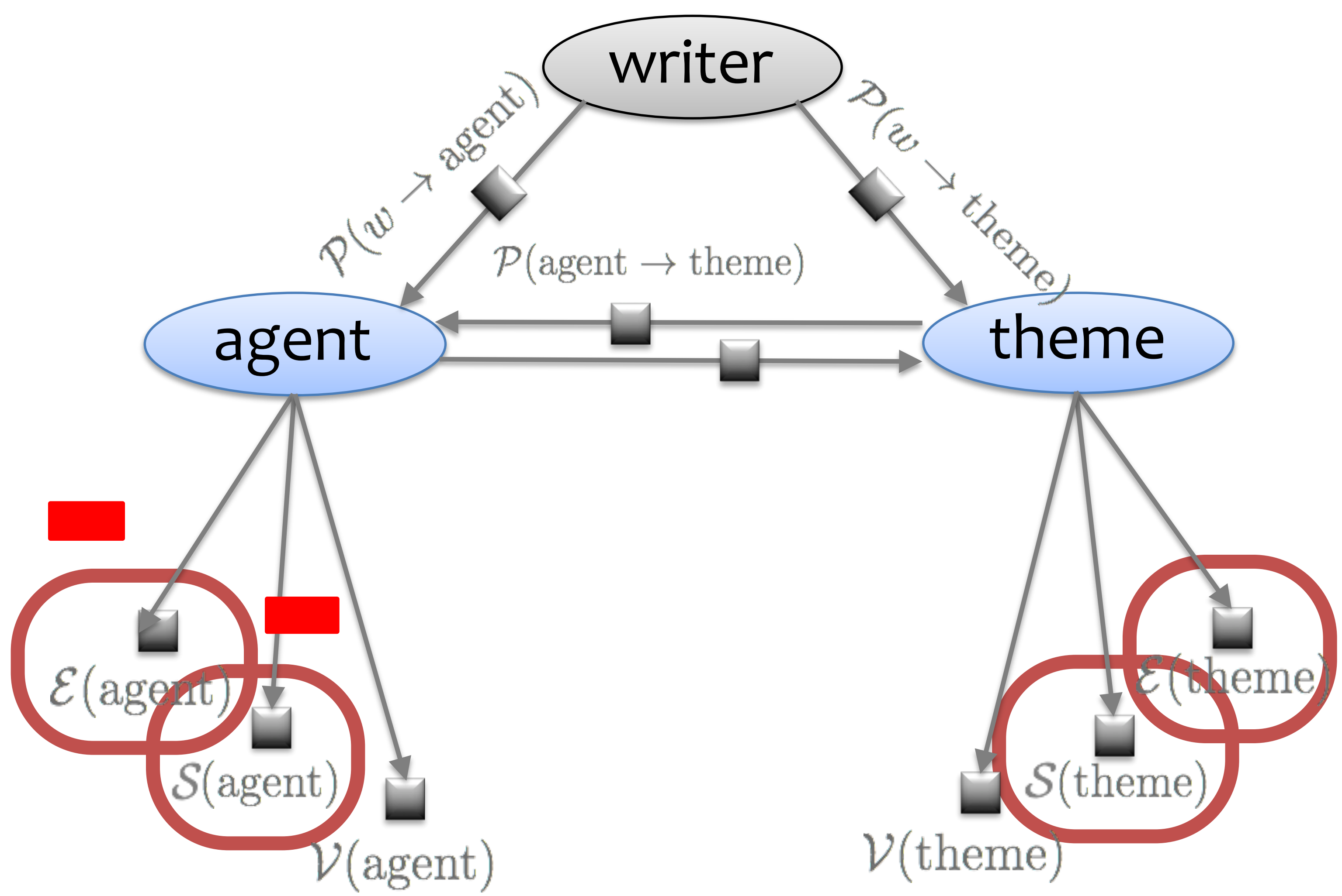
1. **Perspective Triads:**

$$\mathcal{P}_{w \to x_i} = \neg \left(\mathcal{P}_{w \to x_j} \oplus \mathcal{P}_{x_i \to x_j} \right)$$

~ social balance theory (Cartwright 1956)

2. **Effect – Mental State:**

$$\mathcal{S}_{x_i} = \mathcal{E}_{x_i}$$



Connotation Dynamics

1. **Perspective Triads:**

$$\mathcal{P}_{w \rightarrow x_i} = \neg \left(\mathcal{P}_{w \rightarrow x_j} \oplus \mathcal{P}_{x_i \rightarrow x_j} \right)$$

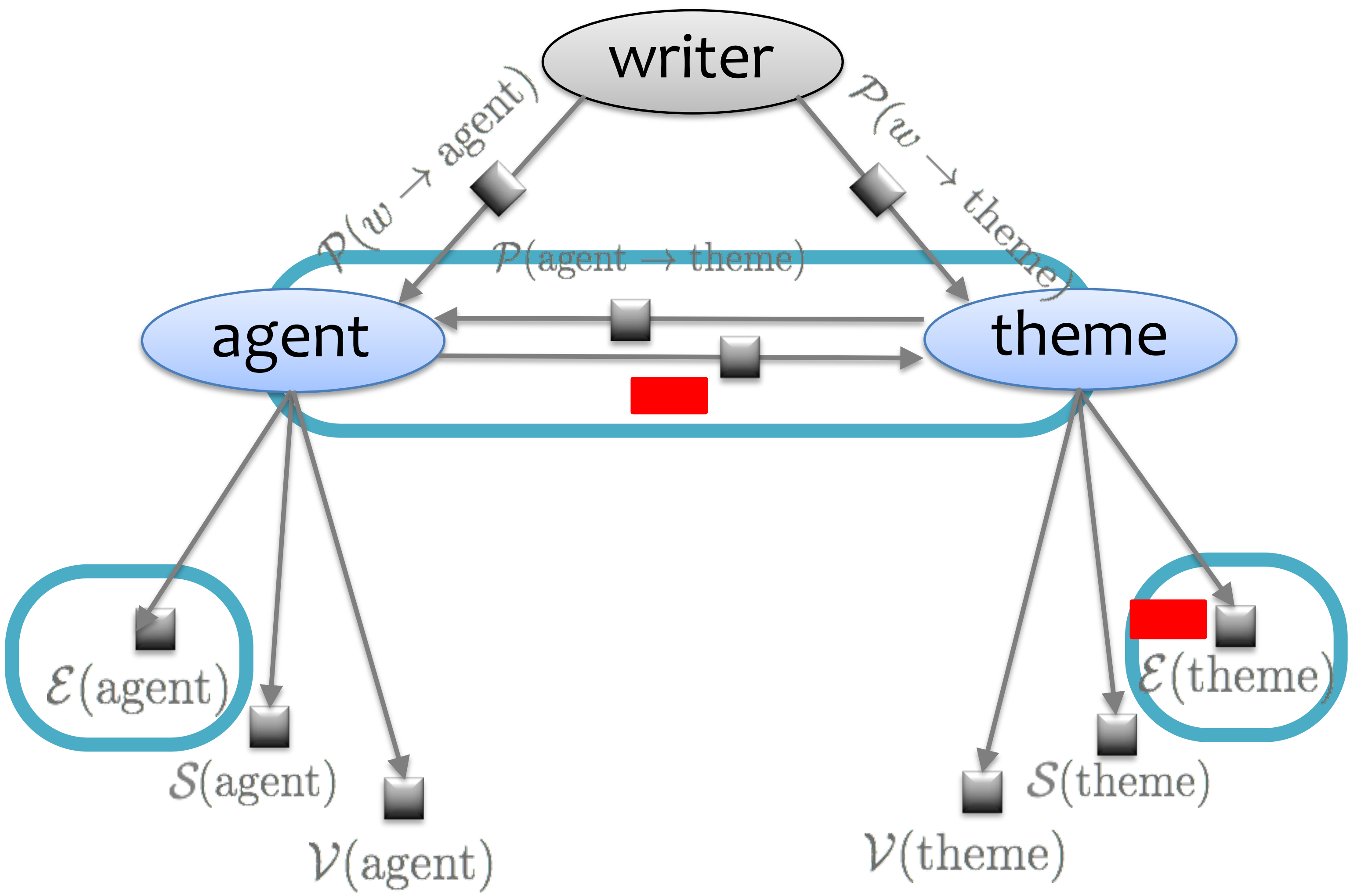
~ social balance theory (Cartwright 1956)

2. **Effect – Mental State:**

$$\mathcal{S}_{x_i} = \mathcal{E}_{x_i}$$

3. **Effect – Perspective:**

$$\mathcal{E}_{x_i} = \mathcal{P}_{x_j \rightarrow x_i}$$



Connotation Dynamics

1. **Perspective Triads:**

$$\mathcal{P}_{w \rightarrow x_i} = \neg \left(\mathcal{P}_{w \rightarrow x_j} \oplus \mathcal{P}_{x_i \rightarrow x_j} \right)$$

~ social balance theory (Cartwright 1956)

2. **Effect – Mental State:**

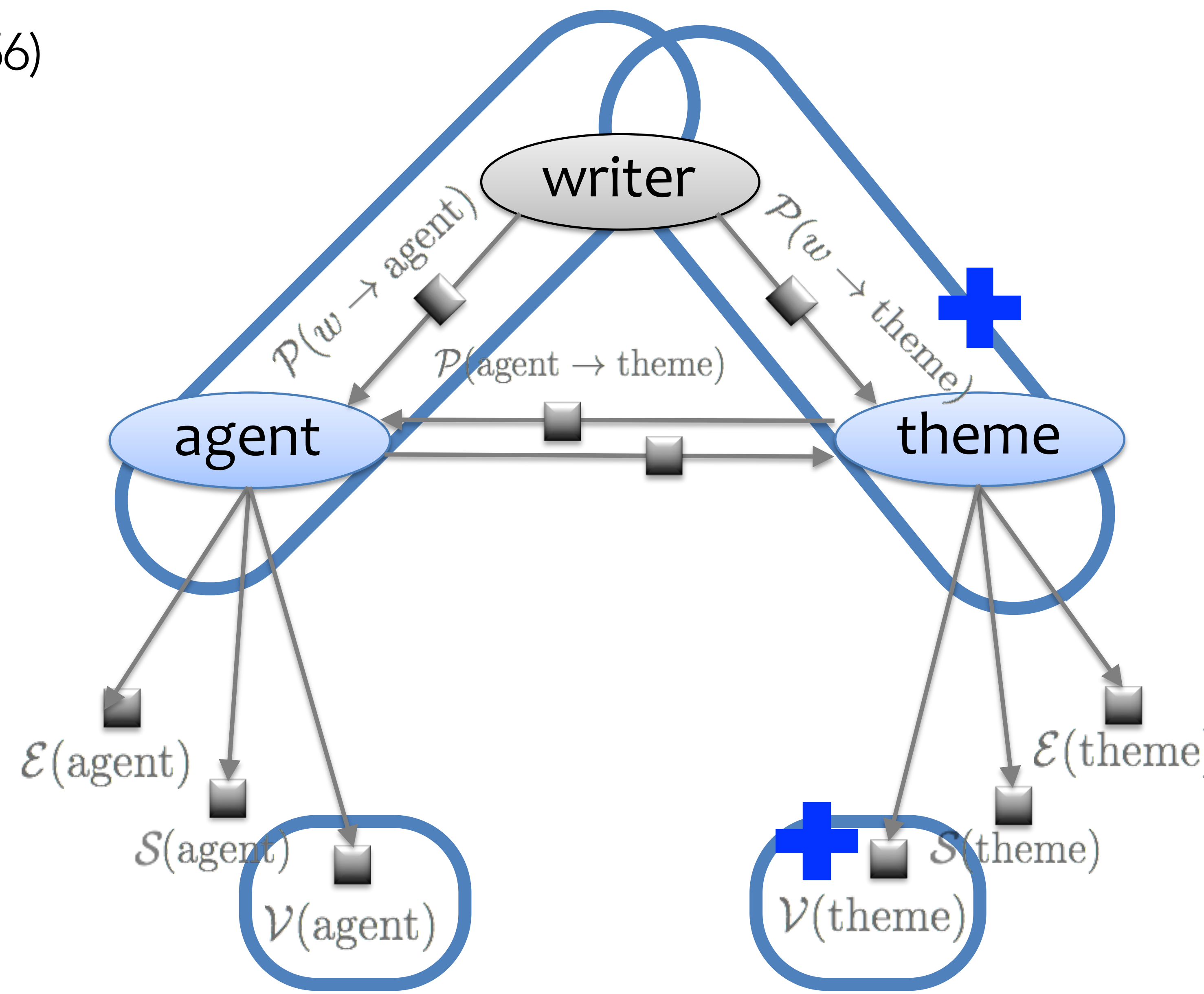
$$\mathcal{S}_{x_i} = \mathcal{E}_{x_i}$$

3. **Effect – Perspective:**

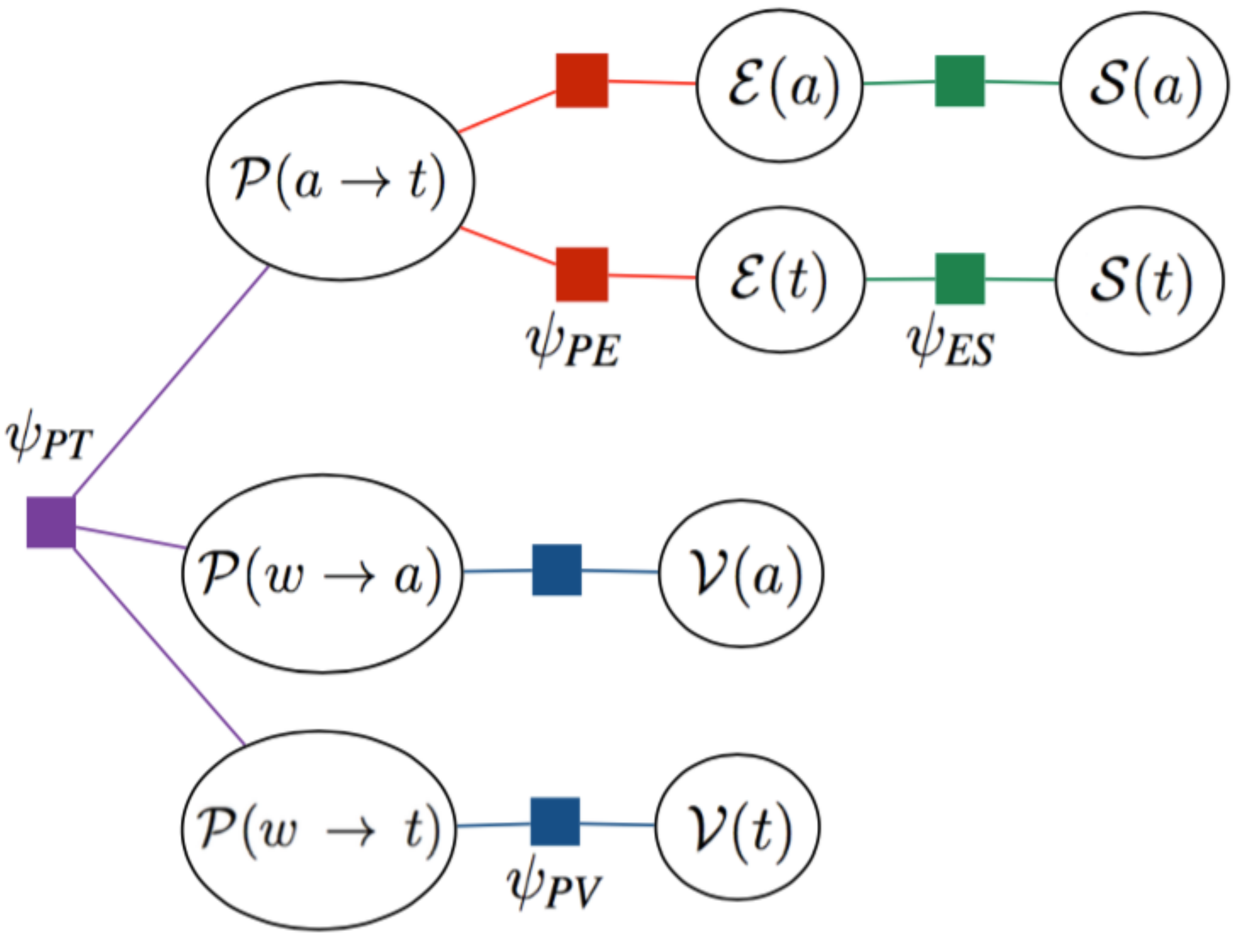
$$\mathcal{E}_{x_i} = \mathcal{P}_{x_j \rightarrow x_i}$$

