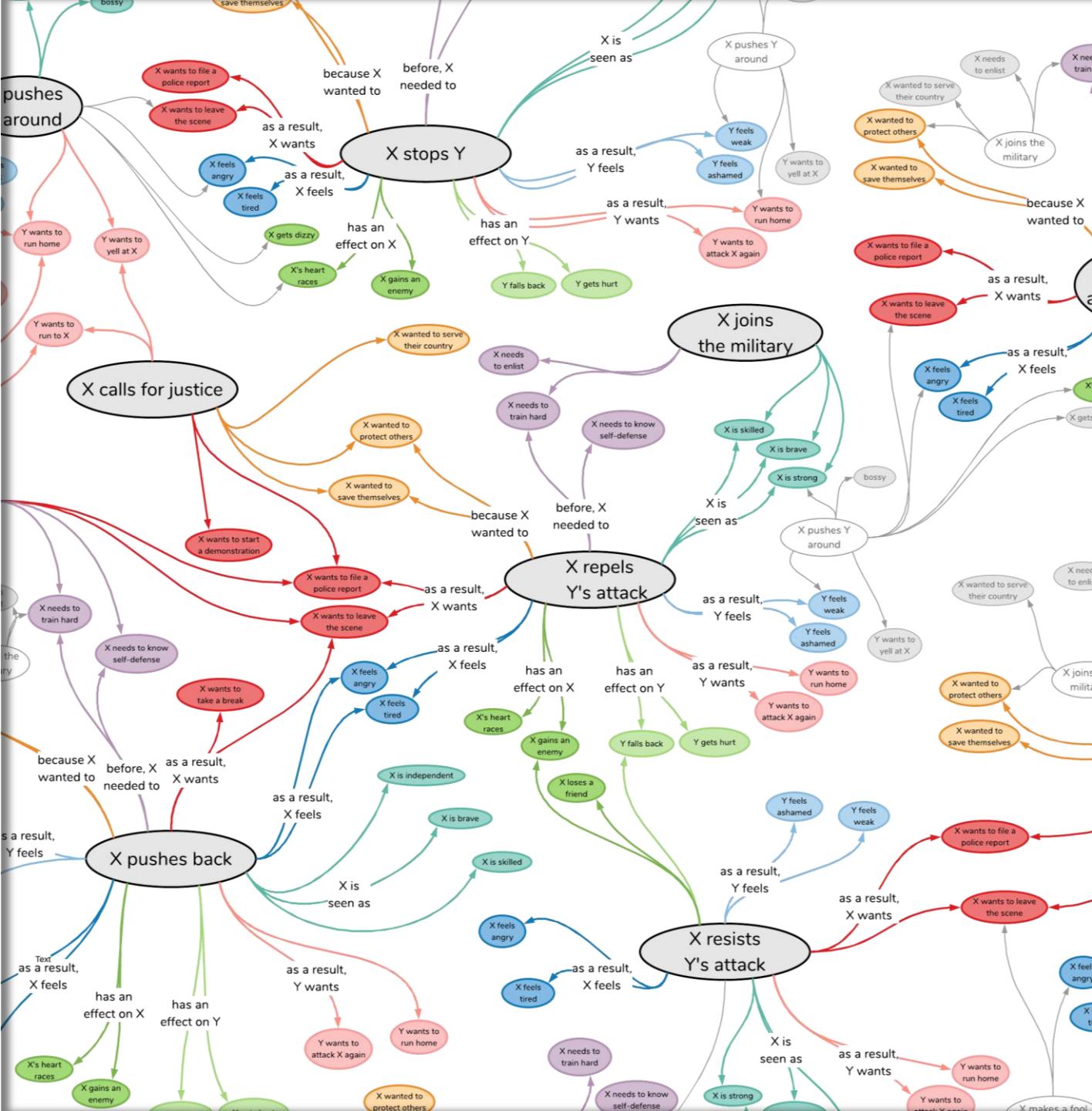


# Commonsense resources



Tom's grandma was reading a new book, when she dropped her glasses.

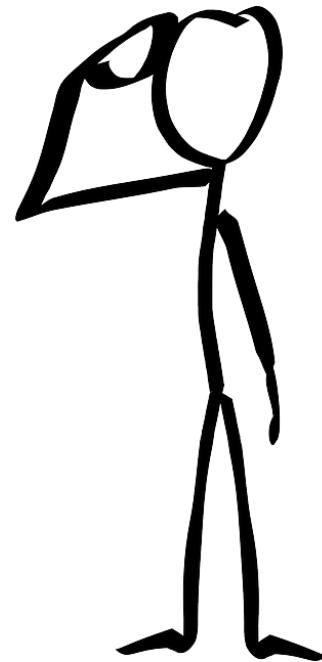
She couldn't pick them up, so she called Tom for help.

Tom rushed to help her look for them, they heard a loud crack.

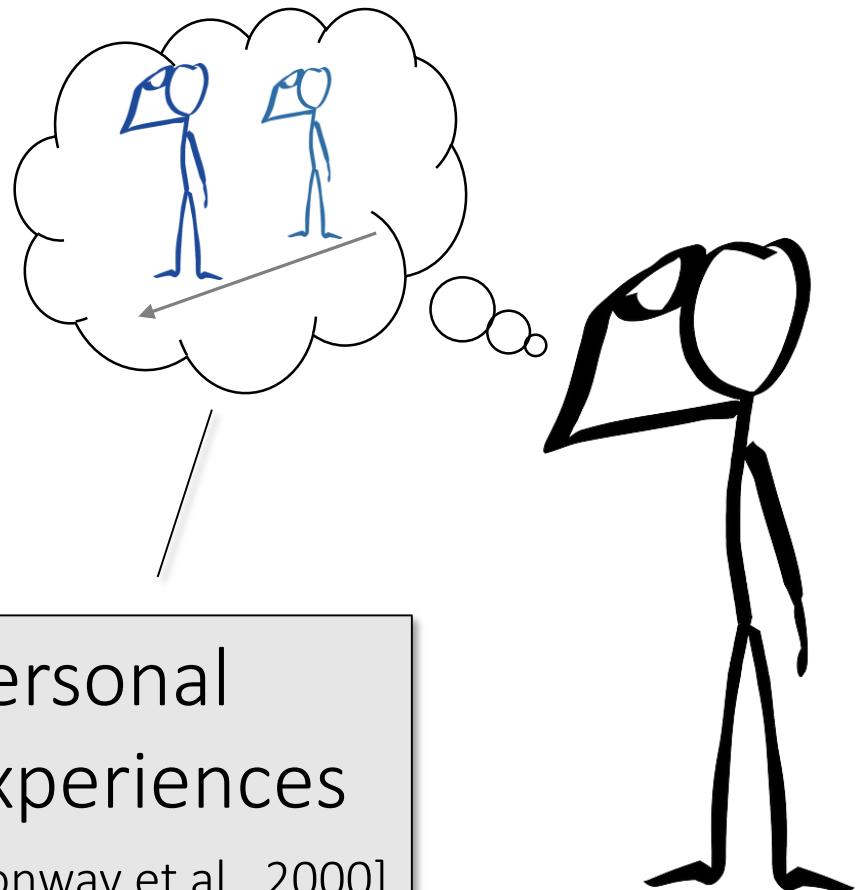
They realized that Tom broke her glasses by stepping on them.

Promptly, his grandma yelled at Tom to go get her a new pair.

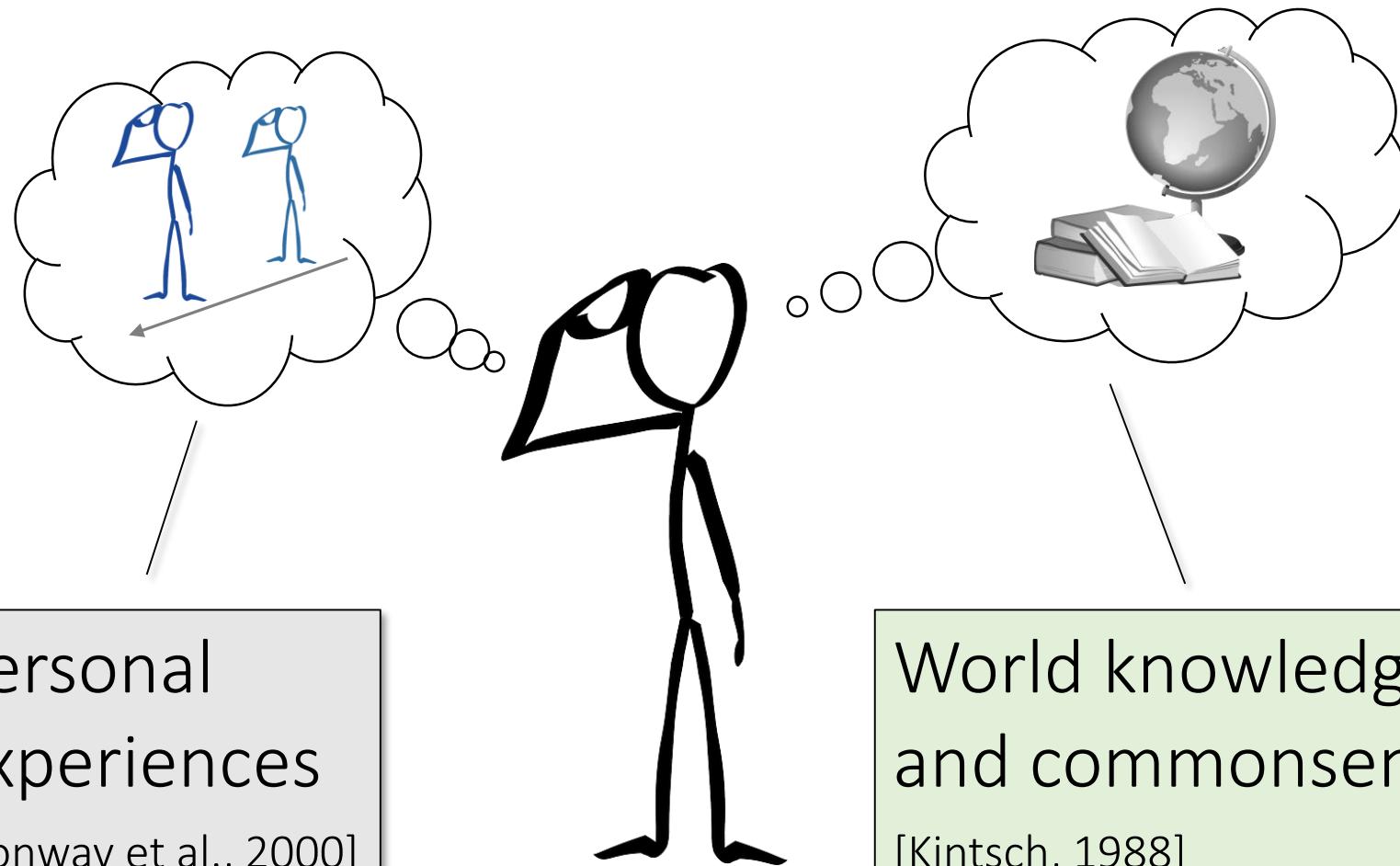
Humans reason about the world with  
mental models [Graesser, 1994]