4. **Value – Perspective:**

$$\mathcal{V}_{x_i} = \mathcal{P}_{w \rightarrow x_i}$$



Factor Graph Model



Inference: Message passing:

$$\mu_{v \to a}(x) \propto \prod_{a^* \in N(v)\ a} \mu_{a^* \to v}(x)$$

$$\mu_{a \to v} \propto \sum_{x', x'_v = x} \psi(x') \prod_{v^* \in N(a)\ v} \mu_{v^* \to a}(x'_{v^*})$$

Factor potentials:

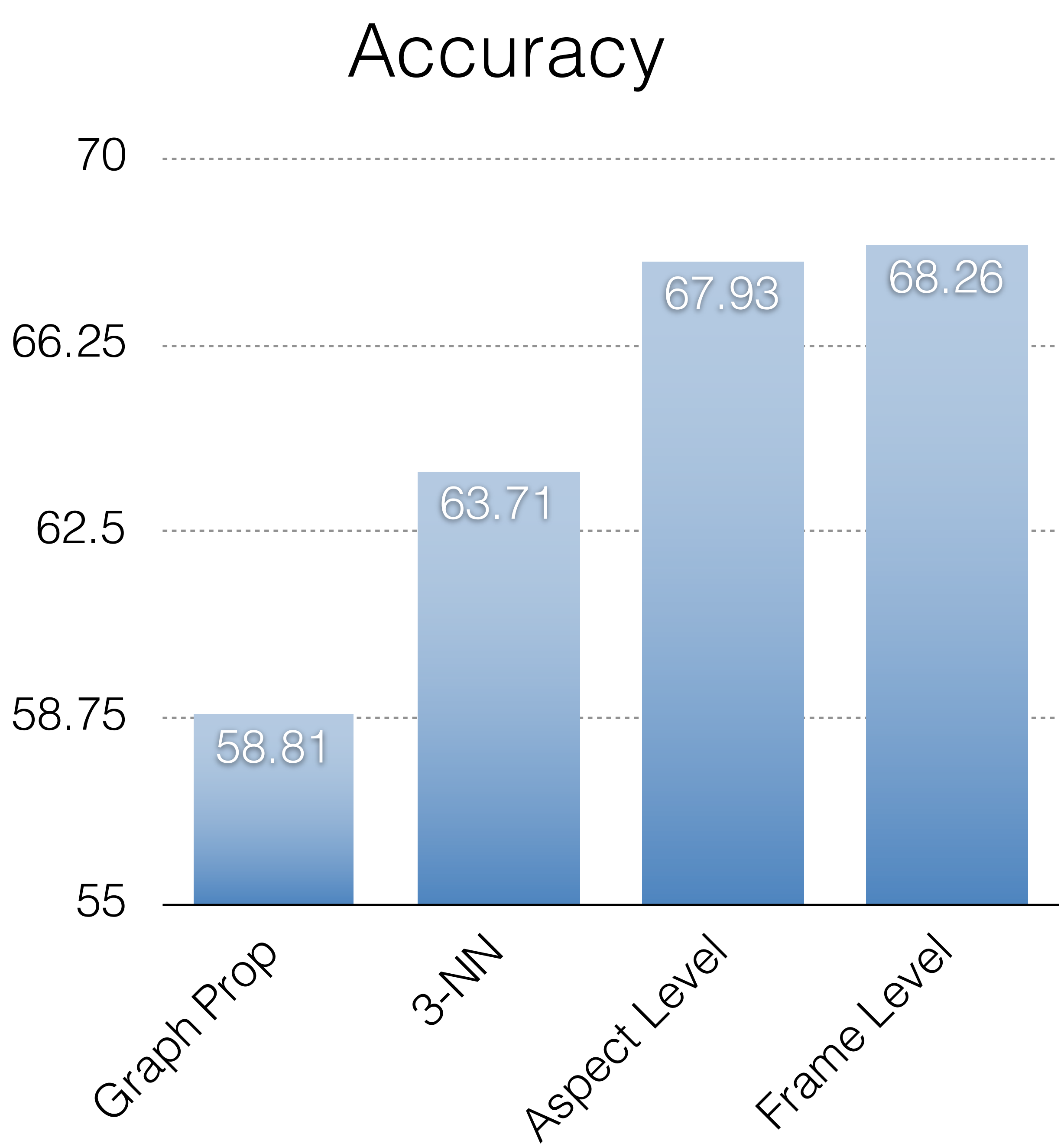
$$\psi_{ES}(\mathcal{E}_a, \mathcal{S}_a) = e^{\theta_{ES,a} \cdot f(\mathcal{E}_a, \mathcal{S}_a)}$$

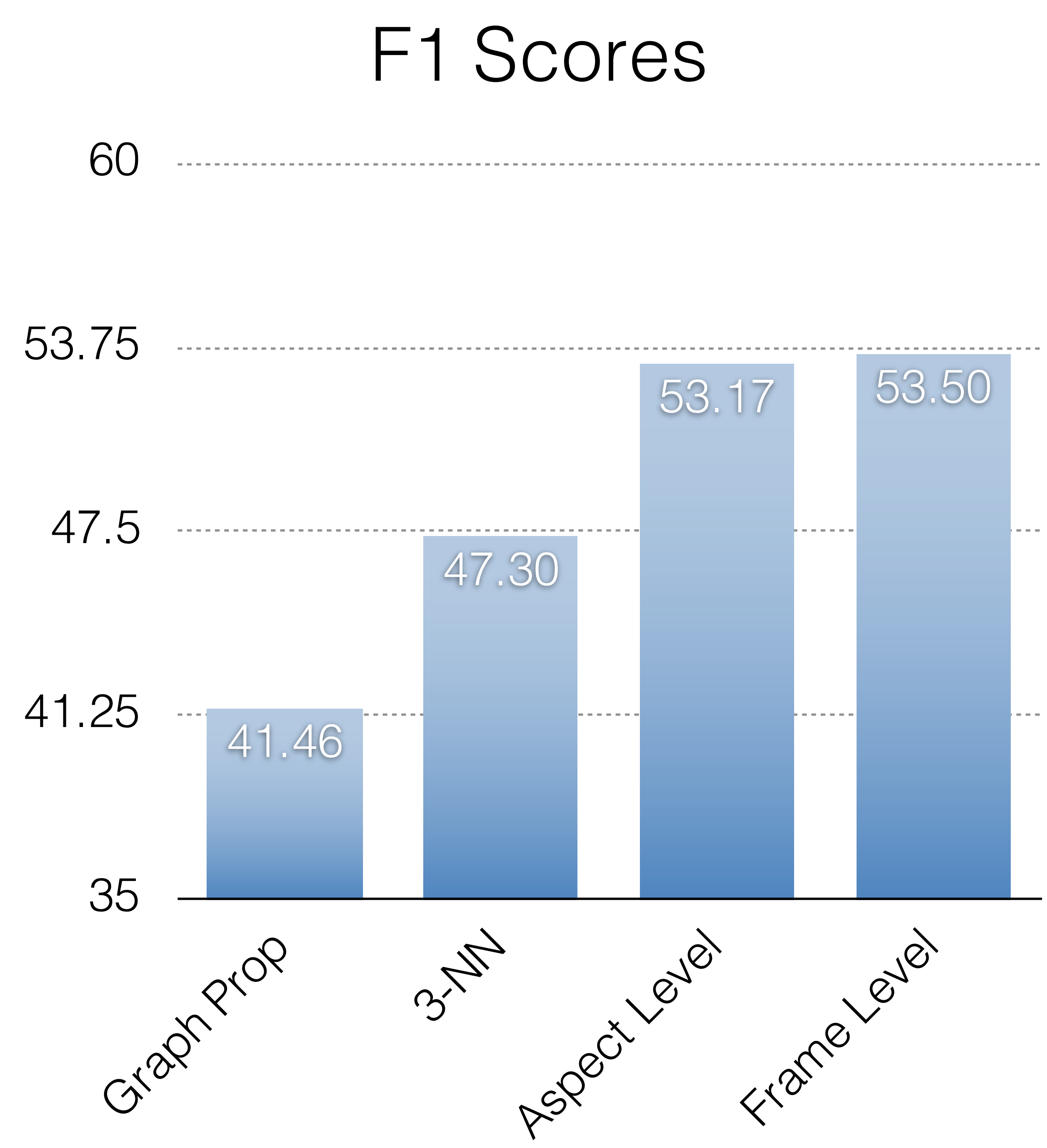
Marginal Probability:

$$\prod_{a \in N(v)} \mu_{a \to v}(x)$$

Experimental Results

train/dev/test sets of 300 predicates each, 3-way classification





Power and Agency Frames in Modern Films

Maarten Sap et al. (EMNLP 2017 in submission)

Bias in Modern Films

- Not a news per say, but more insights from recent studies ((https://www.google.com/intl/en/about/main/gender-equality-films/)
- Bechdel test:
  - At least two women
  - talk to each other
  - about something other than a man.
- Agarwal et al. 2015 report 40% of the films pass the Bechdel test.
- However, those that pass the test still have problems…

A character in *Dykes to Watch Out For* explains the rules that later came to be known as the Bechdel test (1985)

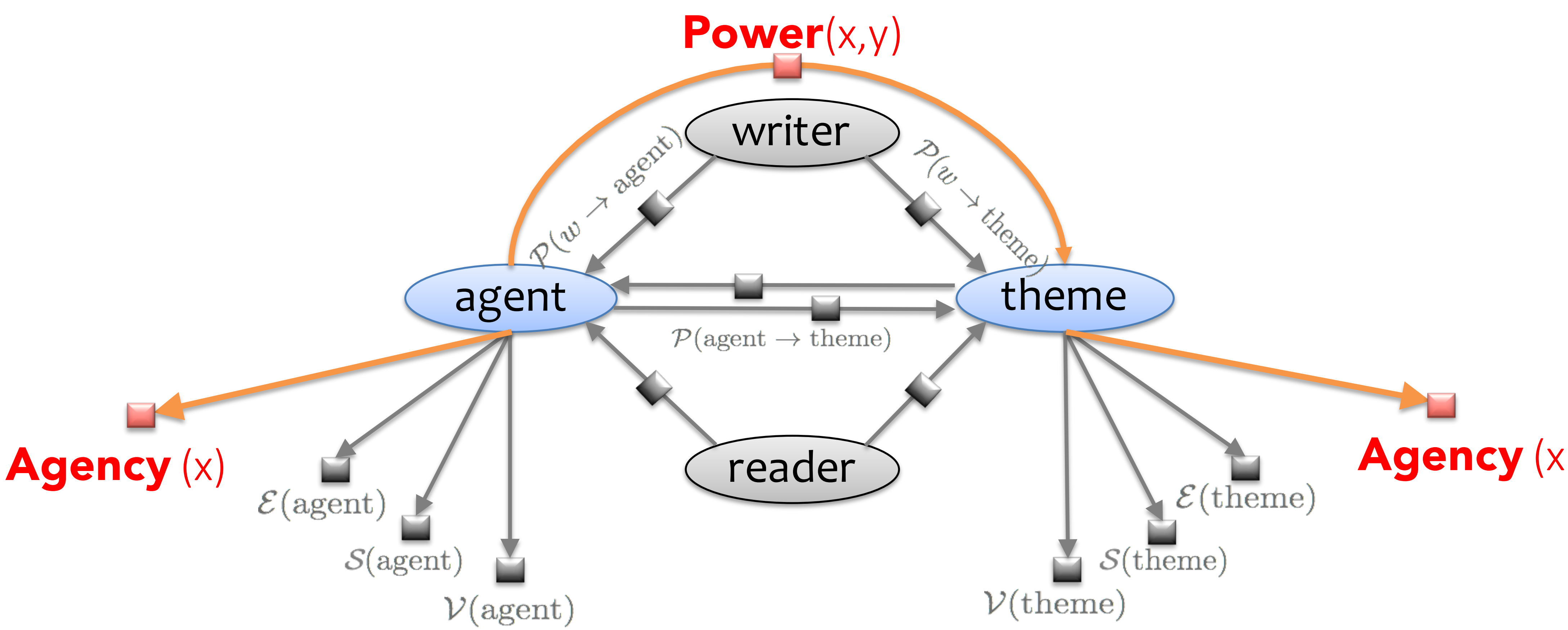
Character Portrayal in Films

"… Elsa objects, unleashing her powers… Elsa discards her crown and creates a palace… Elsa […] freezes Anna … She then summons… Elsa ends the winter"

"… accidentally injures Anna … Hans proposes to Anna … She gets lost … chases Anna … Anna spots Hans … He locks Anna in a room…"

"Cinderella lives an unhappy life… agrees to let Cinderella go… he sees Cinderella… Cinderella hears the clock… allowing the mice to free Cinderella… Cinderella appears…"

Power and Agency Frame

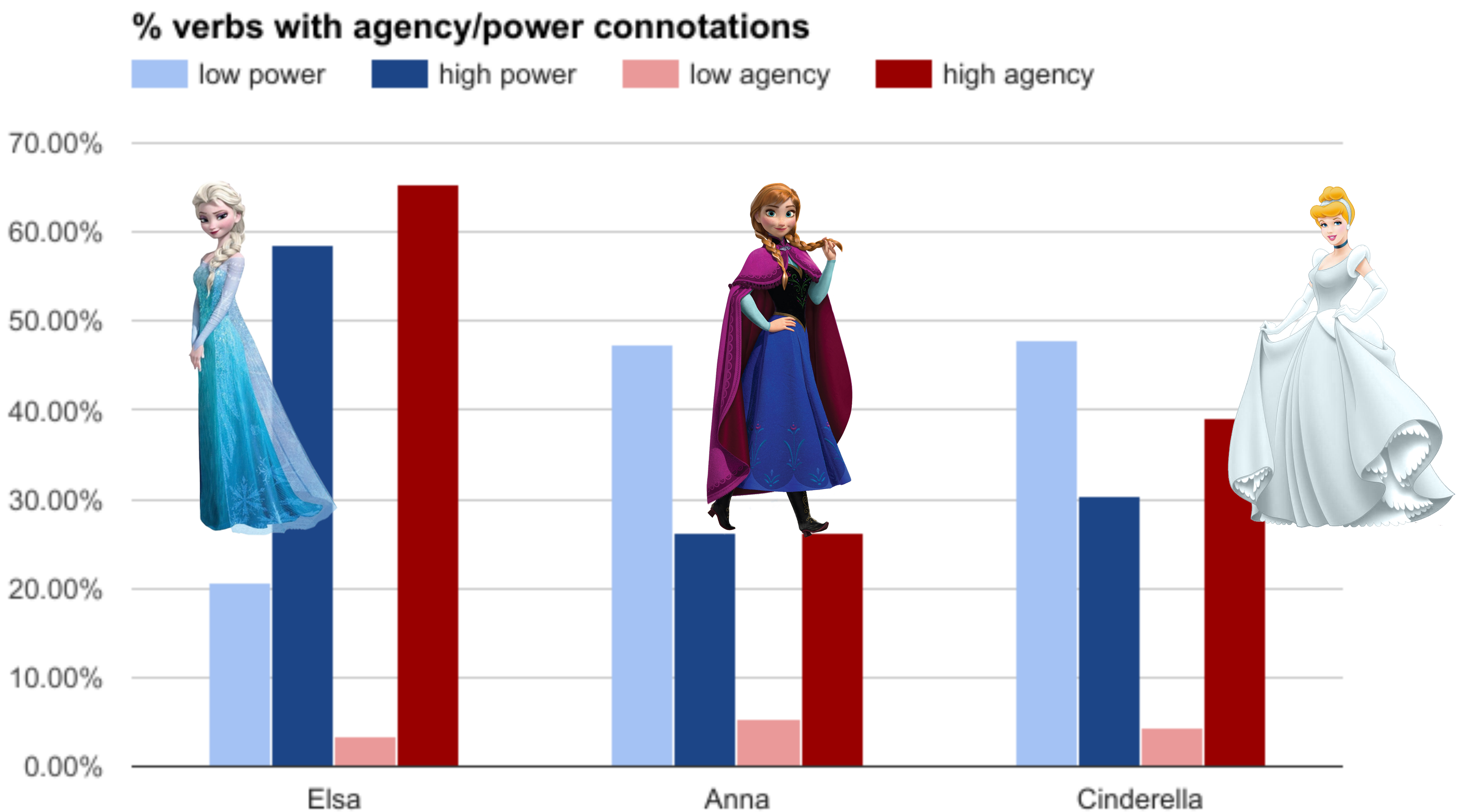


- **Agency (x)**
  - Powerful, decisive, capable of pushing forward their own storyline
  - "waiting" v.s. "searching" for your prince.
- **Power (x,y)**
  - Differential authority level between the verb agent and theme
  - "demanding" v.s. "begging" for mercy

Power and Agency Frame

Collected using Mturk, annotated ~2000 verbs

Frame	Verbs
Agency (+)	terminate, silence, sue, command, invade, throw, organize, authorize, persuade, battle
Agency (-)	shimmer, remain, consist of, border, plummet, age, whimper, inherit, recede, resemble
Power (agent)	squash, seduce, assign, order, suspend, regulate, possess, eject, prohibit, ignite, destroy
Power (theme)	obey, dread, ask, worship, belong to, trust, bow to, serve, implore, require, salute



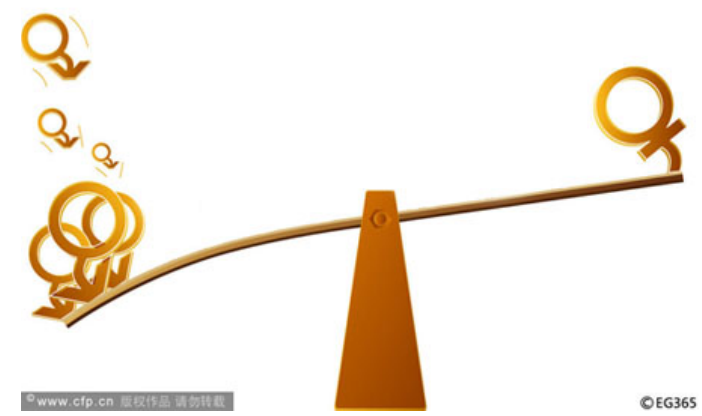
Power and agency connotation frames run on the Wikipedia plot summaries of *Frozen (2013)* and *Cinderella (1950)*

Bias in Modern Films

772 scripts (Gorinski & Lapata, 2015), 21K characters with automatically extracted gender

- Male characters tend to drive the plot more through *high agency* verbs

- Connotation frames provide more nuanced analysis compared to existing tests (Bechdel, Mako Mori, Furiosa, Sexy Lamp test)

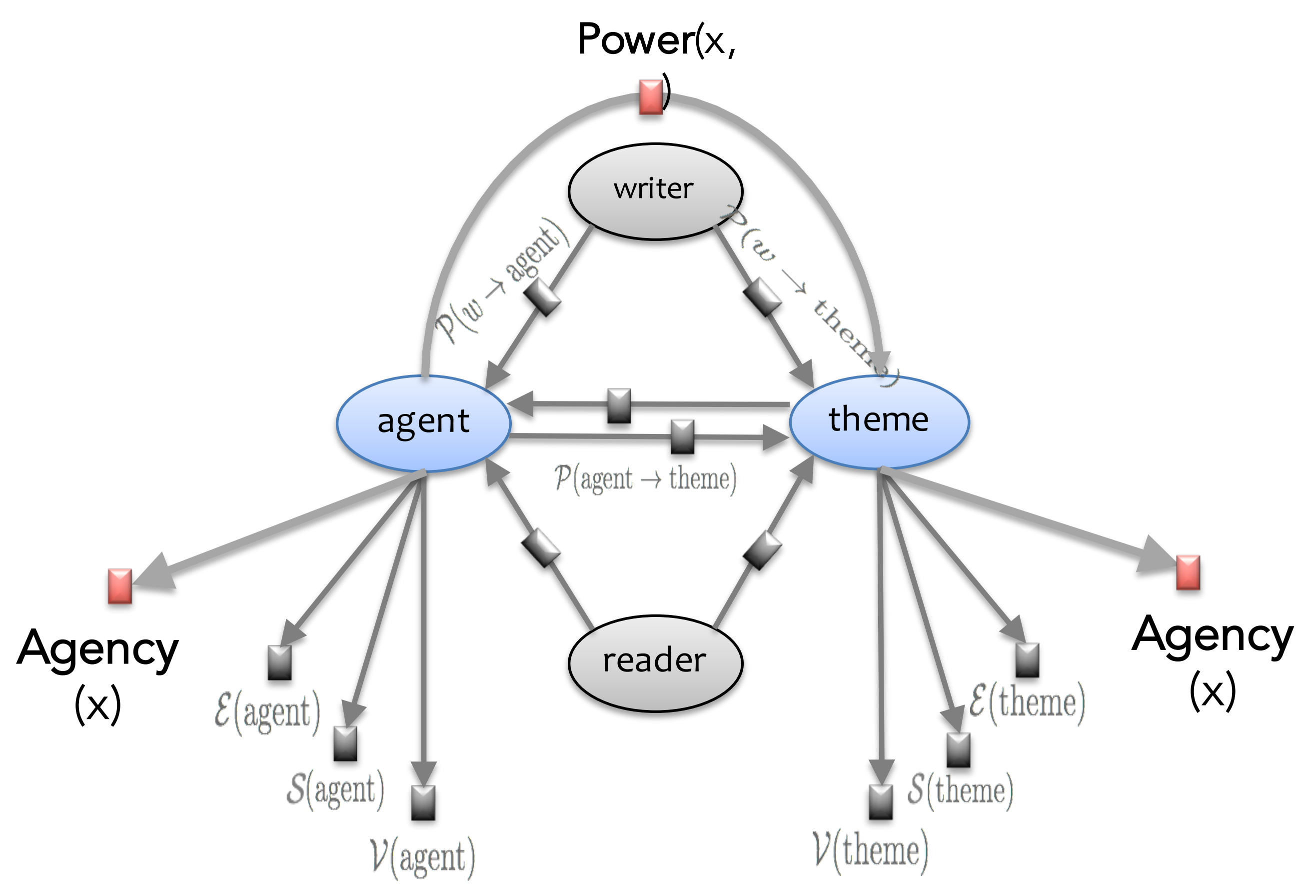


Frame	Gender association
Agency (+)	M**
Agency (-)	F**
Power (agent)	M**
Power (theme)	*n.s.*

Gender association with connotation frames using a logistic regression, controlling for number of words. **: $p<0.001$ (Holm corrected)

To Conclude

- Aspects of connotation can be packed into lexical and frame semantics



NLP for Social Good!

- Identifying unwanted biases in language
  - (Sap et al. @ EMNLP 2017!)
- graded deception detection in media
  - www.politifact.org
  - (Rashkin et al. @ EMNLP 2017!)

Game Plan

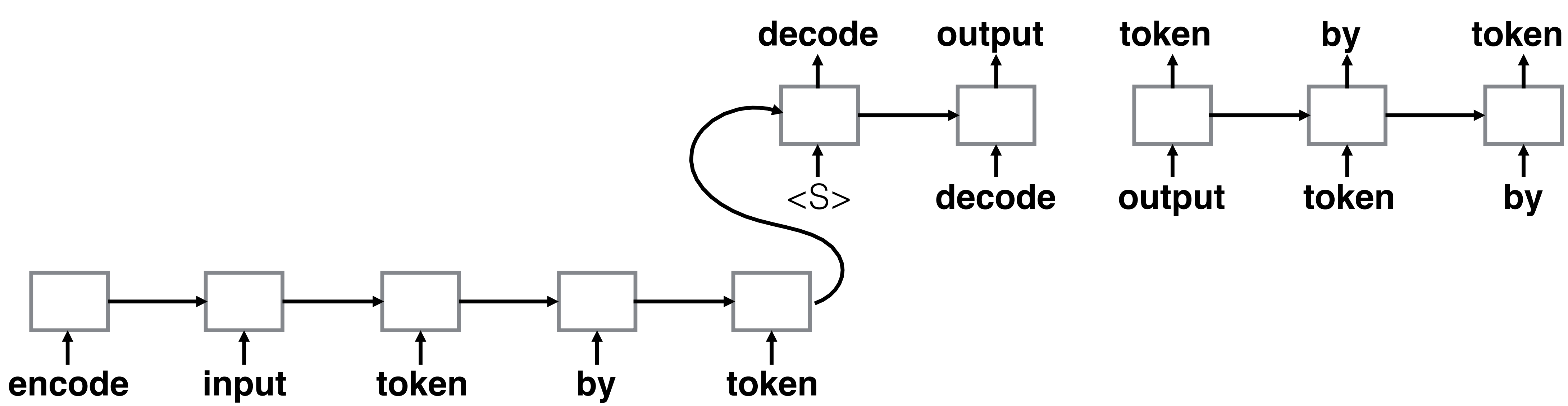
1. Commonsense Frame Semantics
   - Naive physics
   - Social commonsense

Modeling the World, not just Language
   - Dynamic entities
   - Action causalities

~~one vector~~

You can't cram the meaning of a whole %&!$# sentence into a single $&!#* vector!

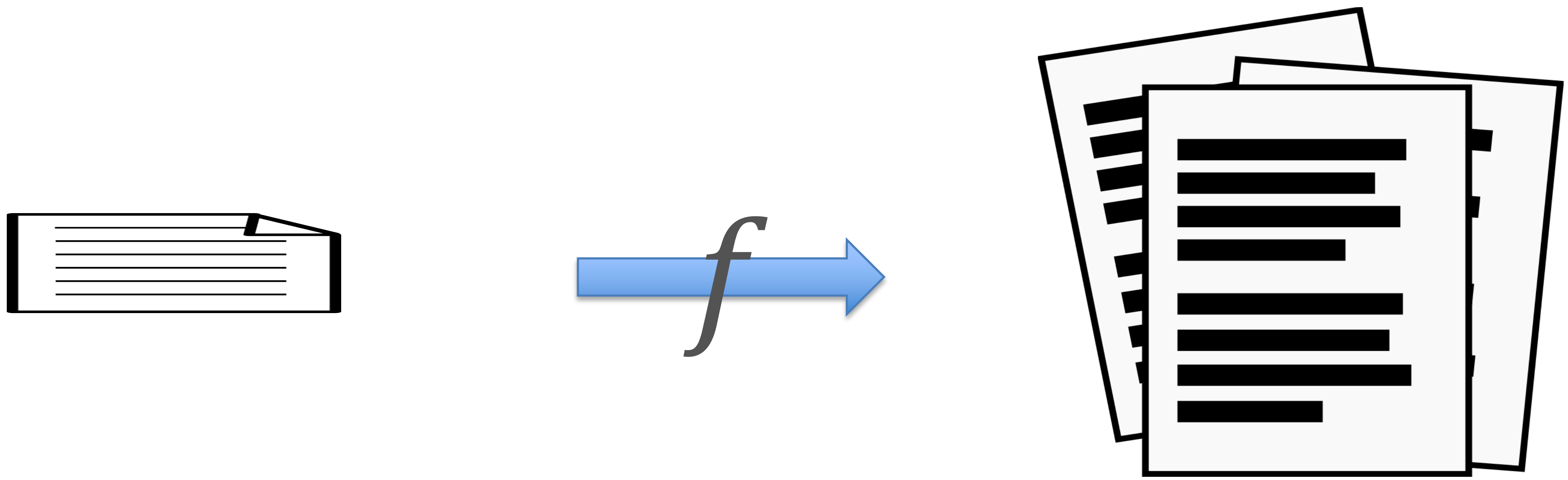
a sequence of vectors?



Neural Checklist Models

Chloe Kiddon et al. (EMNLP 2016)

Task Definition



- Input:
  - Goal: an (optional) title phrase
  - Agenda: a list of items to talk about
- Output: long text with many sentences

Cooking Instructions

Blueberry Muffins

Ingredients

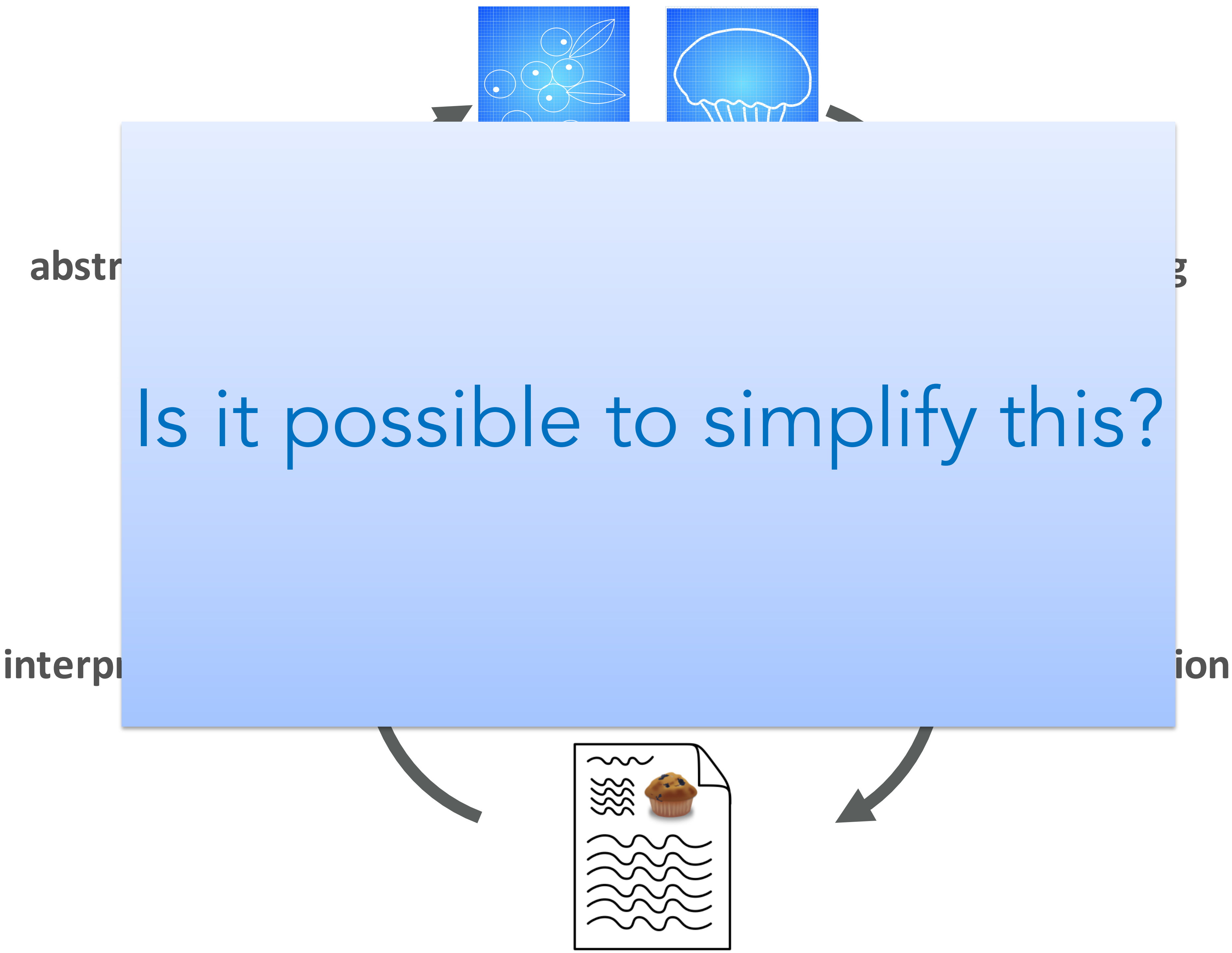
1 cup milk
1 egg
1/3 cup vegetable oil
2 cups all-purpose flour
2 teaspoons baking powder
1/2 cup white sugar
1/2 cup fresh blueberries