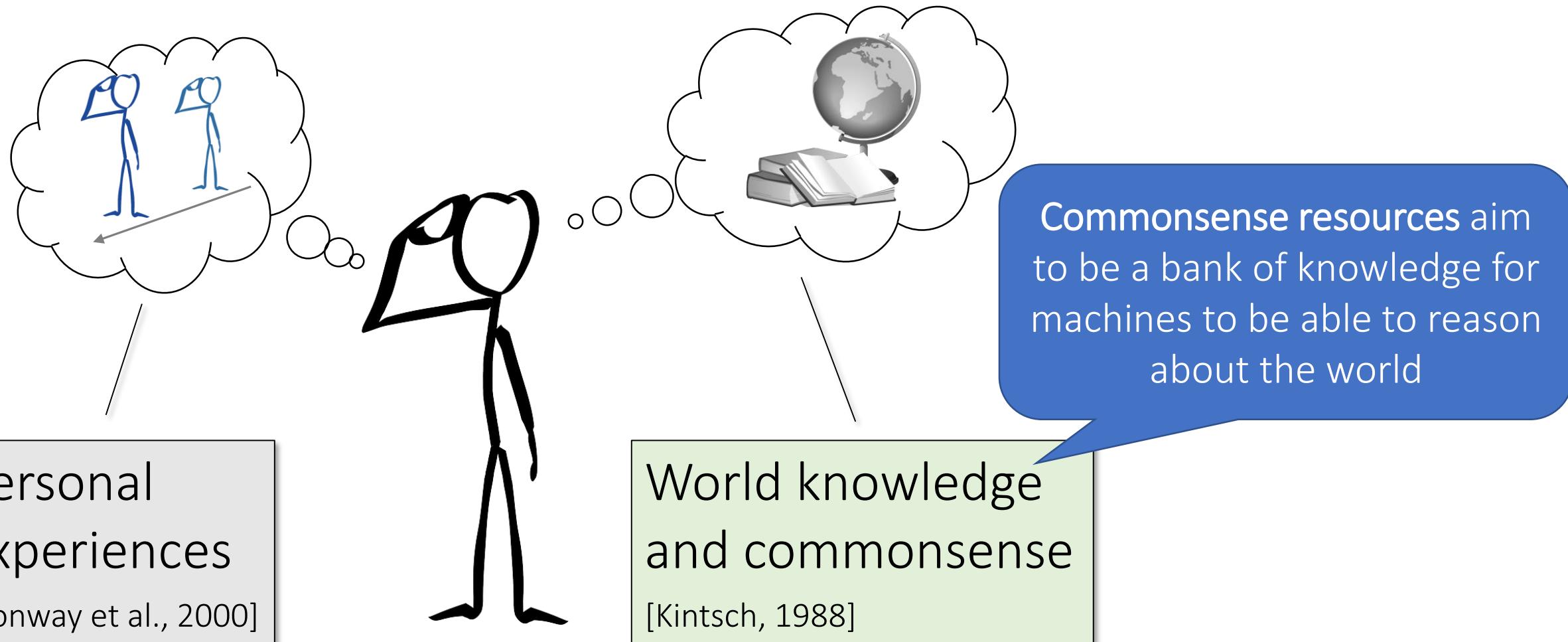
Humans reason about the world with  
mental models [Graesser, 1994]



# Humans reason about the world with mental models [Graesser, 1994]



# Humans reason about the world with mental models [Graesser, 1994]



Tom's grandma was reading a new book, when she dropped her glasses.

She couldn't pick them up, so she called Tom for help.

Tom rushed to help her look for them, they heard a loud crack.

They realized that Tom broke her glasses by stepping on them.

Promptly, his grandma yelled at Tom to go get her a new pair.

*ConceptNet*

Tom's grandma was reading a new book, when she dropped her glasses.

```
graph LR; reading[reading] -- usedFor --> glasses[glasses]
```

*ATOMIC*

She couldn't pick them up, so she called Tom for help.

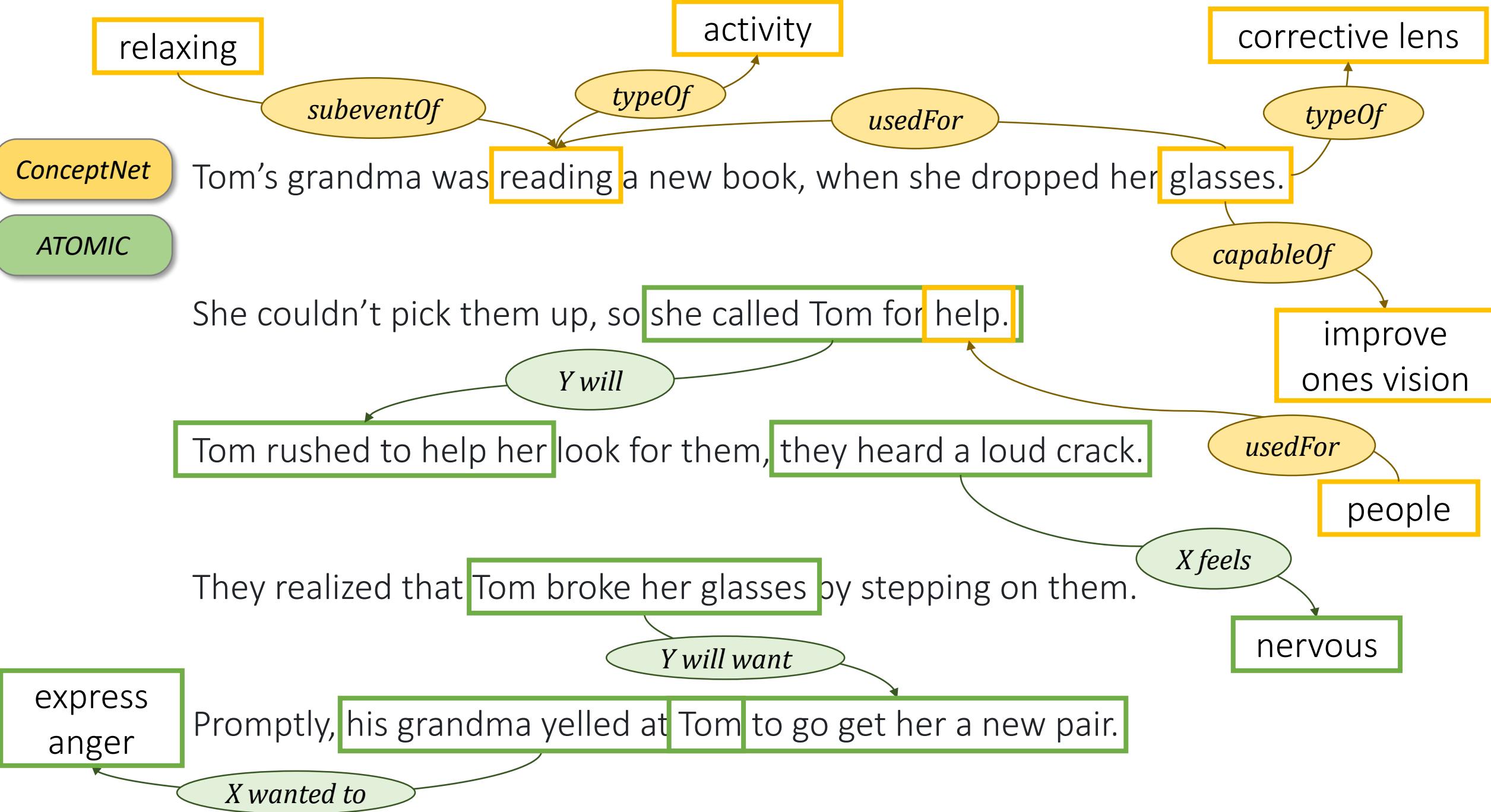
*Y will*

Tom rushed to help her look for them, they heard a loud crack.

They realized that Tom broke her glasses by stepping on them.

*Y will want*

Promptly, his grandma yelled at Tom to go get her a new pair.



# Desiderata for a good commonsense resource

## Coverage

- Large scale
- Diverse knowledge types

## Useful

- High quality knowledge
- Usable in tasks

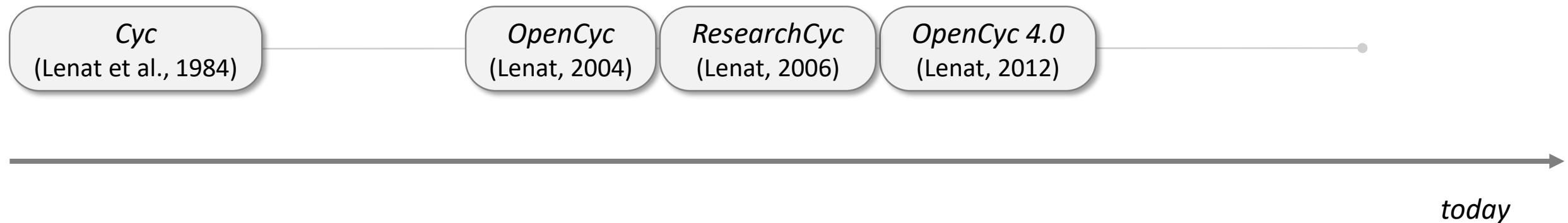
Multiple resources tackle different  
knowledge types