Procedure

1. Preheat oven to 400 degrees F. Line a 12-cup muffin tin with paper liners.
2. In a large bowl, stir together milk, egg, and oil. Add flour, baking powder, sugar, and blueberries; gently mix the batter with only a few strokes. Spoon batter into cups.
3. Bake for 20 minutes. Serve hot.

http://allrecipes.com/Recipe/Blueberry-Muffins-I/

How to generate recipes

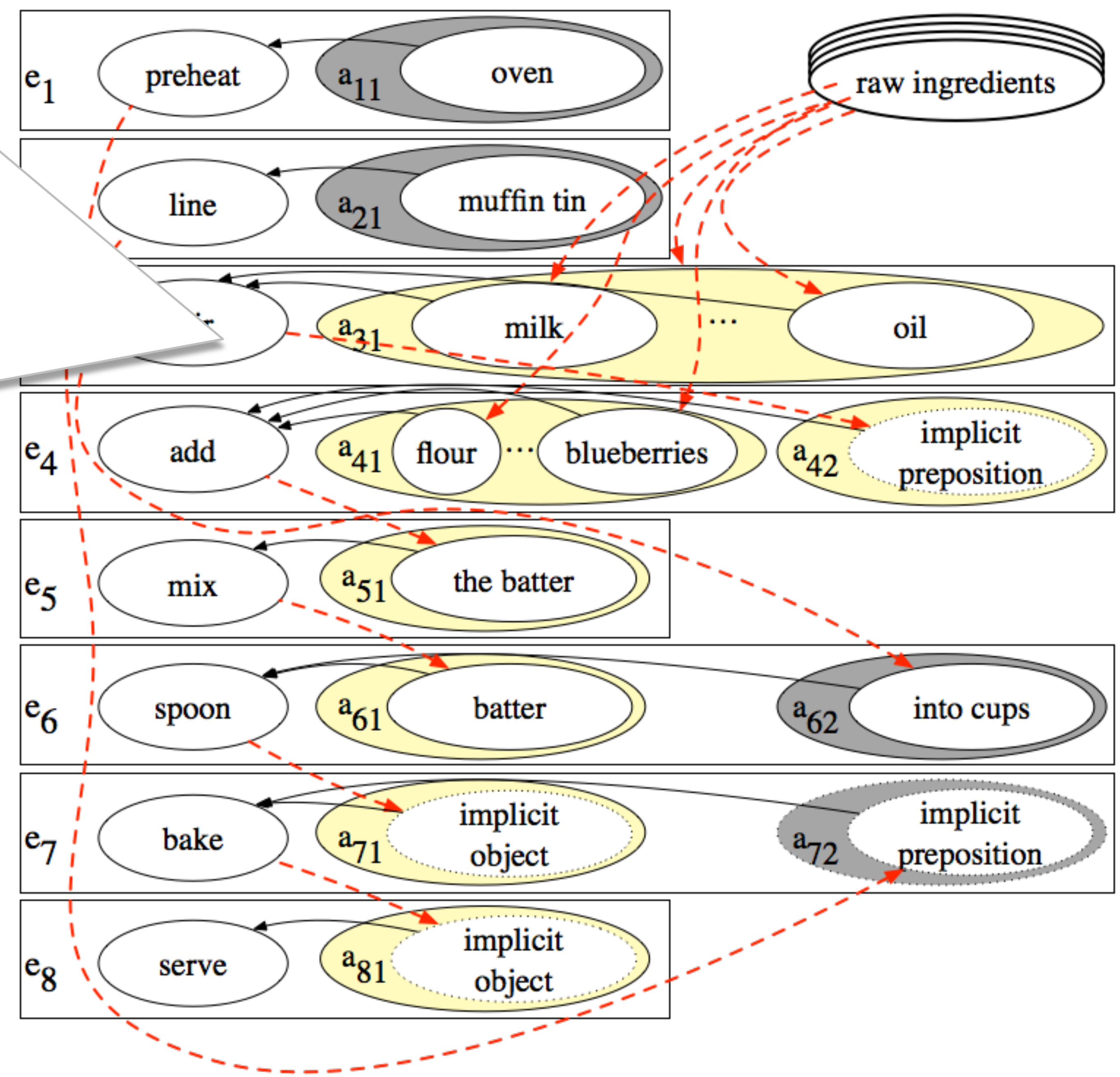


Is it possible to simplify this?

Document-level Graph Parsing

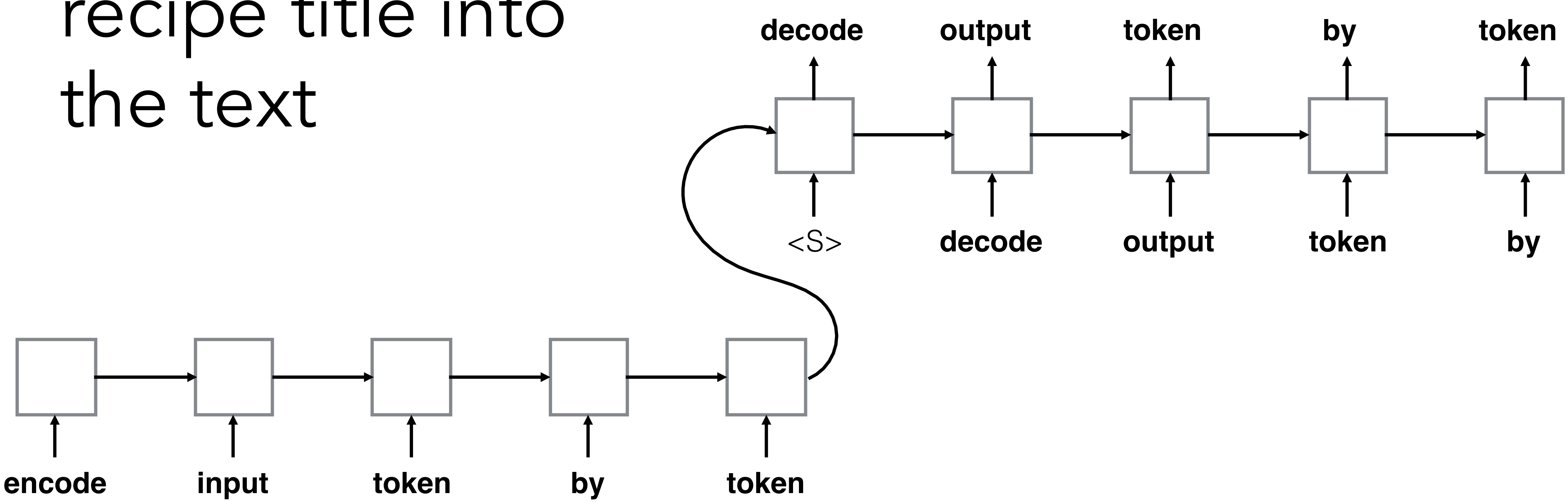
(Kiddon et al., 2015)

1. Predicate argument structure with implicit arguments (~ SRL)
2. How entities are introduced and then flow through different actions (~ reference resolution)
3. Discourse-level parsing



Recipe generation as machine translation?

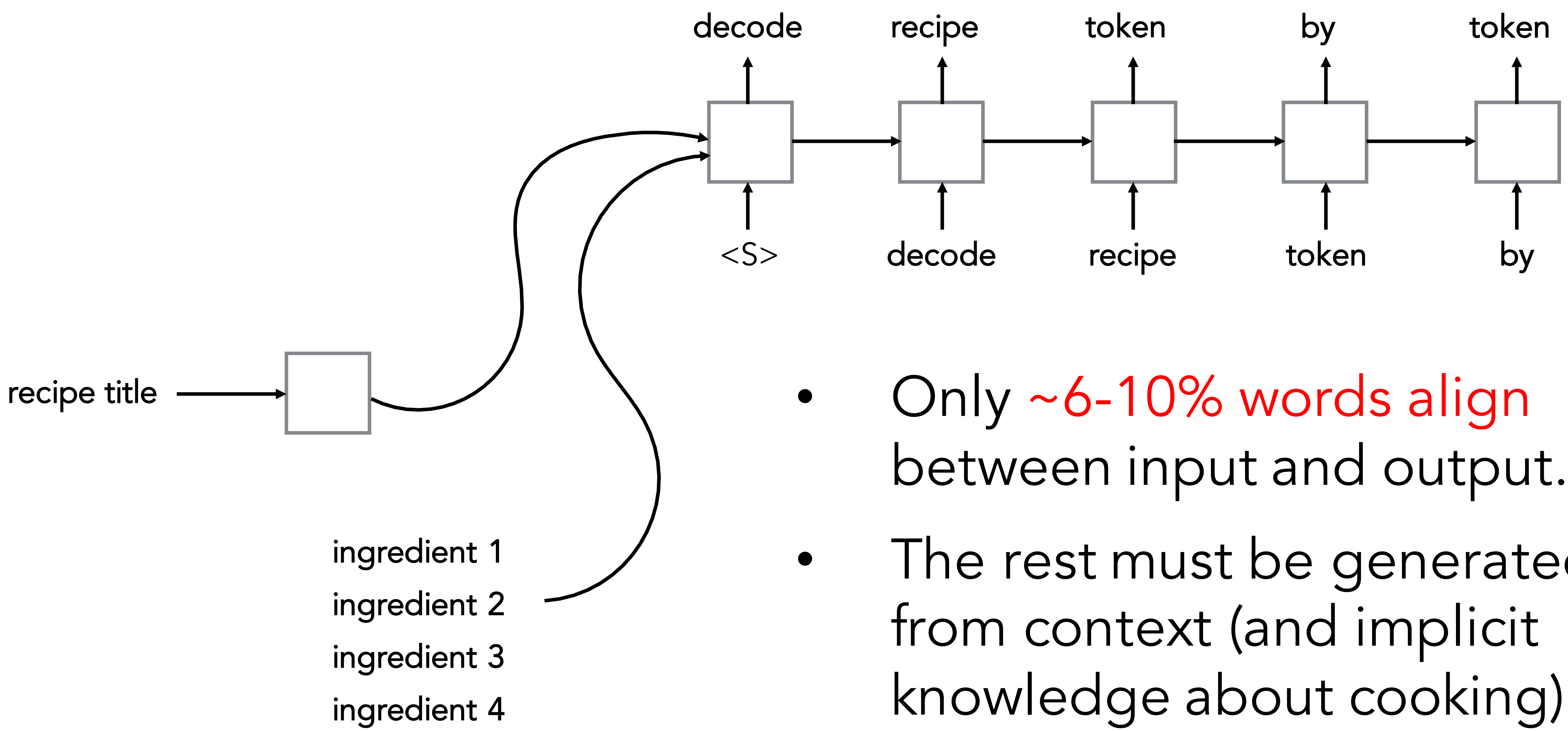
- Encoder-decoder NNs work great for machine translation (Cho et al.2014, Sutskever et al. 2014)
- We can *translate* a recipe title into the text



Encode title - decode recipe

sausage sandwiches → Cut each sandwich in halves.
Sandwiches with sandwiches.
Sandwiches, sandwiches, Sandwiches, sandwiches, sandwiches
sandwiches, sandwiches, sandwiches, sandwiches, sandwiches, sandwiches, or sandwiches or triangles, a griddle, each sandwich.
Top each with a slice of cheese, tomato, and cheese.
Top with remaining cheese mixture.
Top with remaining cheese.
Broil until tops are bubbly and cheese is melted, about 5 minutes.

Recipe generation **vs** machine translation



- Only ~6-10% words align between input and output.
- The rest must be generated from context (and implicit knowledge about cooking)
- Contextual switch between two different input sources

Two input sources

Challenges

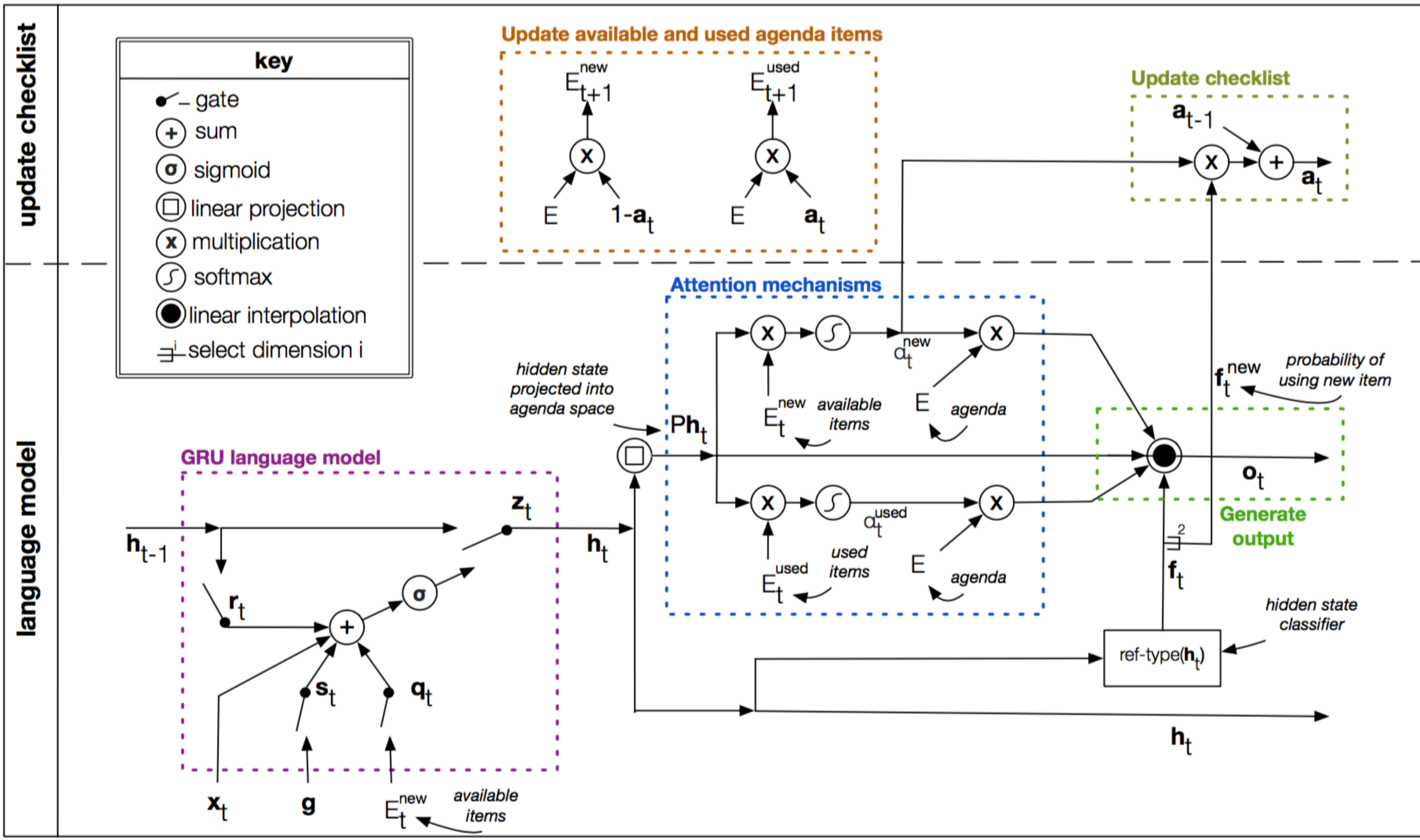
1. Modeling global coherence and coverage
   – May not be possible to compress everything on to one hidden vector!

2. Need a mechanism to track an “agenda”

   -- which ingredients have been already used

   -- what need to be yet introduced

Neural checklist model

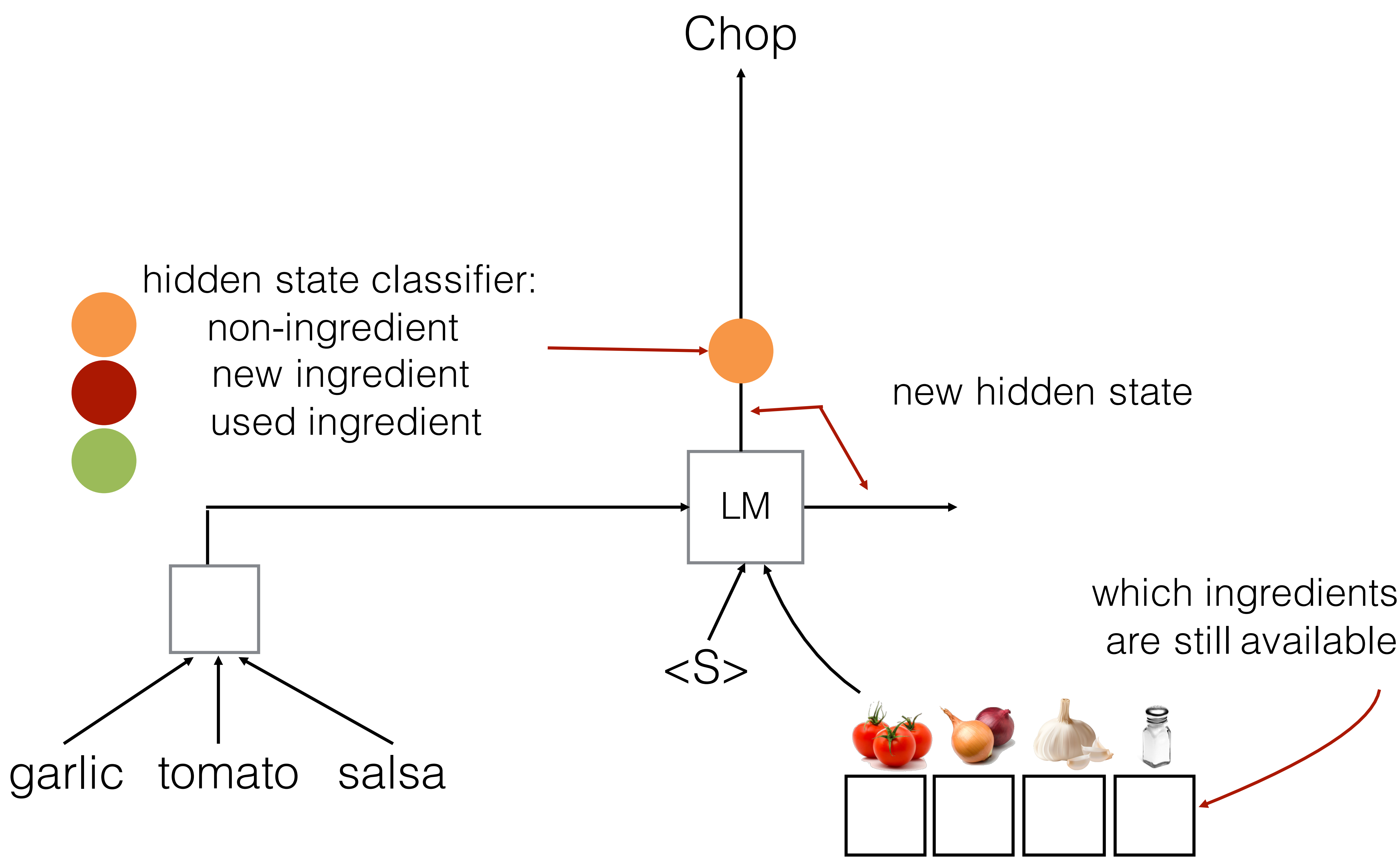


Let’s make salsa!