# Overview of existing resources

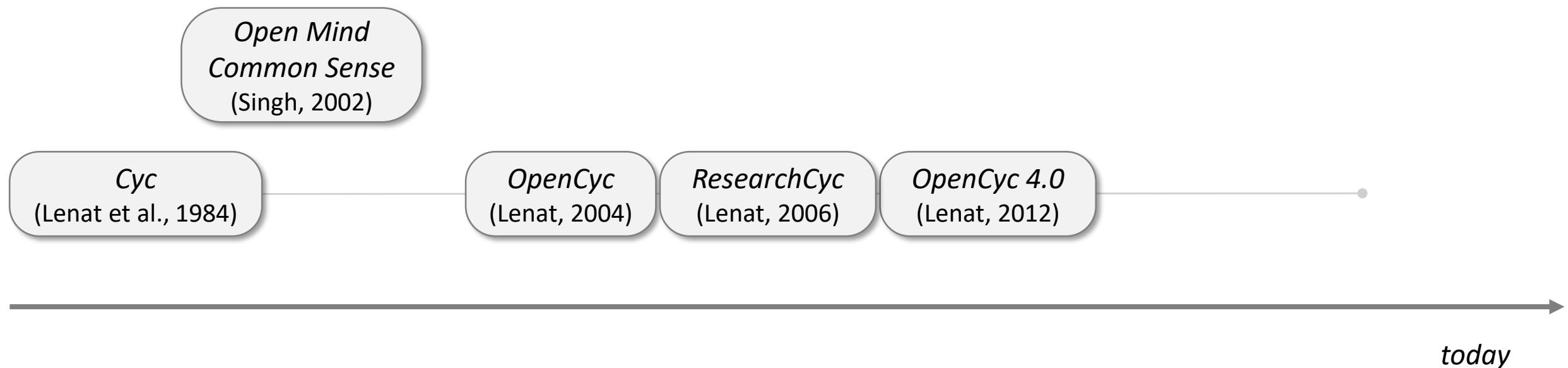
*Cyc*  
(Lenat et al., 1984)

*today*

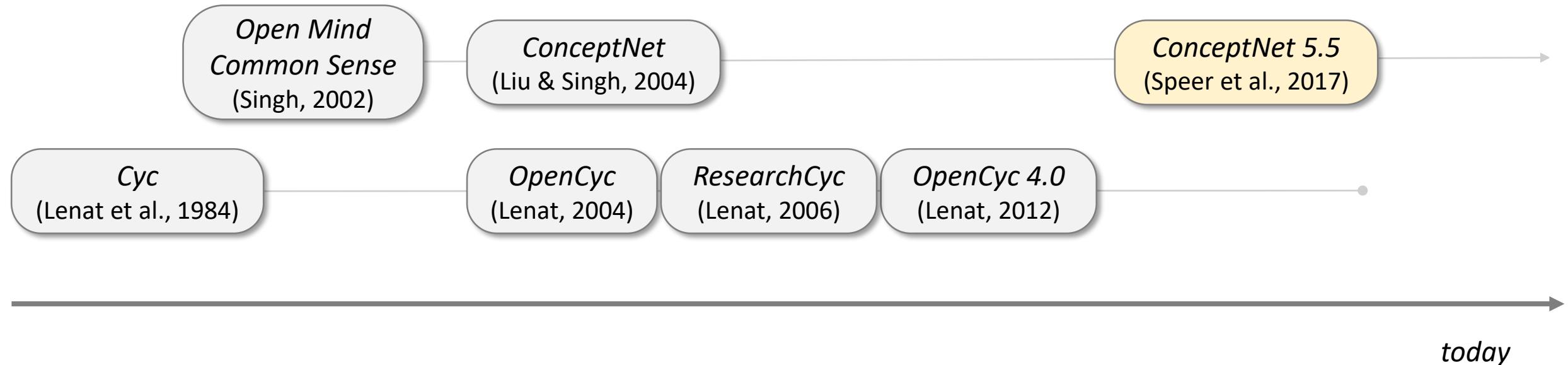
# Overview of existing resources



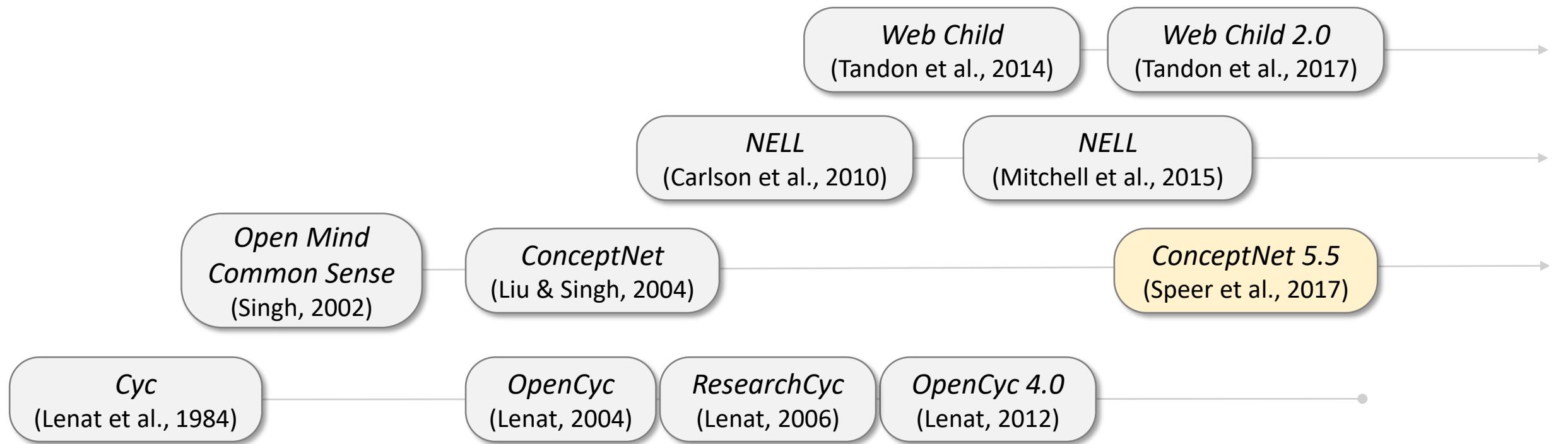
# Overview of existing resources



# Overview of existing resources

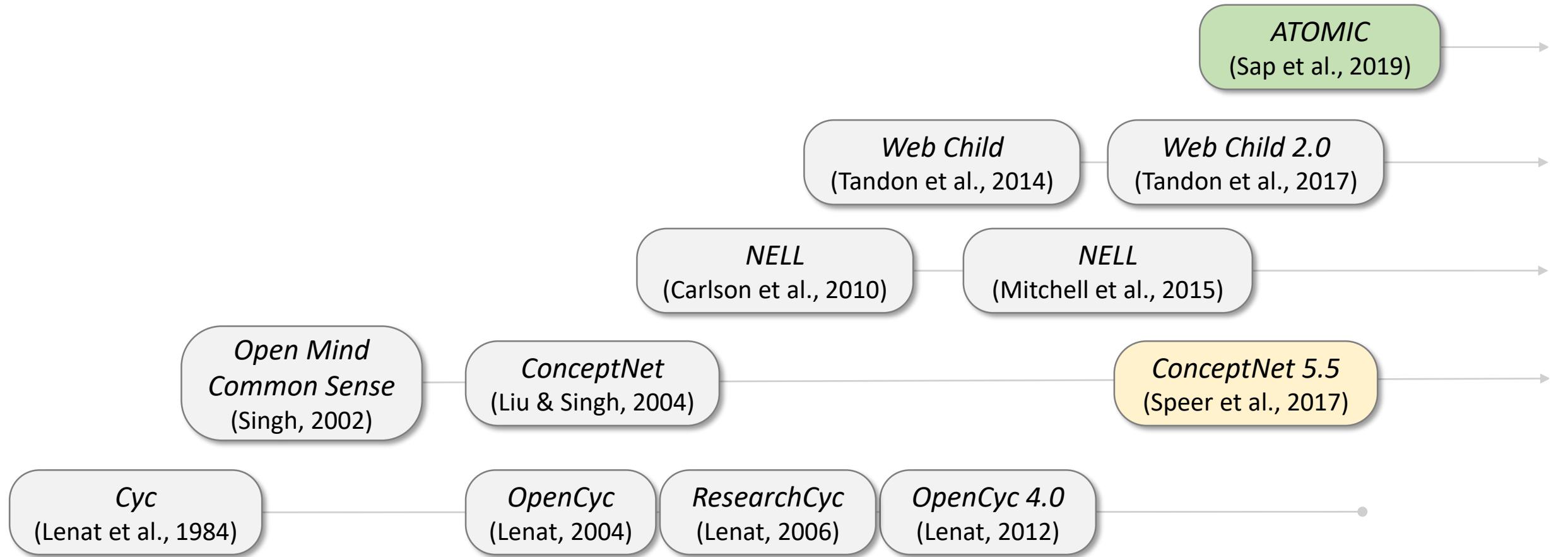


# Overview of existing resources



today

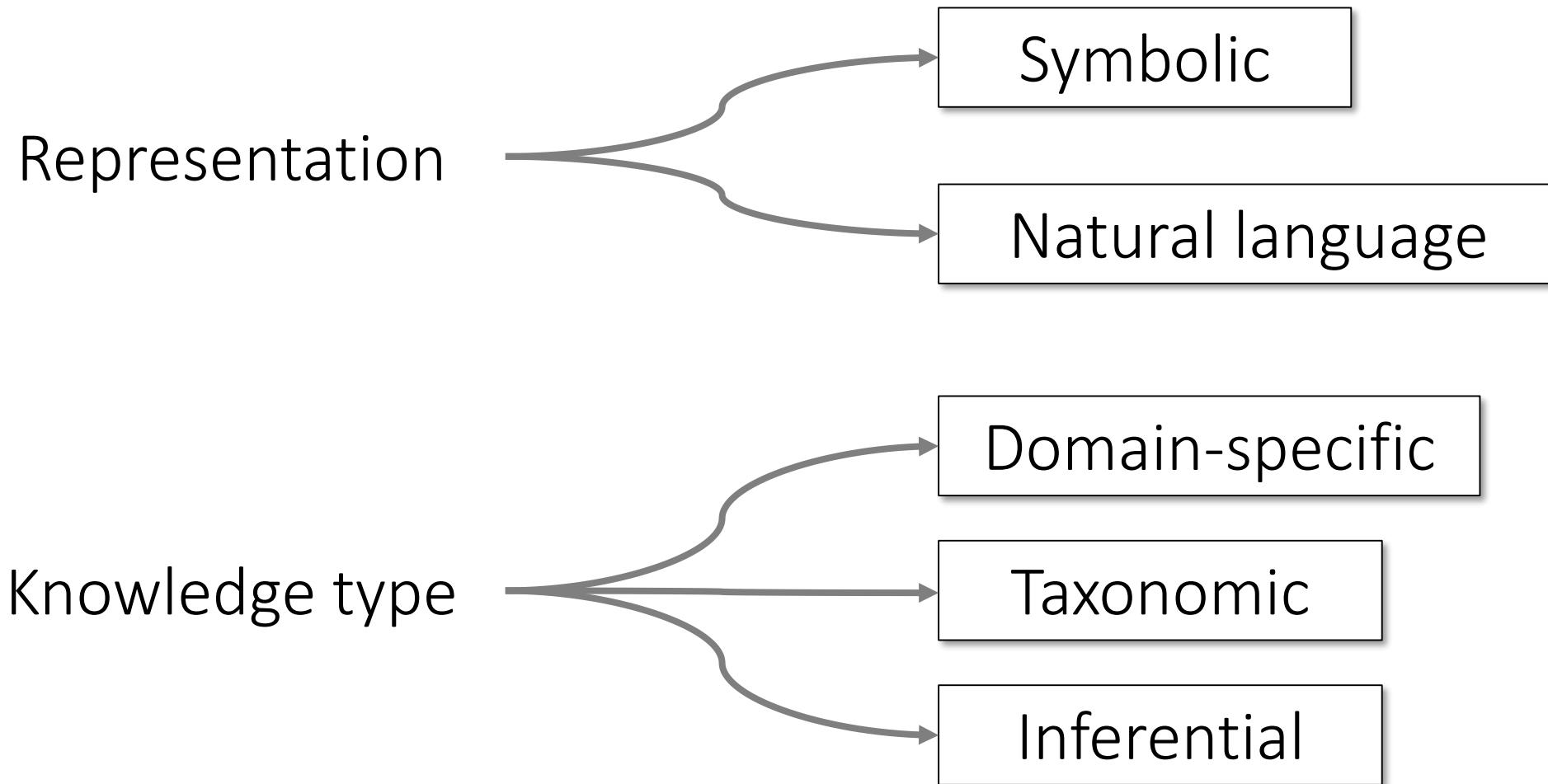
# Overview of existing resources



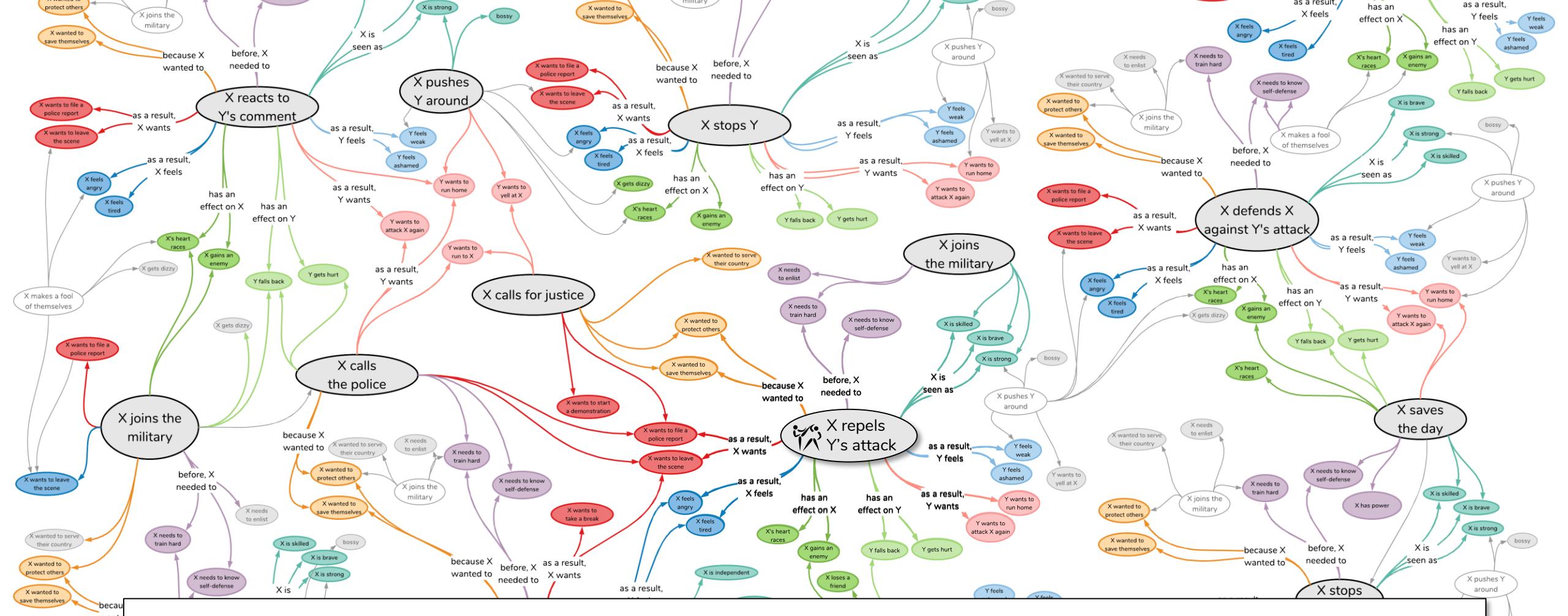
today

How do you create a commonsense resource?