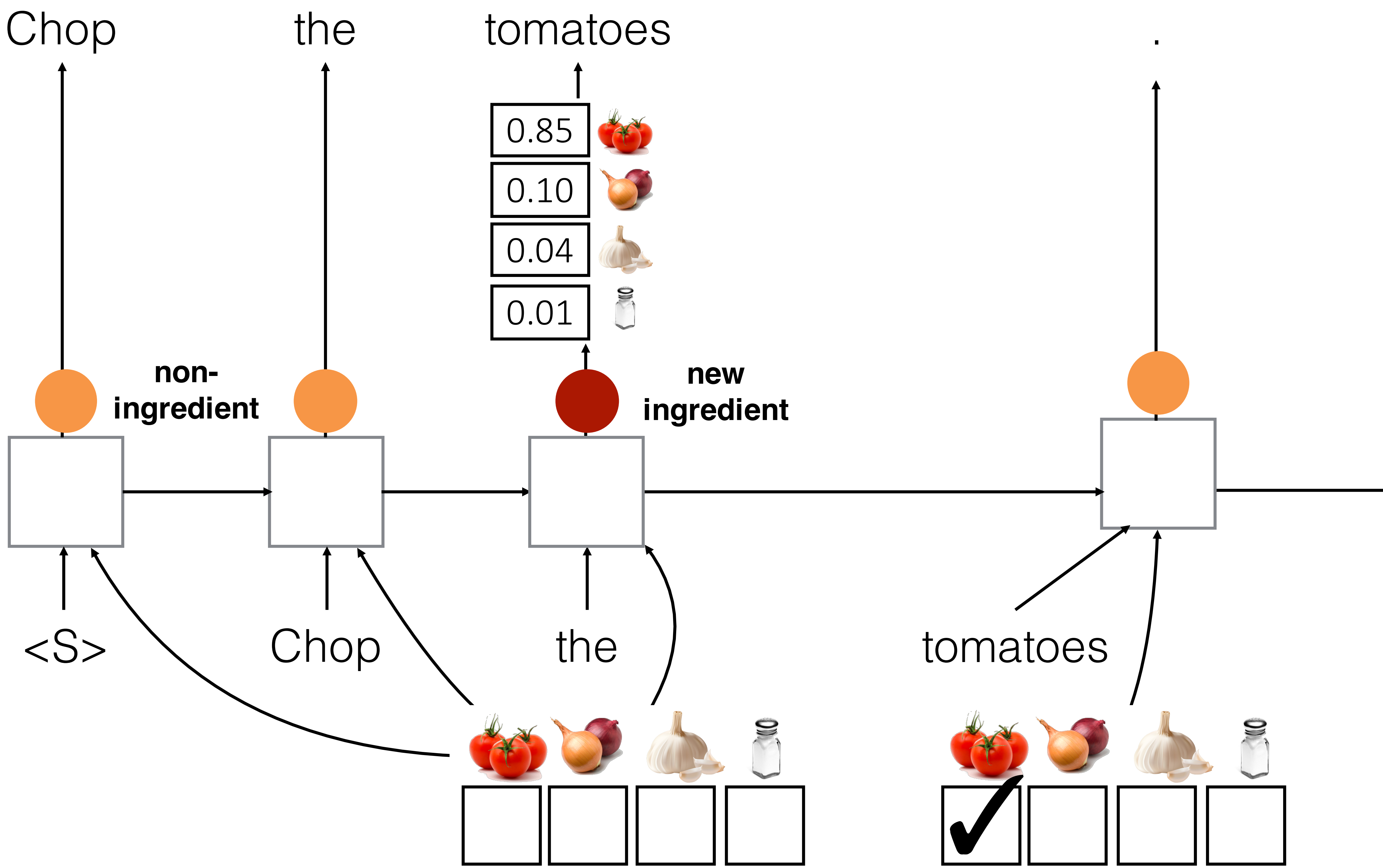
Garlic tomato salsa

tomatoes
onions
garlic
salt

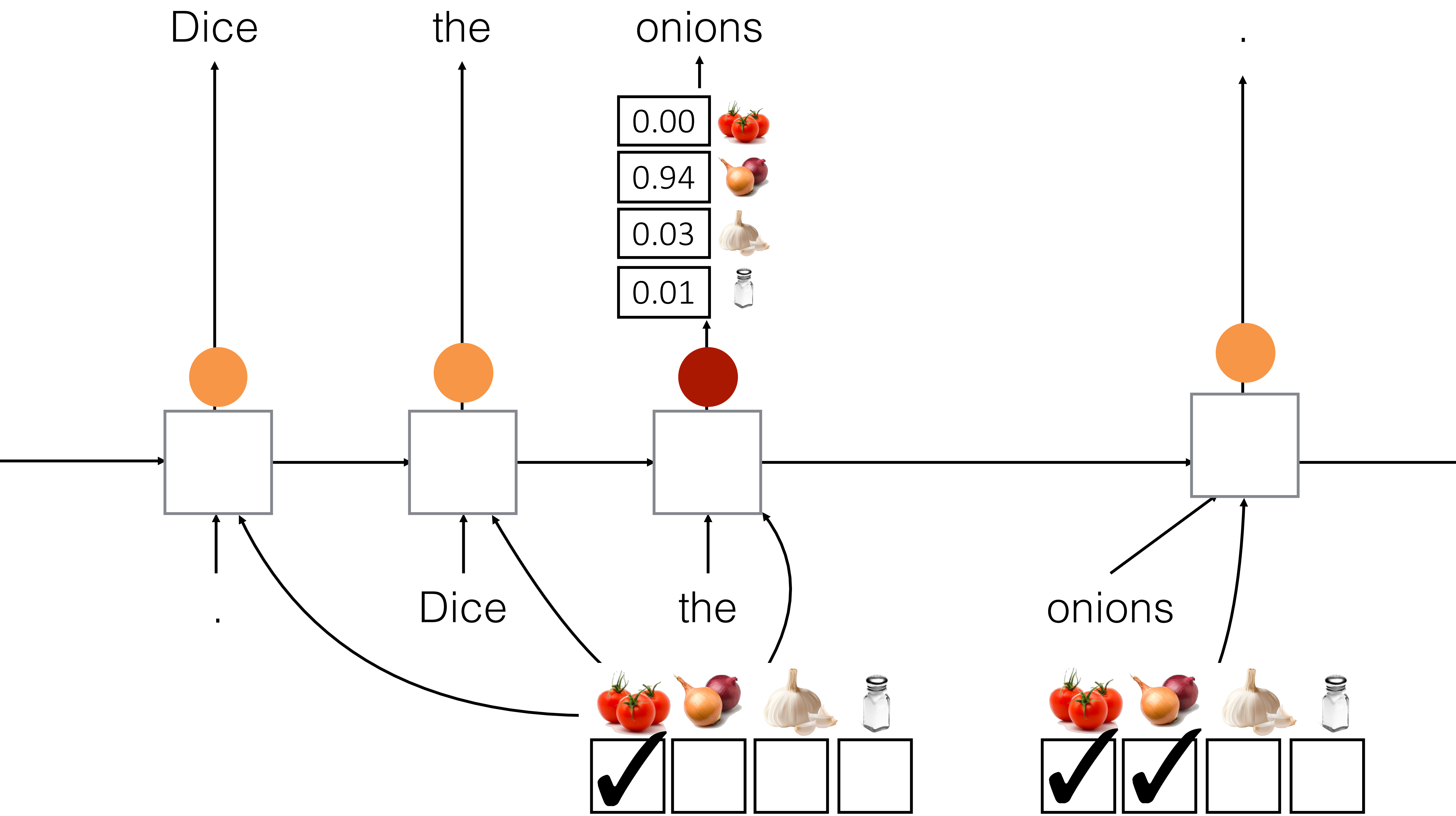
Neural checklist model



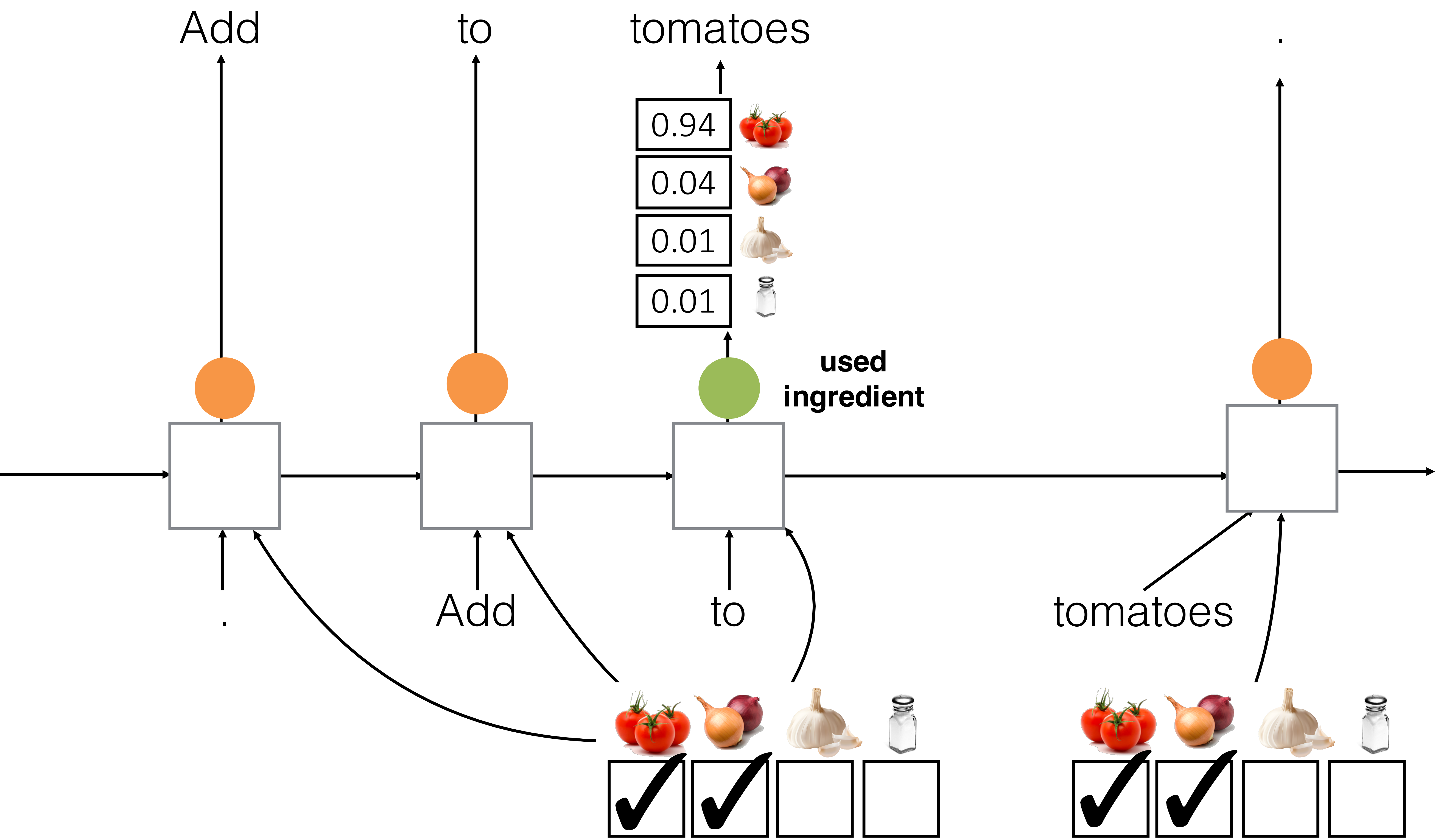
Neural checklist model



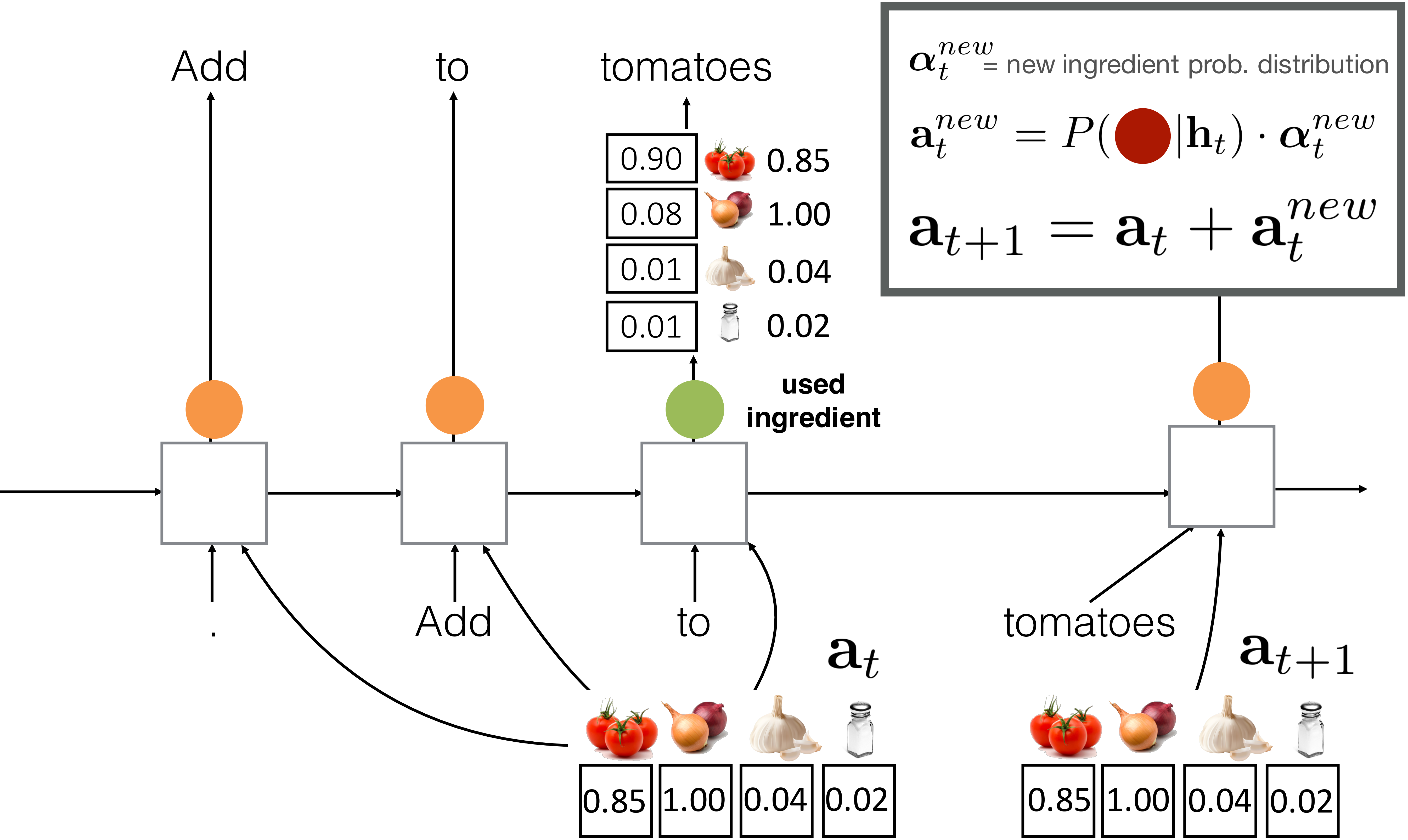
Neural checklist model



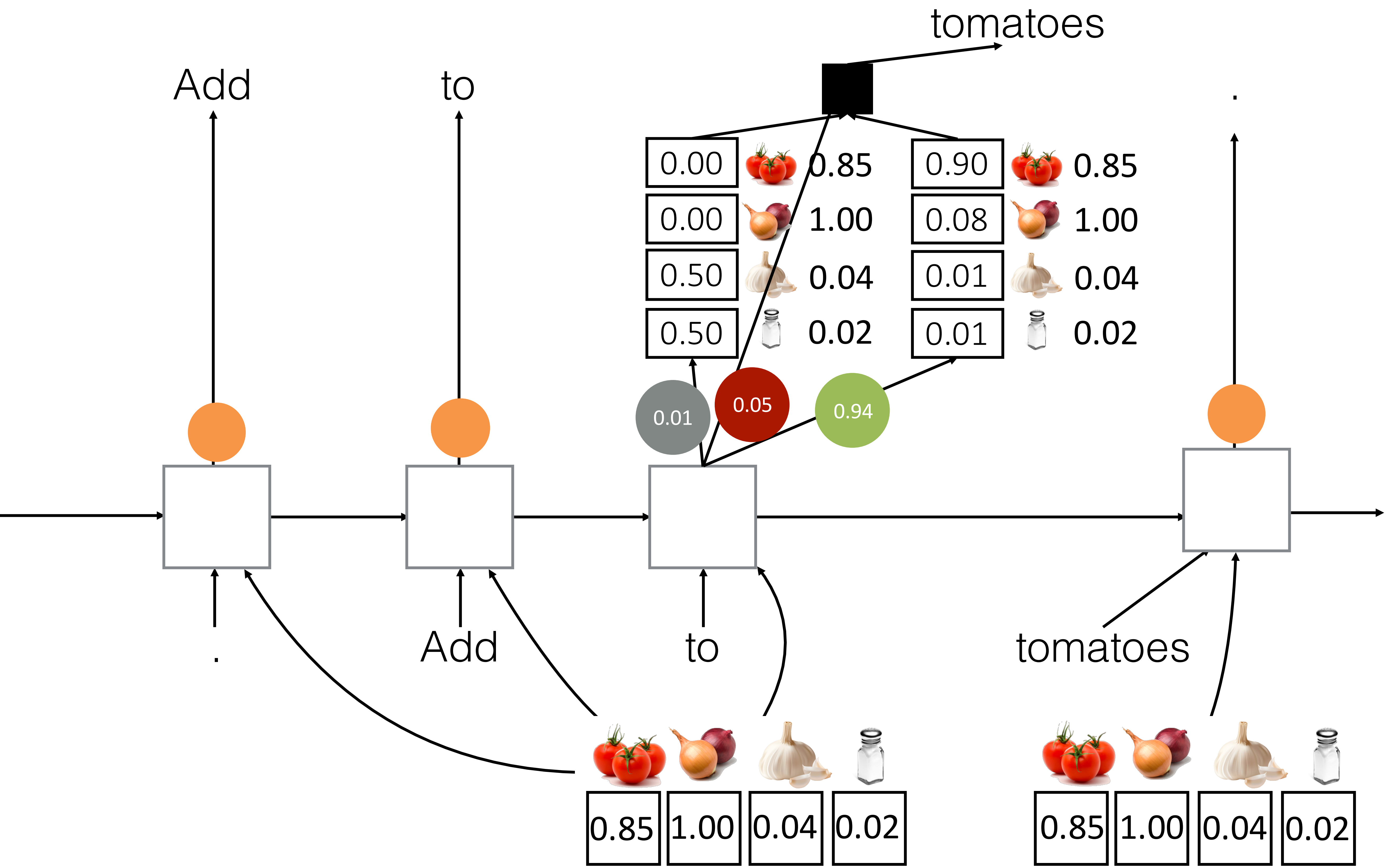
Neural checklist model



Checklist is probabilistic



Hidden state classifier is soft



Neural Recipe Example #1

title: oven eggplant

In a small bowl , combine the cheese , eggplant , basil , oregano , tomato sauce and onion . Mix well .
Shape mixture into 6 patties , each about 3/4-inch thick.
Place on baking sheet .
Bake at 350 degrees for 30 minutes or until lightly browned .
Southern living magazine , sometime in 1980 .
Typed for you by nancy coleman .

eggplant
cheese cottage lowfat
chopped onion
bay ground leaf
basil
oregano
tomato sauce
provolone

Cook eggplant in boiling water , covered , for 10 min .
Drain and cut in half lengthwise . scoop out insides leaving 1/2 '' shell . Mash insides with cottage cheese , onion , bay leaf , basil , oregano and tomato sauce .
Preheat oven to 350 ^ stuff eggplant halves , place in casserole dish and bake covered for 15 min .
Add a little water to bottom of pan to keep eggplant moist . top with provolone cheese .
Bake 5 more min uncovered 1 serving =

Example: skillet chicken rice

Ingredients

2 lb chicken pieces skinned
3 cups mushrooms fresh sliced
4 carrots peeled sliced 1/2"
3/4 cups long grain rice
1/2 c onion chopped
1 tsp poultry seasoning
1 tsp bouillon chicken granules
1/4 tsp salt

True recipe

Spray a 12-inch skillet with nonstick spray coating.
Brown chicken pieces on all sides over medium heat about 15 minutes.
Remove chicken.
Drain fat from skillet, if neccessary.
Add mushrooms, carrots, rice, onion, bouillon, poultry seasoning, 2 cups water, salt.
Place chicken atop rice mixture.
Cover; simmer 30 minutes or till chicken and rice are done.

Baseline

In a large skillet, brown **chicken** in oil.
Add **chicken** and **rice**; cook, stirring, until chicken is no longer pink.
Add rice and **broth**; bring to a boil.
Reduce heat to low; cover and cook 5 minutes or until rice is tender and liquid is absorbed. **Stir in rice**.
Heat to boiling; reduce heat.
Cover and simmer until rice is tender, about 20 minutes.
Stir in rice. Heat to boiling; reduce heat.
Cover and simmer 20 minutes or until rice is tender and liquid is absorbed and rice is tender.

Checklist

In large skillet, heat rice and onion. Add carrots and mushrooms; cook, stirring constantly, until hot and bubbly. Stir in seasonings. Cook and stir over medium heat until hot and bubbly. Stir in chicken and salt. Cover and cook 5 minutes or until chicken is no longer pink in center. Serve over **rice**.

Example: chocolate covered potato chips

True recipe

Ingredients

8 oz sweet milk chocolate

24 thick sliced ridged potato chips

Melt 6 ounces of chocolate. Remove from heat. While chocolate is still hot, mix the 2 ounces of chocolate into the melted chocolate, in 3 additions, stirring until each addition is completely melted before adding the next. **Dip the potato chips**, 1 at a time, in the chocolate. Coat completely and lift with a small fork. Shake off excess chocolate by rapping the fork on the edge of the bowl lightly. Remove any drips from the bottom by running the fork across the edge of the bowl. Slide the chips onto a cookie sheet lined with parchment or wax paper. **Allow to cool** until solid. Let chips sit at room temperature or in the refrigerator.

Baseline

Checklist

Preheat oven to 350 degrees. Grease and flour a 9 x 13 inch pan. Bake at 350 degrees for 30 minutes.

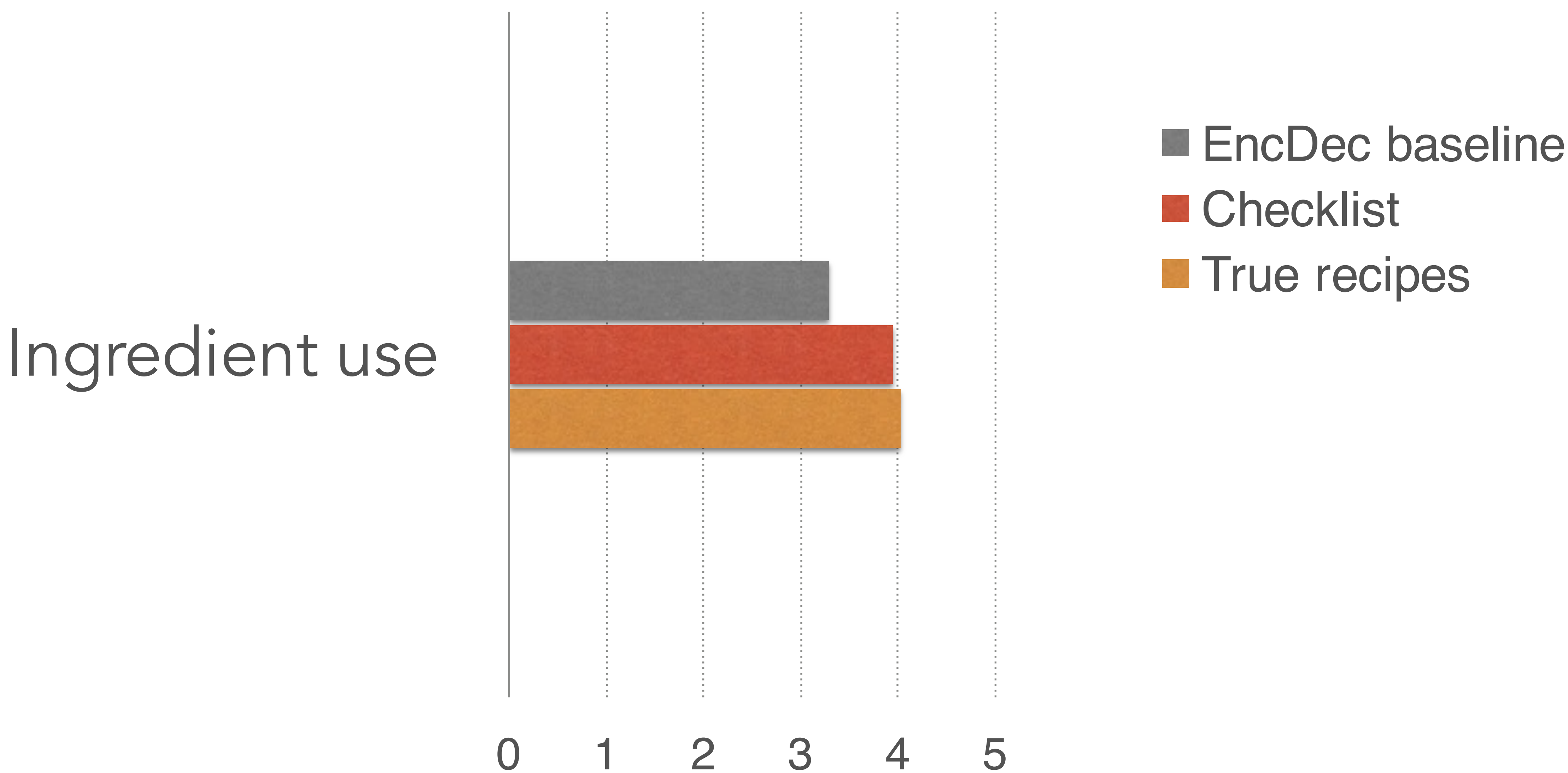
Bake in a 400 degree f oven for 15 minutes or until lightly browned. Drain on paper towels. **Melt chocolate** in top of double boiler over simmering water. Remove from heat and cool to room temp. **Add potato mixture to the potato mixture**; mix well. Cover tightly with plastic wrap. Refrigerate until firm enough to handle. Shape into croquettes. **Fry in hot oil** until golden brown. Drain on paper towels. Serve hot.

Baselines

- Encoder-Decoder baseline (EncDec)
  - Encode title and ingredients as sequence, decode recipe
- Attention model baseline
  - Use attention embedding to ingredient list during output computation at each step

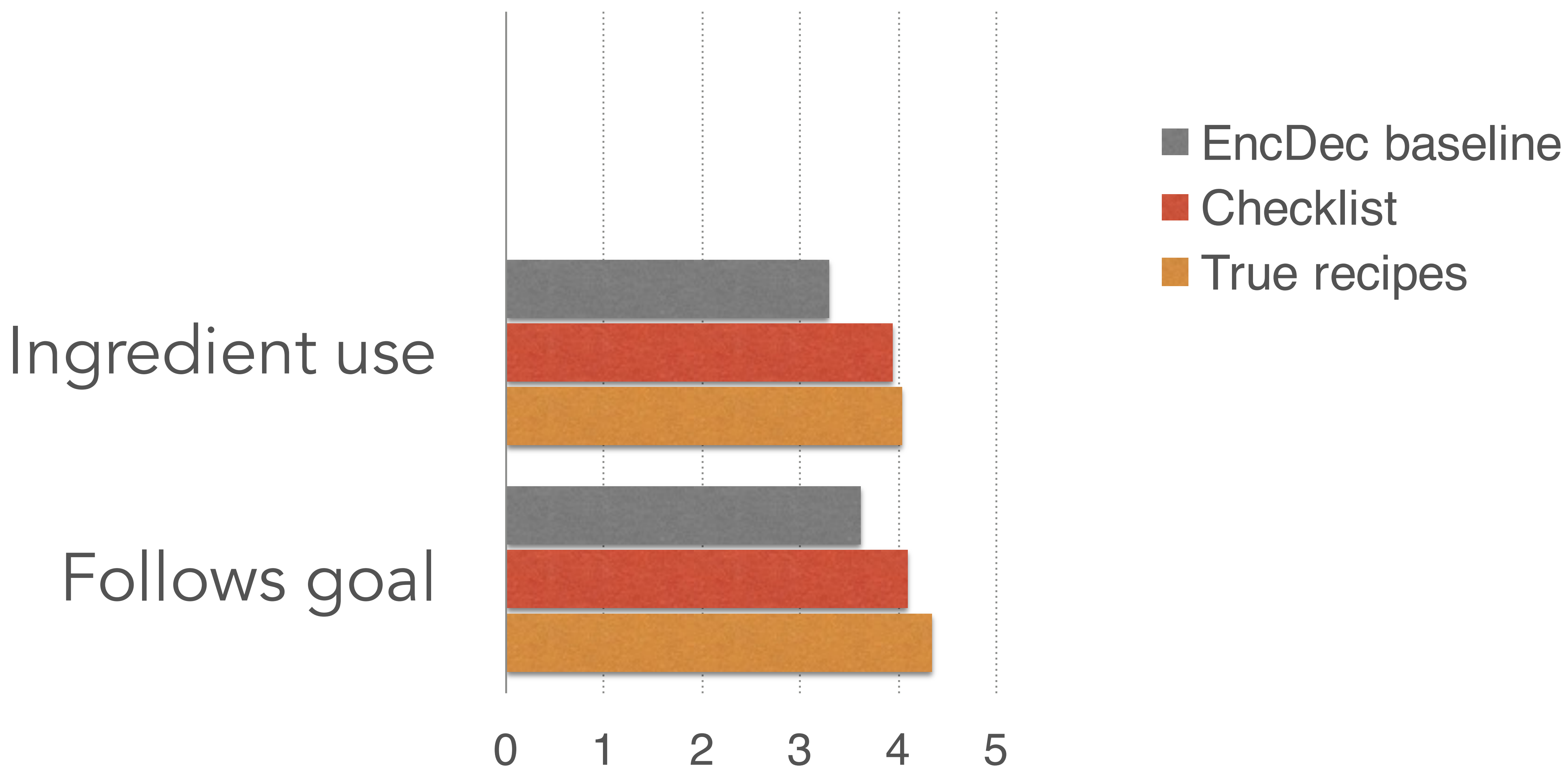
Human evaluation

- Amazon Mechanical Turk



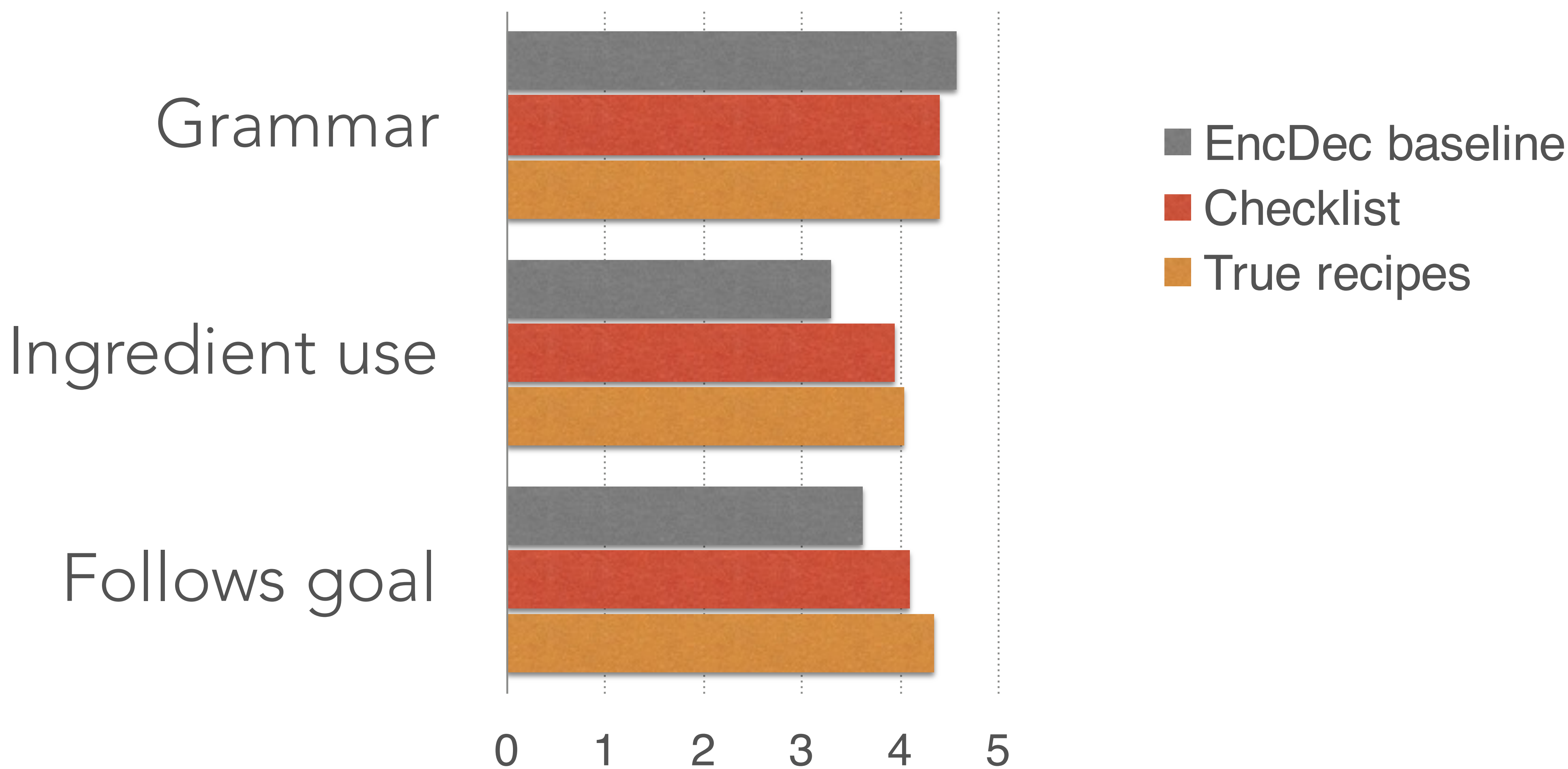
Human evaluation

- Amazon Mechanical Turk

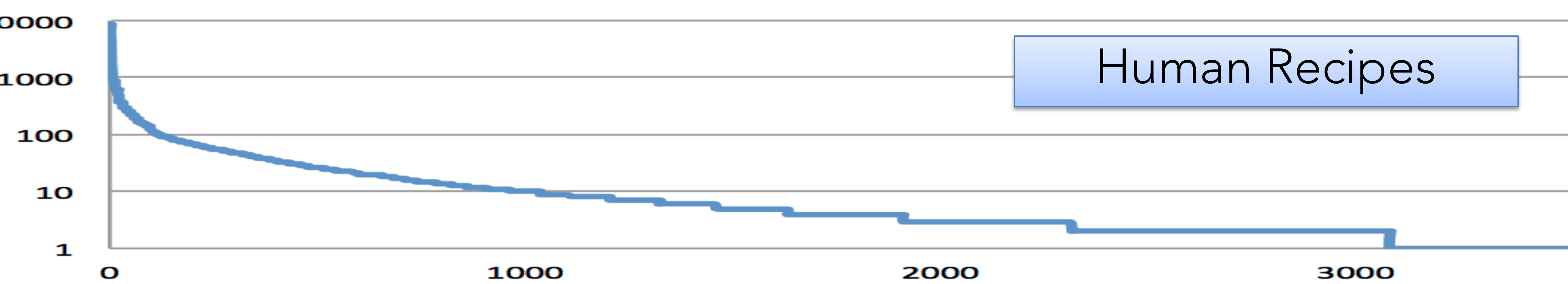


Human evaluation

- Amazon Mechanical Turk

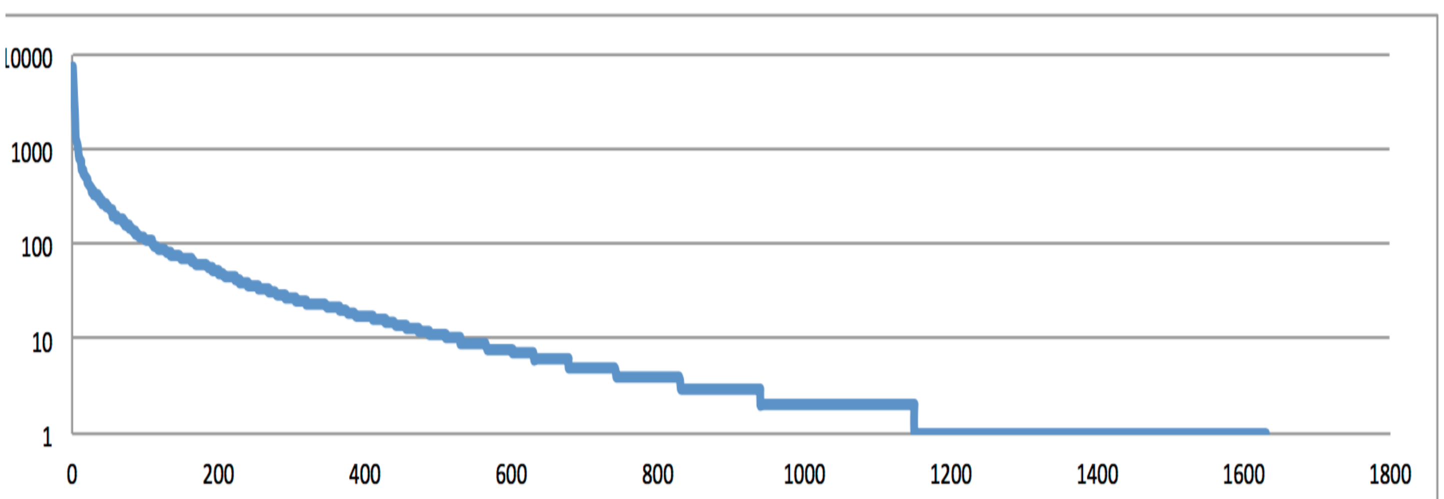


Natural vs. Neural Language (on dev set)



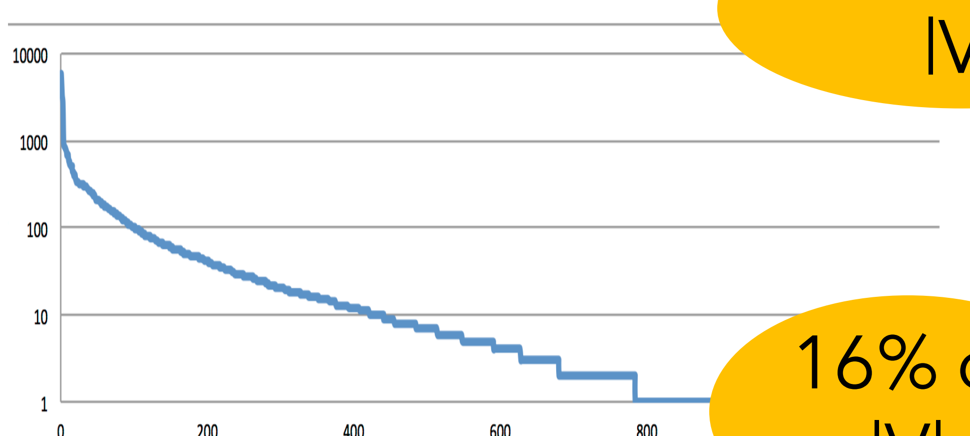
Human Recipes

|V| = 5206



Checklist NN

33% of |V|



16% of |V|

Enc-Dec RNN

Neural checklist model for dialogue generation

- Dialogue system output generation
  - Hotel & restaurant informational systems
  - Wen et al. 2015
  - inform{name(Hotel Stratford),has_internet(no)}
    - "Hotel Stratford does not have internet."

Examples

inform(name=piperade;good_for_meal=dinner;food=basque)

piperade is a basque restaurant that is good for dinner.

inform(name='red door cafe';good_for_meal=breakfast;area='cathedral hill';kids_allowed=no)

the red door cafe is good for breakfast, does not allow children and is in the cathedral hill area.

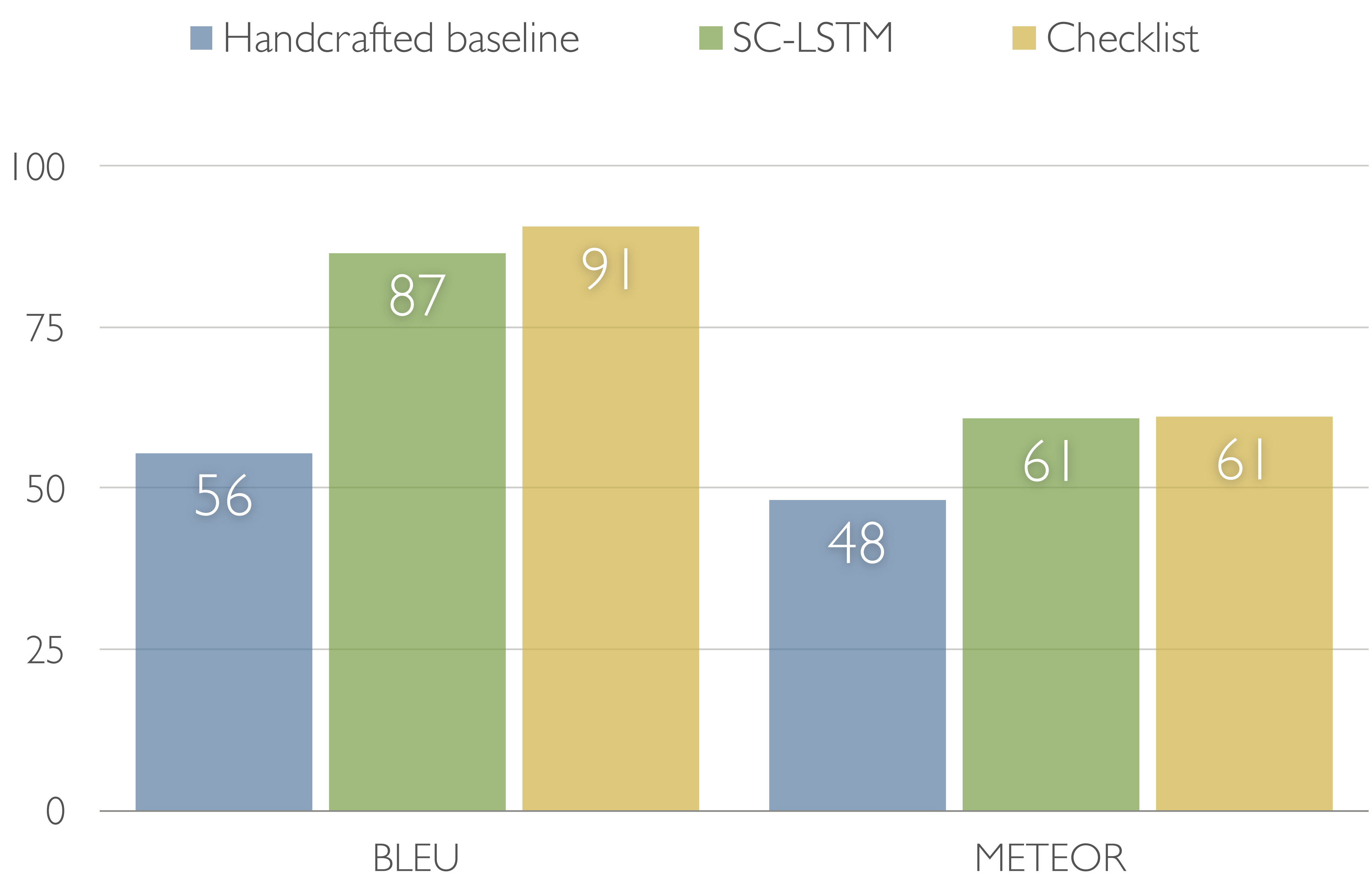
inform(name='manna';area='hayes valley or inset';address='845 irving street')

manna is in the hayes valley or inset area. the address is 845 irving street.