# Creating a commonsense resource

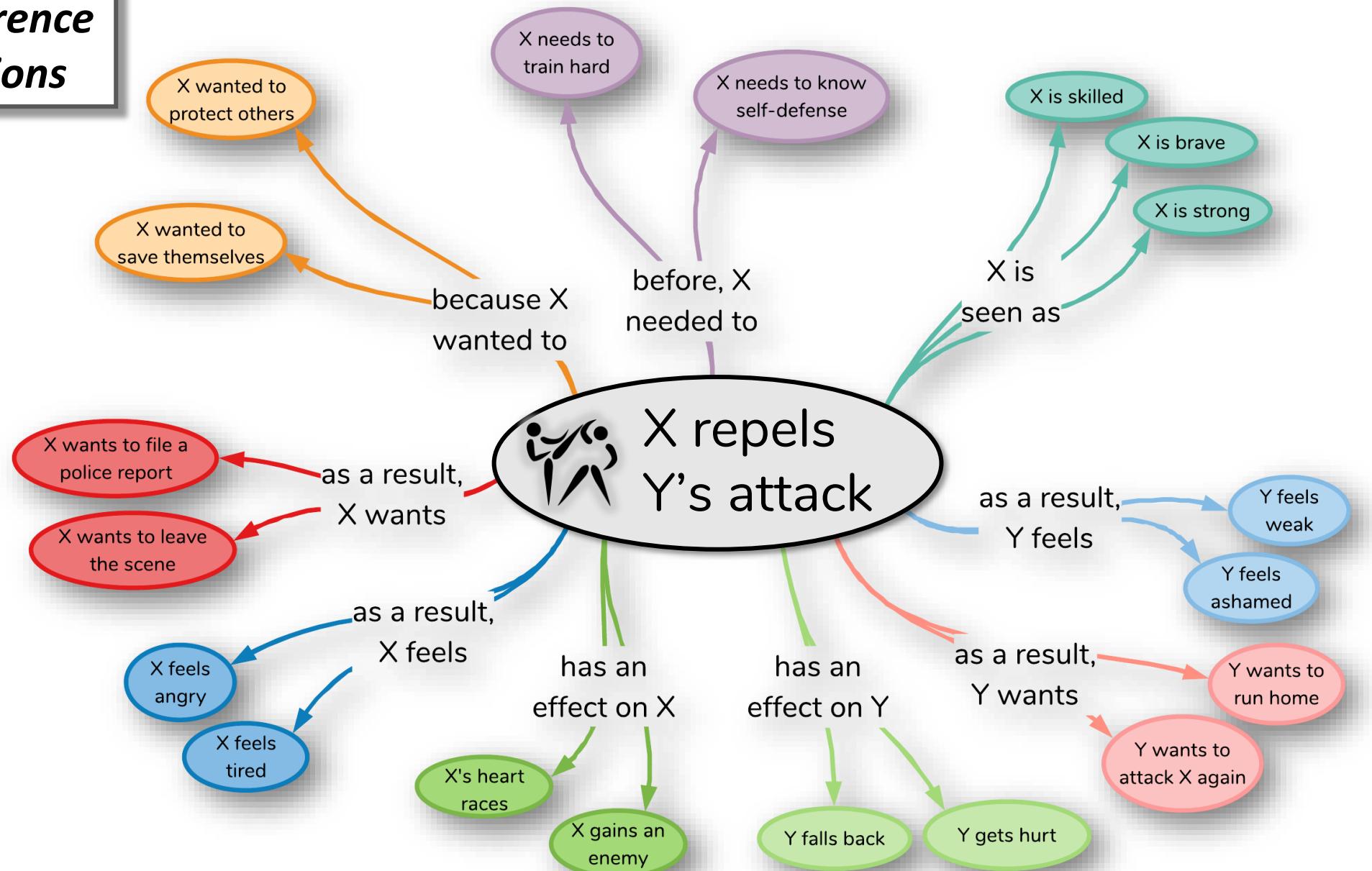


**ATOMIC:**  
*inferential knowledge in natural language form*

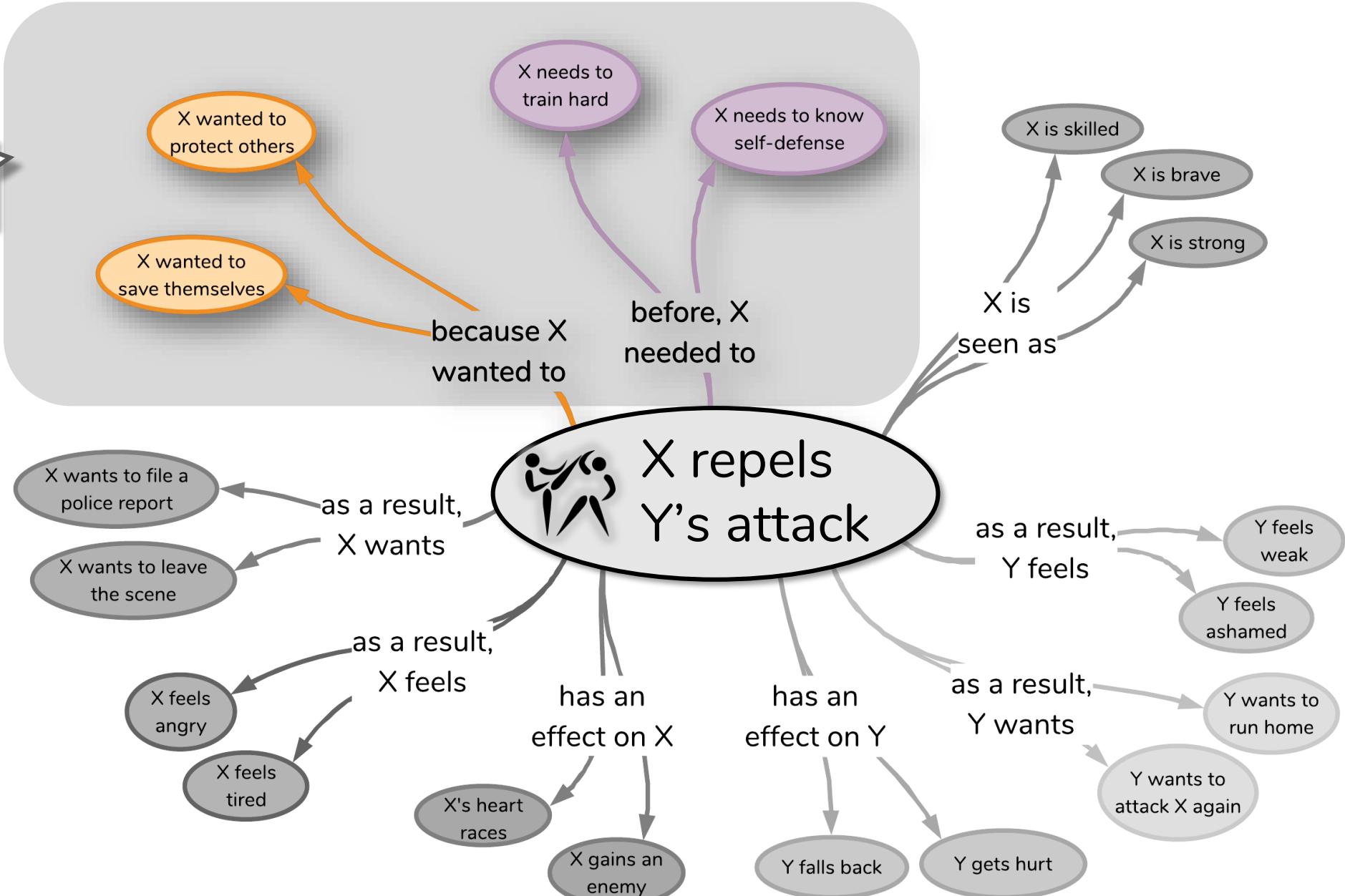


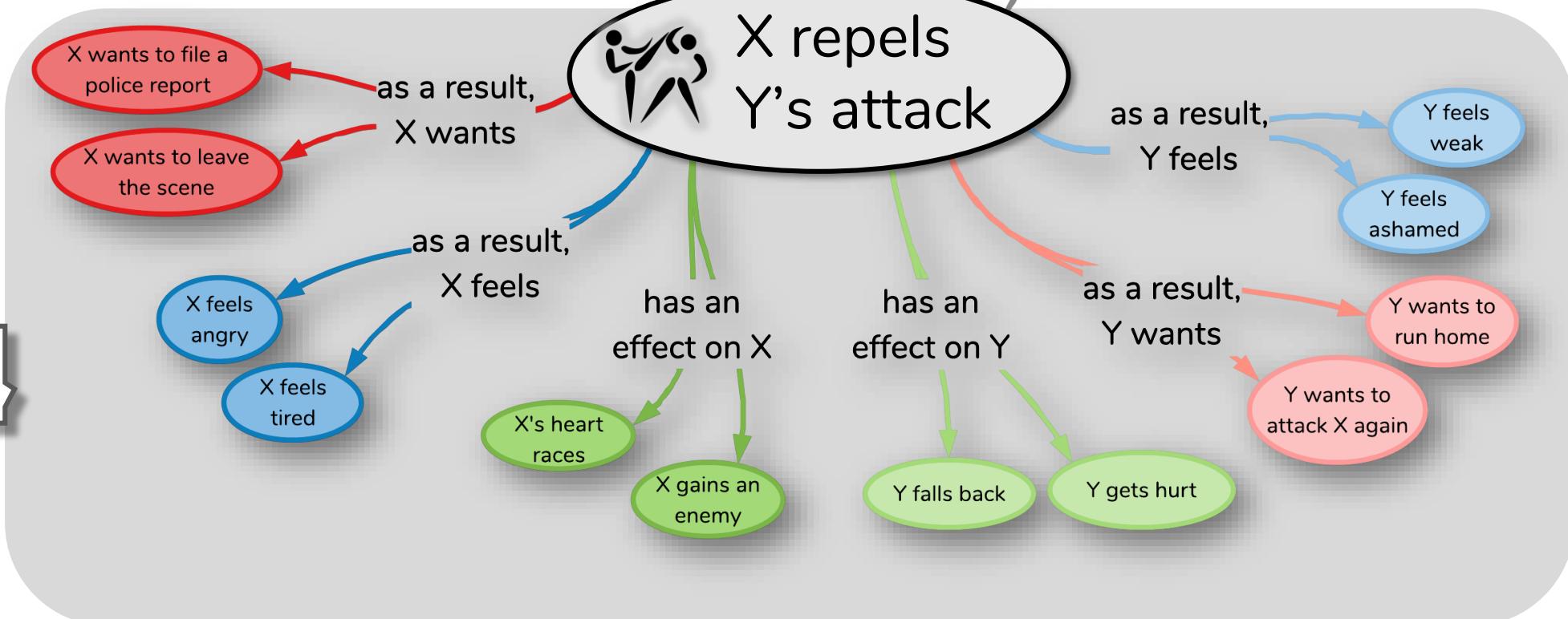
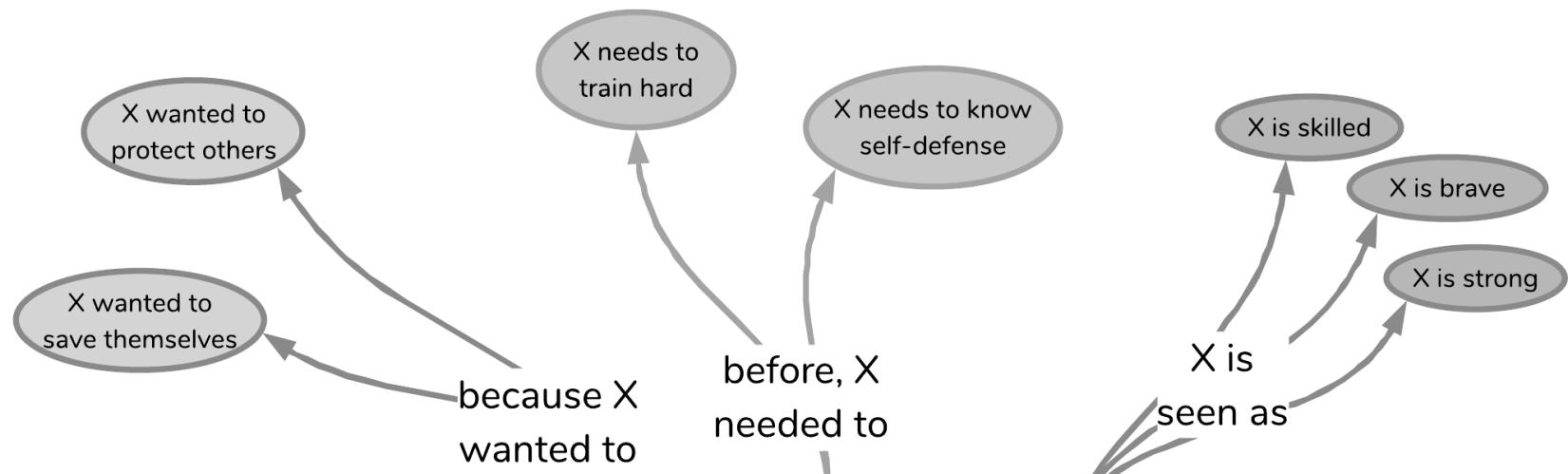
**ATOMIC: 880,000 triples for AI systems to reason about *causes* and *effects* of everyday situations**

## nine inference dimensions

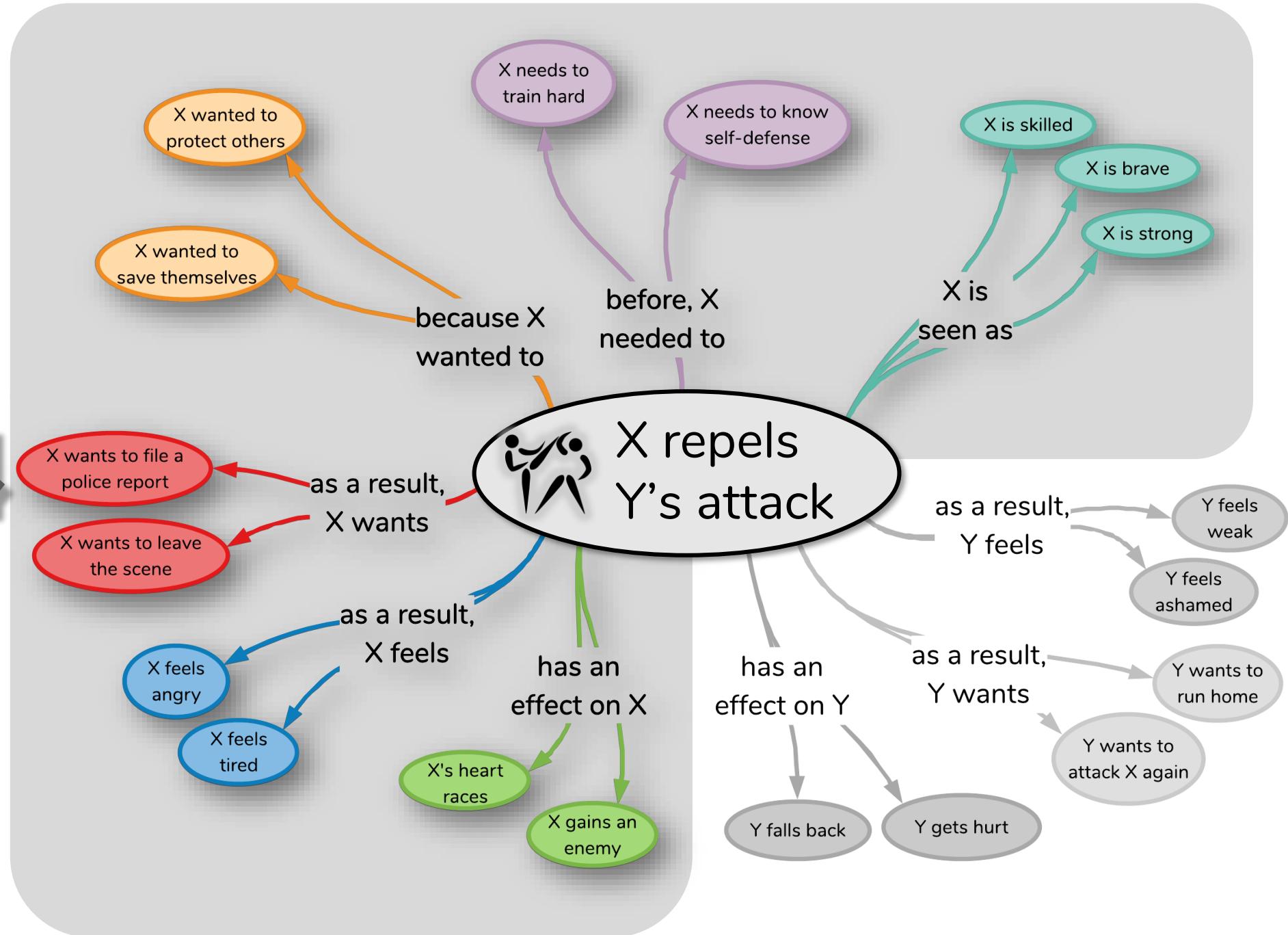


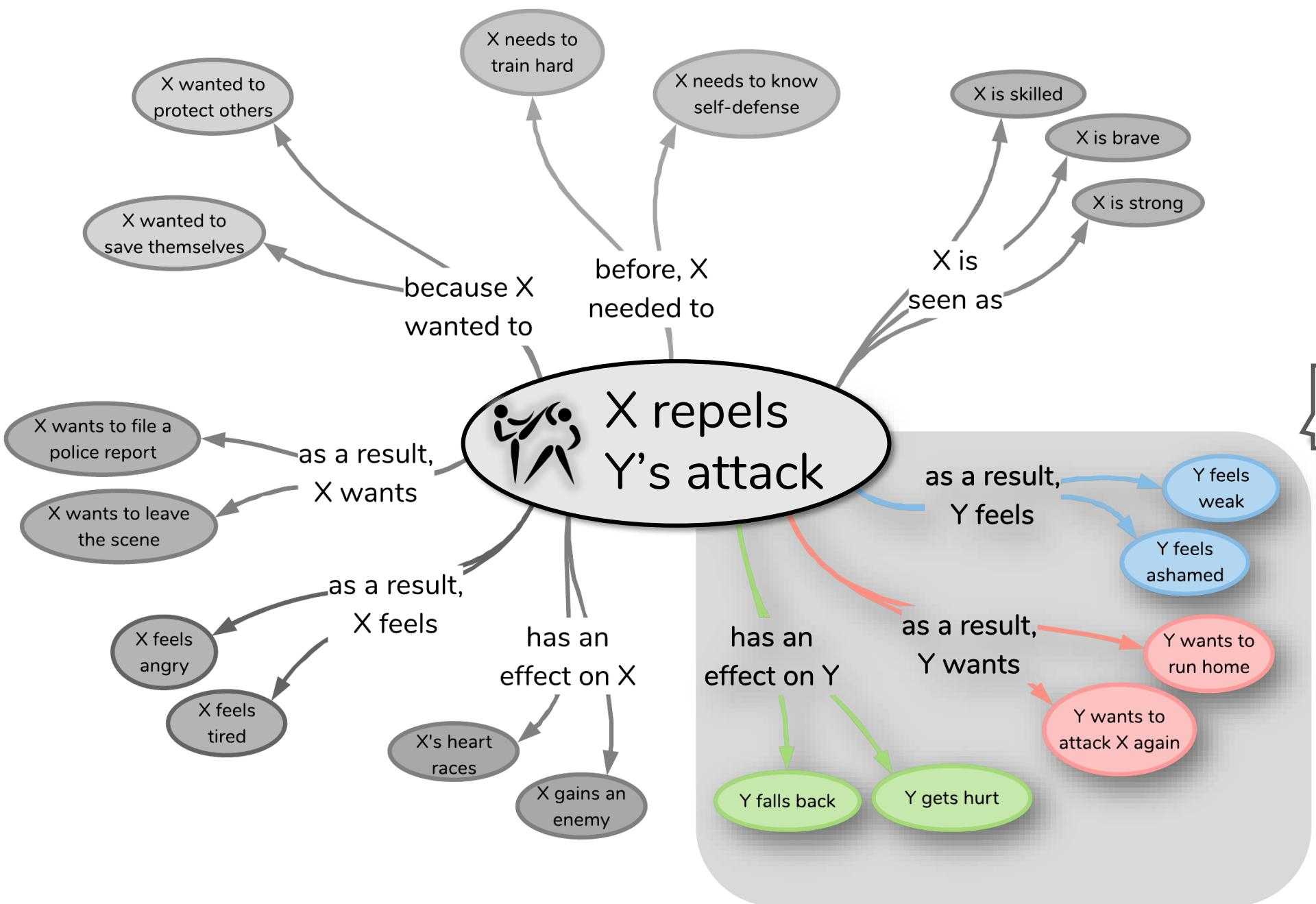
## Causes



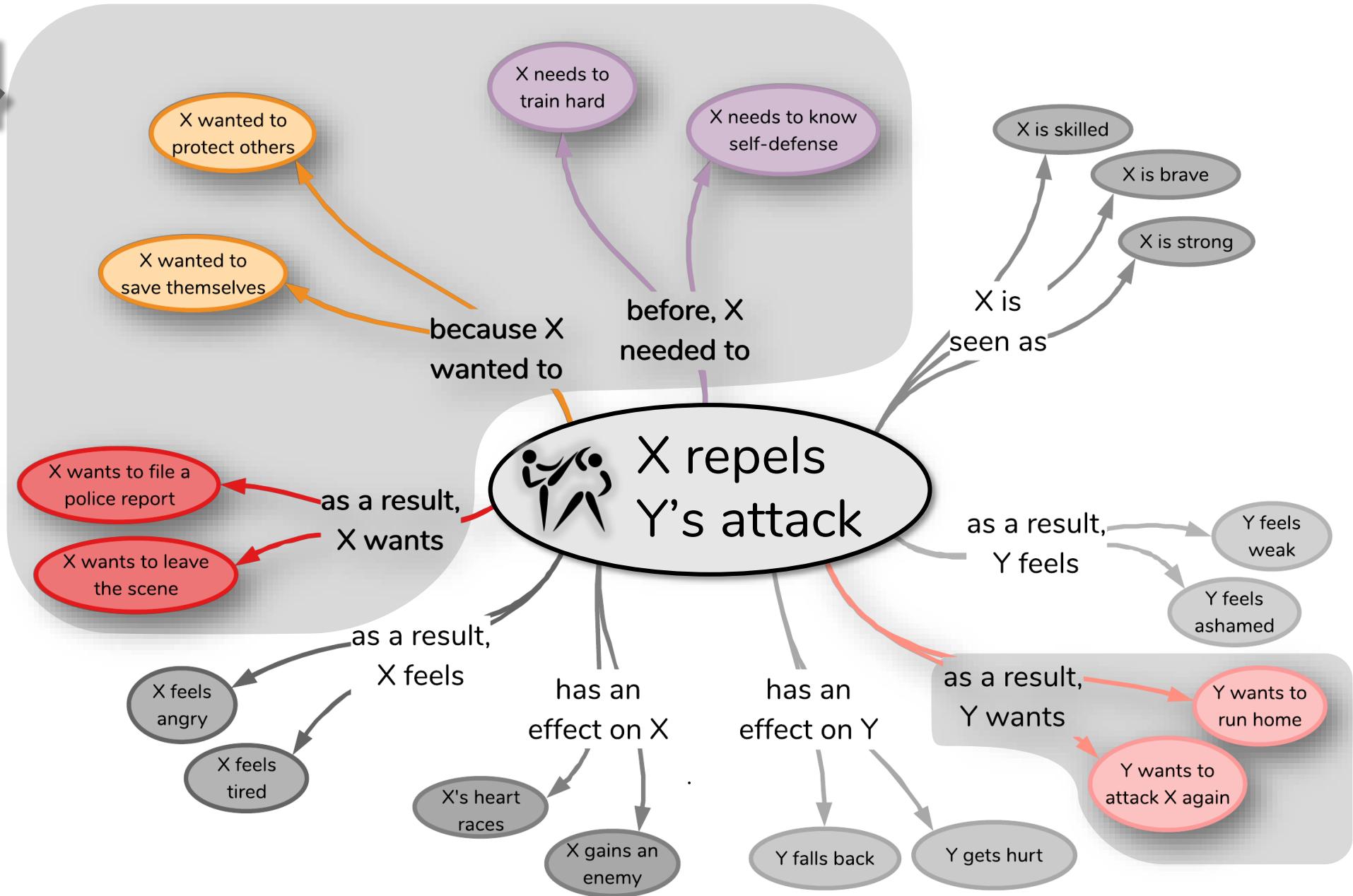


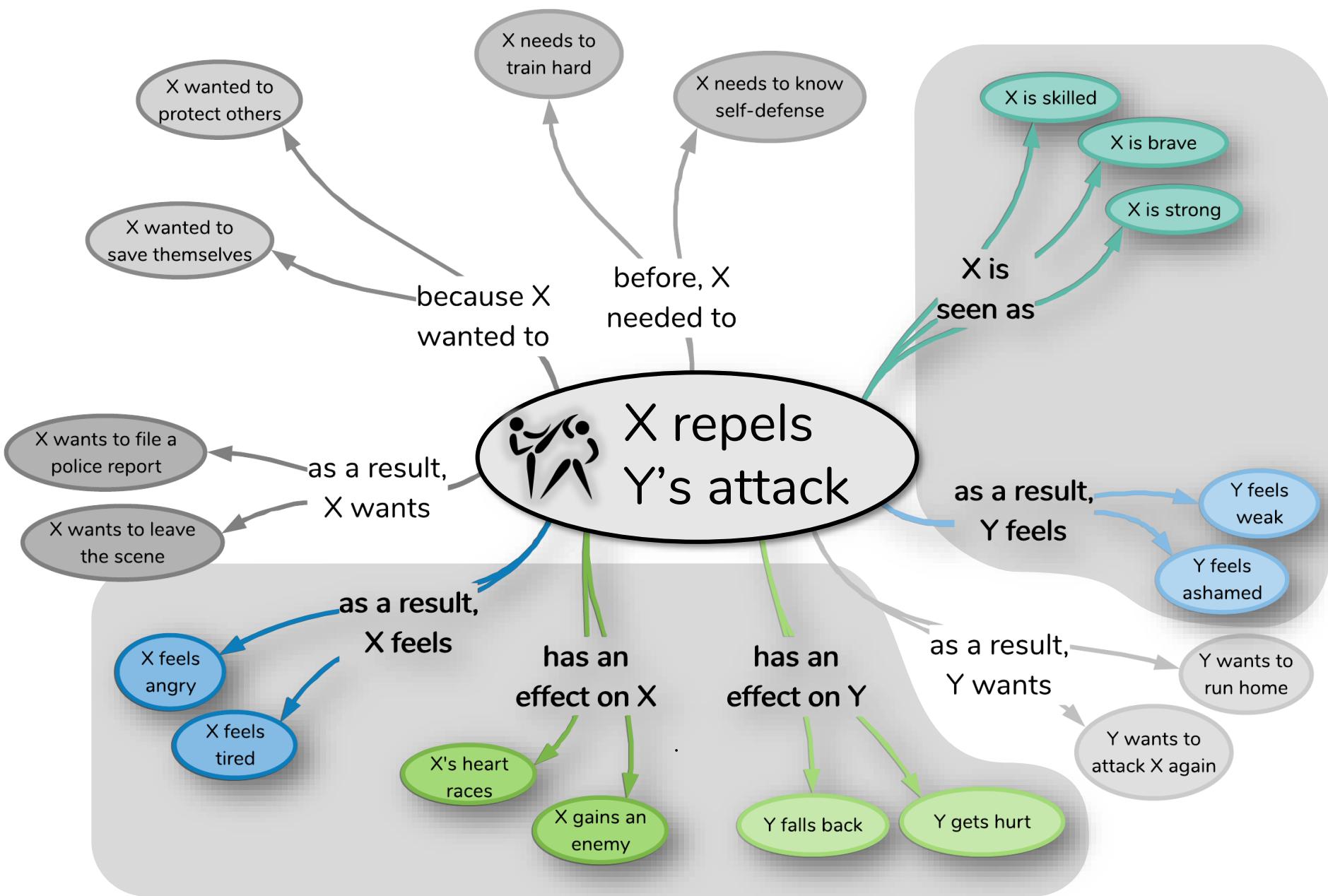
## Agent



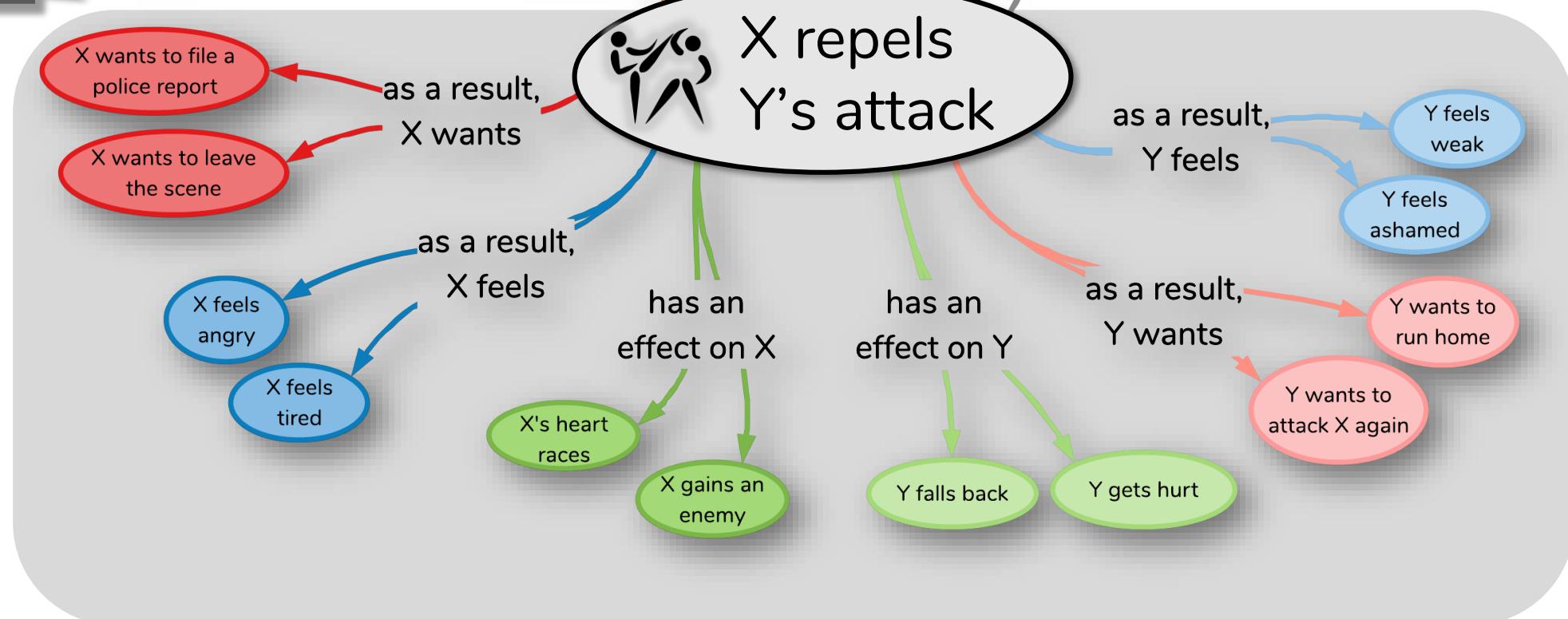


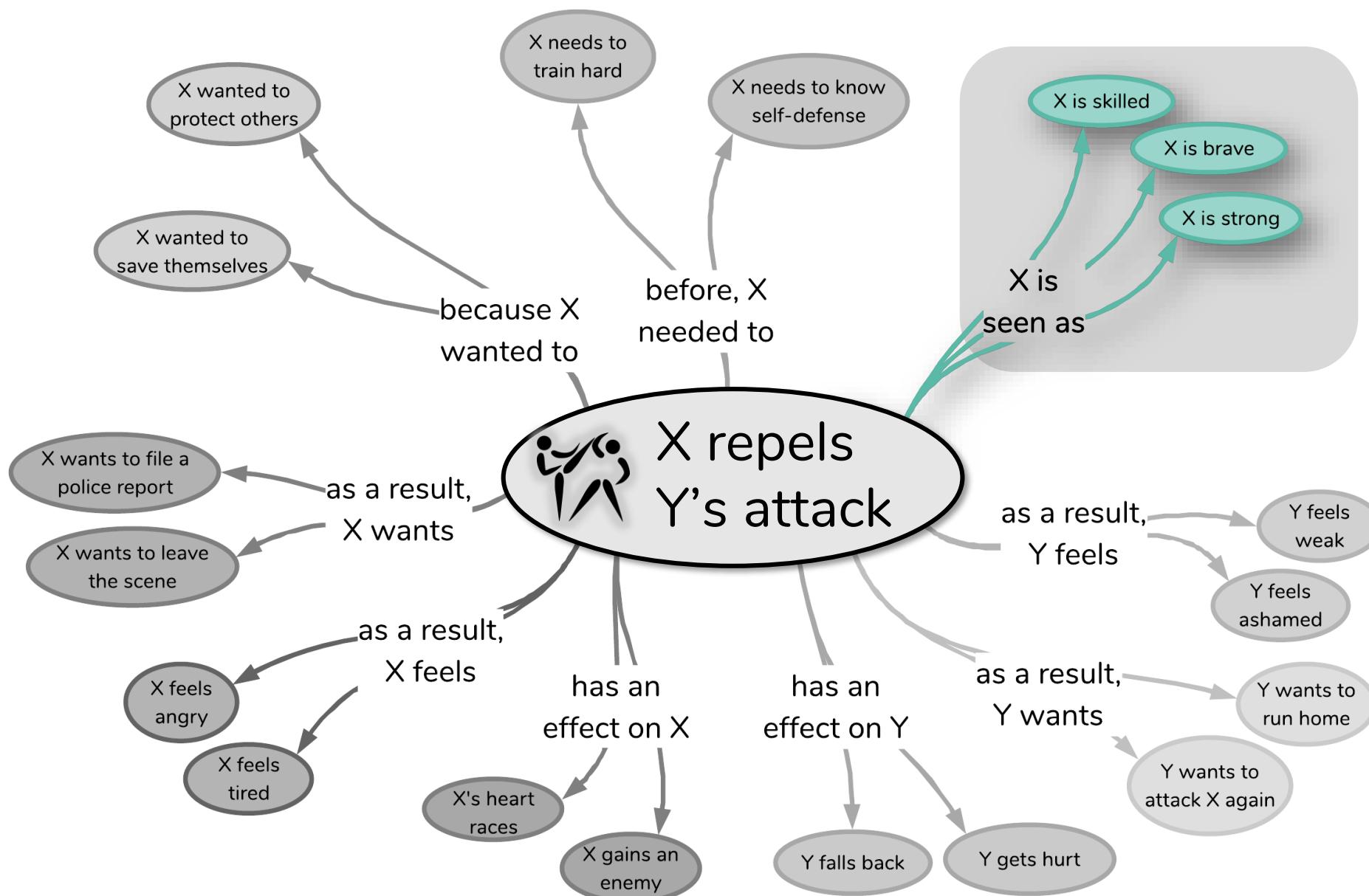
# Voluntary



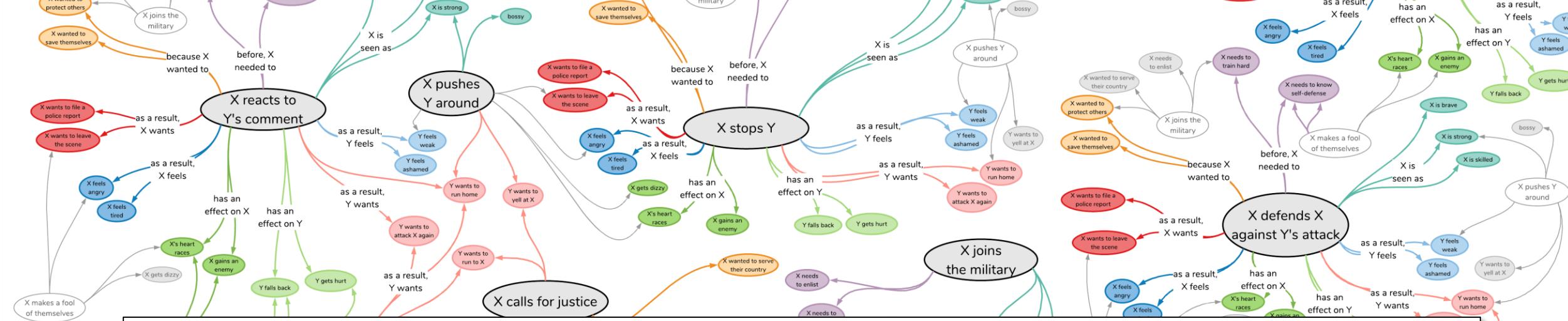


## **Dynamic**



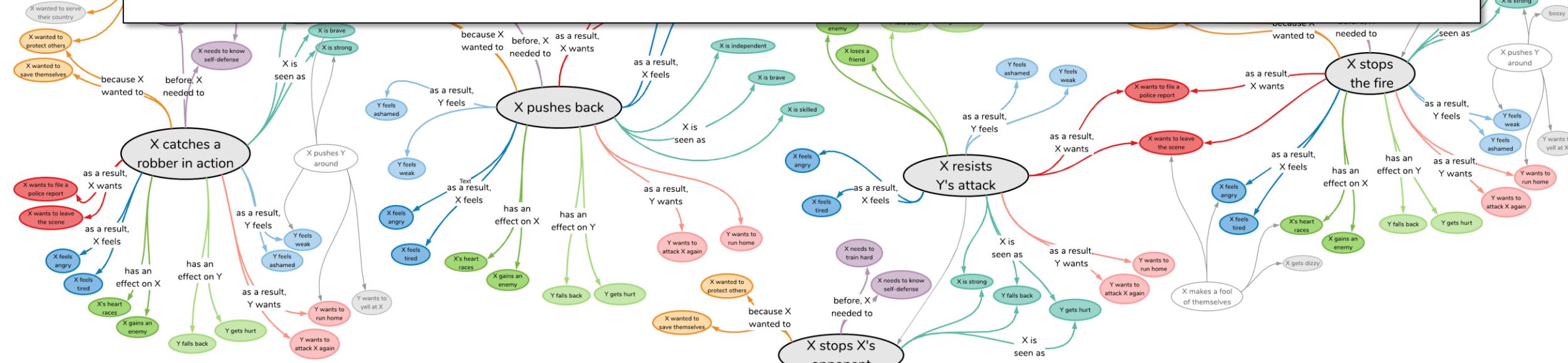


**Static**



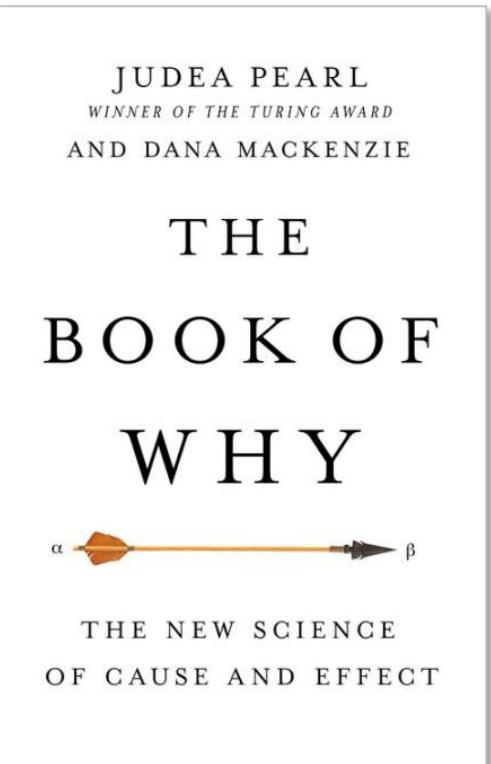
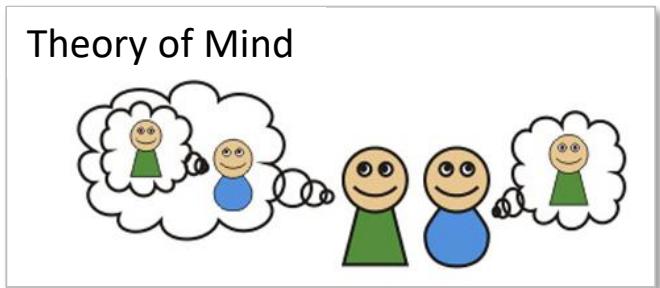
# 300,000 event nodes to date

# 880,000 if-Event-then-\* knowledge triples

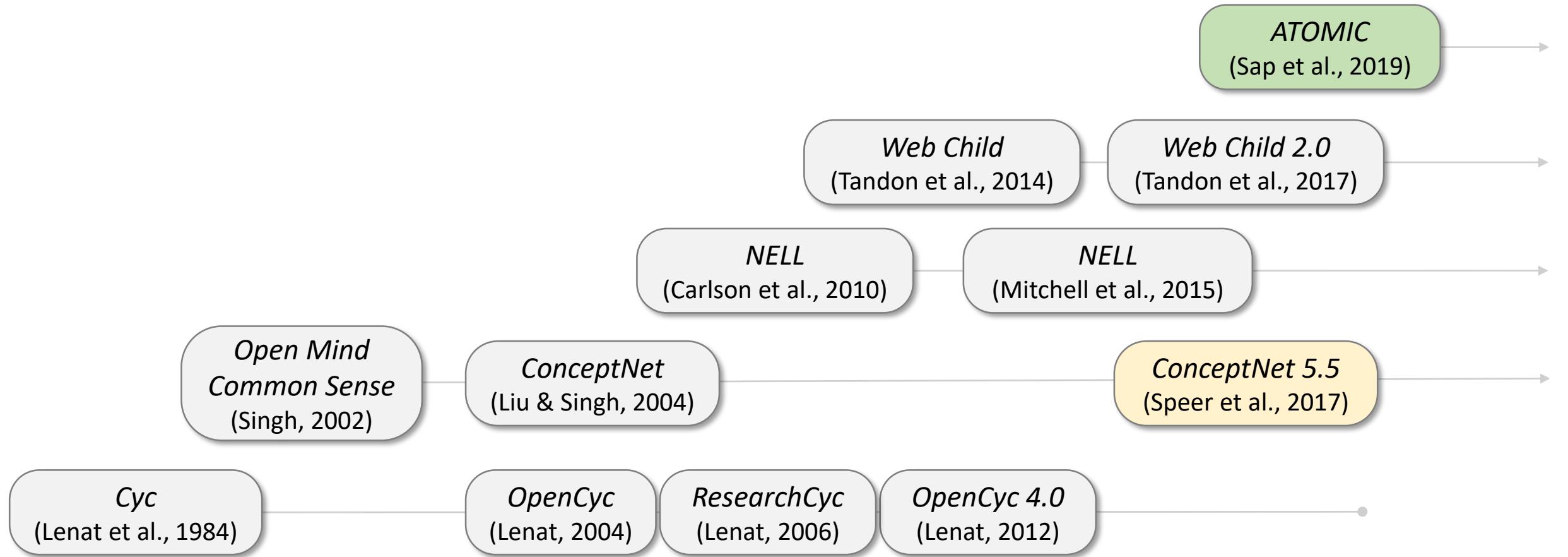


# ATOMIC: knowledge of *cause* and *effect*

- Humans have **theory of mind**, allowing us to
  - make inferences about people's mental states
  - understand **likely events** that precede and follow (Moore, 2013)
- AI systems struggle with ***inferential*** reasoning
  - only find **complex correlational patterns** in data
  - **limited to the domain** they are trained on(Pearl; Davis and Marcus 2015; Lake et al. 2017; Marcus 2018)



# Overview of existing resources



today

# Existing knowledge bases

*ATOMIC*

(Sap et al., 2019)

*NELL*

(Mitchell et al., 2015)

*ConceptNet 5.5*

(Speer et al., 2017)

*OpenCyc 4.0*

(Lenat, 2012)