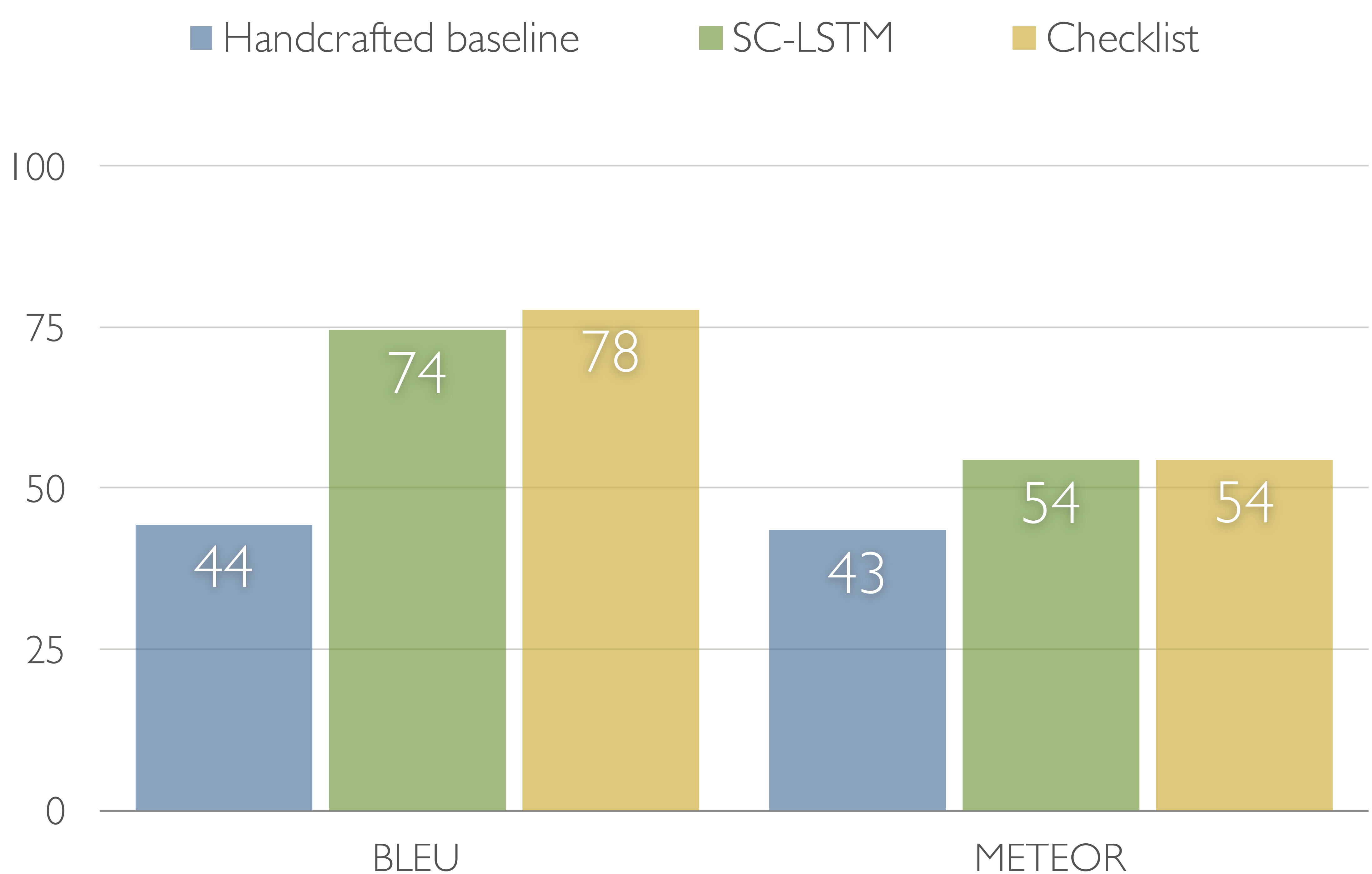
?select(dogs_allowed='yes';dogs_allowed='no')

would you like a hotel that allows dogs?

Hotel domain



Restaurant results

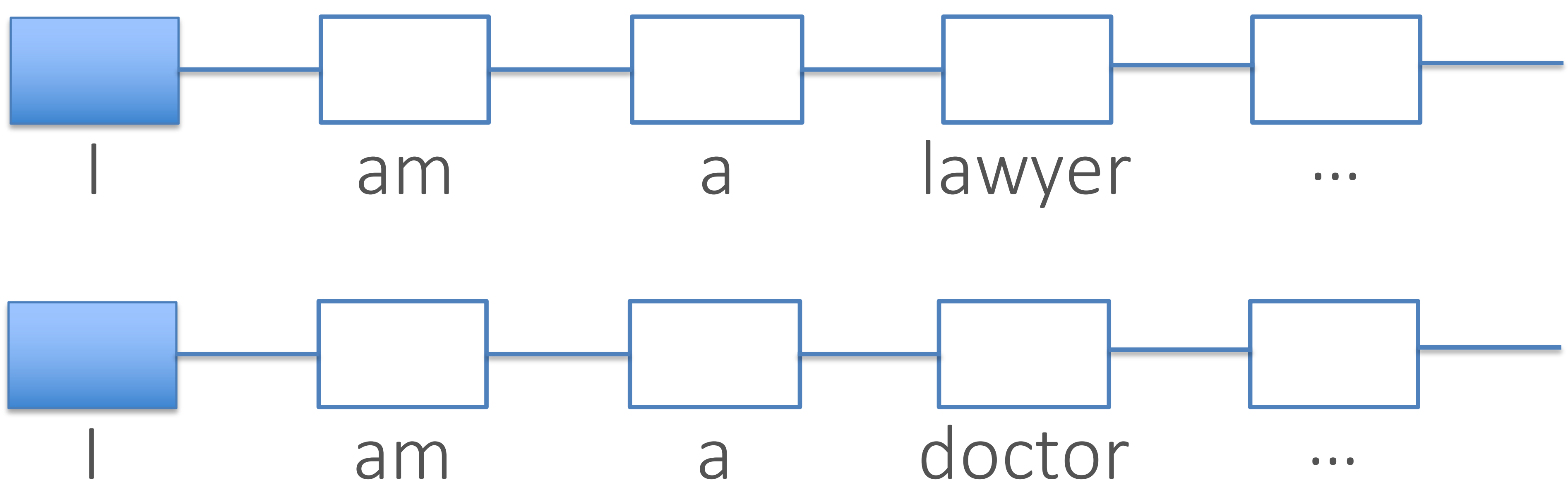


Dynamic Entity Representations in Neural Language Models

Yangfeng Ji et al. @ EMNLP 2017

Static vs Dynamic Entity Representations

- Human: what is your job?
- Machine: i'm a lawyer
- Human: what do you do?
- Machine: i'm a doctor



Example from Paperno et al. 2016

Neural Process Network

Antoine Bosselut et al. (In Preparation)

What’s missing in the end-to-end…

- Title: deep-fried cauliflower
- Ingredients: cauliflower, frying oil, sauce, salt, pepper.

➔

Wash and dry the cauliflower.

Heat the oil in a skillet and fry the sauce until they are golden brown.

Drain on paper towels.

Add the sauce to the sauce and mix well.

Serve hot or cold.

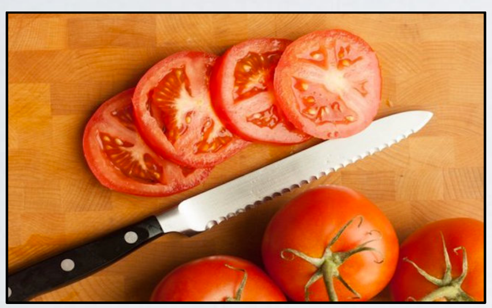
Action Causality

Read:
Slice the tomatoes

Choose Action:
slice

Choose Entity:
tomato

State Changes:
shape → separated



Read:
Bake in pan for 20 minutes.

Choose Action:
bake

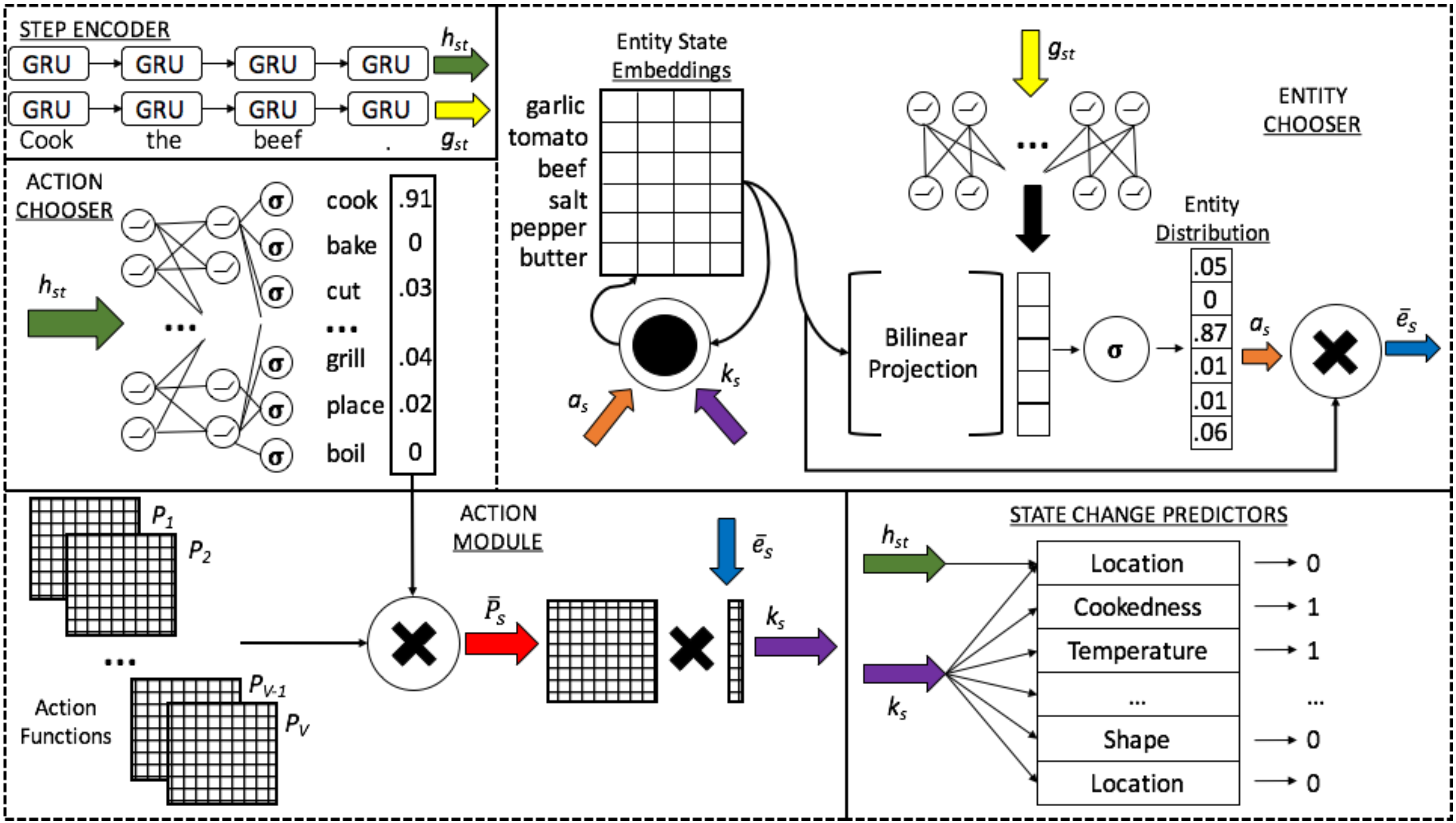
Choose Entity:
<IMPLICIT>

State Changes:
cookedness → cooked
temperature → hot
location → oven

Read:
<END_RECIPE>

Neural Process Networks

"understanding by simulation"



Concluding Remarks

Toward Intelligent Communication

1. Commonsense Frame Semantics
   - naive physics
   - Social commonsense
   - Packing knowledge into lexical/frame semantics

2. Modeling the World, not just Language
   - Dynamic entities
   - Action causalities

Fillmore Tribute Workshop @ ACL 2014

Date: Friday, June 27, 2014

Location: ACL, Baltimore Marriott Waterfront, Baltimore, MD

Frame Semantics in Natural Language Processing

A Workshop in Honor of Chuck Fillmore (1929–2014)

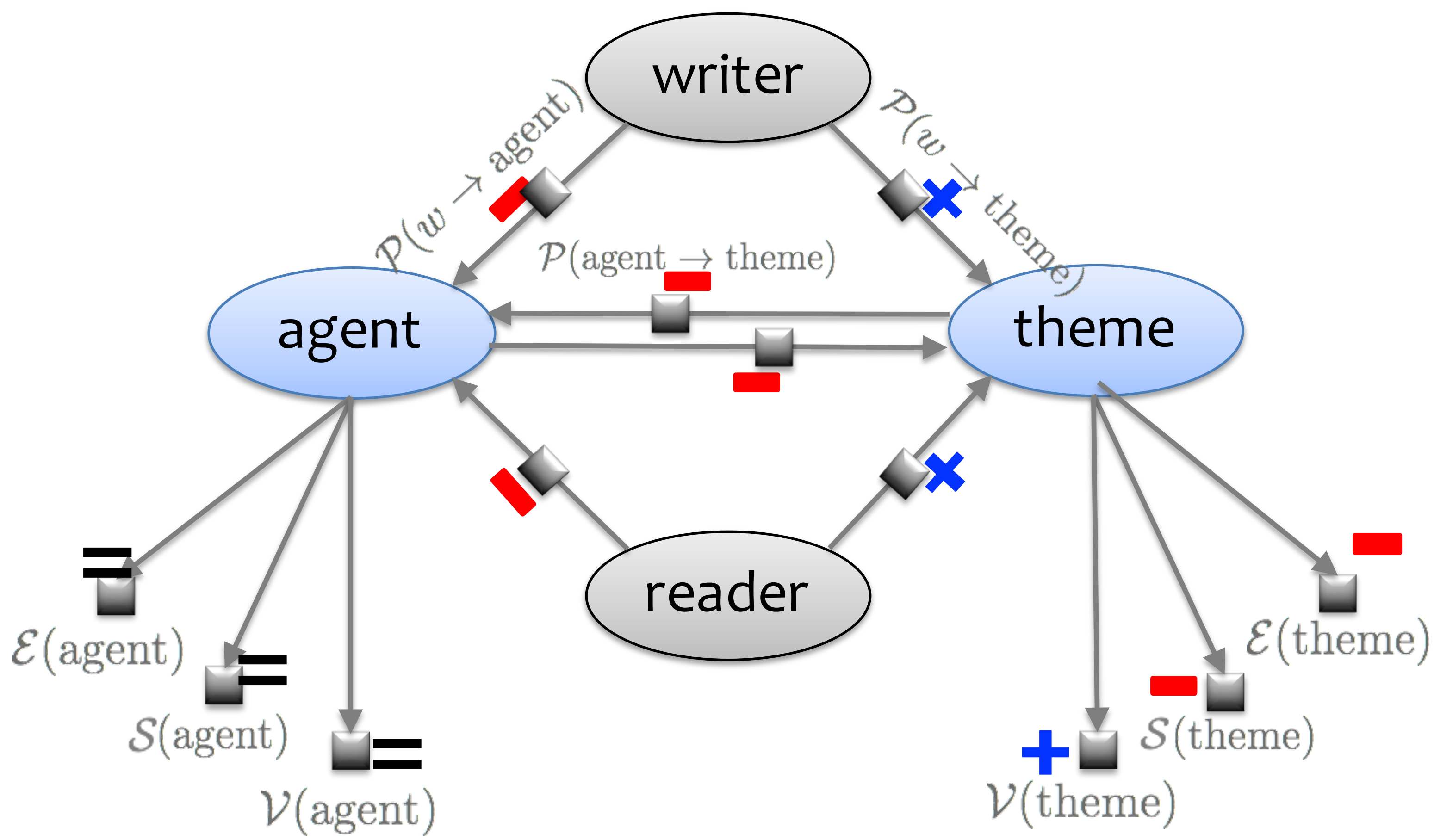
Announcement

Frame Semantics in Natural Language Processing: A Workshop in Honor of Chuck Fillmore (1929–2014) was held at the 52nd Annual Meeting of the Association for Computational Linguistics in Baltimore on June 27, 2014.

The workshop consisted of a number of inspiring invited talks about Chuck's work and his contributions to NLP.

Proceedings now available!

“X stabbing Y with Z”



Connotation Frame	Perspective(w→X) = −	The writer considers X guilty.
	Effect(Y) = −	What happened is bad for Y.
	Value(Y) = +	Y is inherently valuable.
VerbPhysics	Y < Z in RIGID	It’s not possible to stab someone with a cheeseburger.

Thanks! Questions?