# Existing knowledge bases

Represented in **symbolic logic**  
(e.g., LISP-style logic)

*NELL*  
(Mitchell et al., 2015)

*OpenCyc 4.0*  
(Lenat, 2012)

Represented in **natural language**  
(how humans *talk* and *think*)

*ConceptNet 5.5*  
(Speer et al., 2017)

*ATOMIC*  
(Sap et al., 2019)

# Existing knowledge bases

Represented in **symbolic logic**  
(e.g., LISP-style logic)

*NELL*  
(Mitchell et al., 2015)

*OpenCyc 4.0*  
(Lenat, 2012)

Represented in **natural language**  
(how humans *talk* and *think*)

*ConceptNet 5.5*  
(Speer et al., 2017)

*ATOMIC*  
(Sap et al., 2019)

```
(#$implies
  (#$and
    (#$isa ?OBJ ?SUBSET)
    (#$genls ?SUBSET ?SUPERSET))
  (#$isa ?OBJ ?SUPERSET))
```

# Existing knowledge bases

Represented in **symbolic logic**  
(e.g., LISP-style logic)

*NELL*  
(Mitchell et al., 2015)

*OpenCyc 4.0*  
(Lenat, 2012)

Represented in **natural language**  
(how humans *talk* and *think*)

*ConceptNet 5.5*  
(Speer et al., 2017)

Knowledge of “**what**”  
(taxonomic: A *isA* B)

Knowledge of “**why**” and “**how**”  
(inferential: *causes* and *effects*)

*ATOMIC*  
(Sap et al., 2019)

**Q:** How do you gather commonsense knowledge at scale?

**A:** It depends on the type of knowledge

# Extracting commonsense from text

Based on information extraction (IE) methods

1. Read text
  2. Create candidate rules
  3. Filter rules

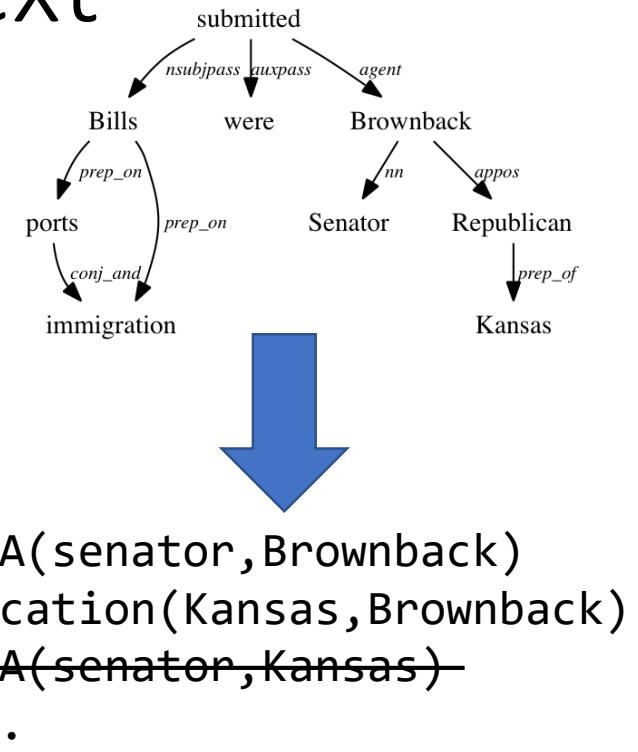
## Advantage:

can extract knowledge automatically

## Example system:

## Never Ending Language Learner (*NELL*; Carlson et al., 2010)

*... more on this later with temporal commonsense*



# Some commonsense cannot be extracted

Text is subject to **reporting bias**

(Gordon & Van Durme, 2013)

- Idioms & figurative usage
  - Black sheep problem
- Noteworthy events
  - Murdering 4x more common than exhaling

Commonsense is not often written

-> *Grice's maxim of quantity*

# Some commonsense cannot be extracted

Text is subject to **reporting bias**  
(Gordon & Van Durme, 2013)

- Idioms & figurative usage  
Black sheep problem
- Noteworthy events  
Murdering 4x more common than exhaling



Commonsense is not often written  
-> *Grice's maxim of quantity*

# Some commonsense cannot be extracted

Text is subject to **reporting bias**  
(Gordon & Van Durme, 2013)

- Idioms & figurative usage  
Black sheep problem
- Noteworthy events  
Murdering 4x more common than exhaling

Commonsense is not often written  
-> *Grice's maxim of quantity*



found when extracting commonsense knowledge on four large corpora using Knext (Gordon & Van Durme, 2013)

# Some commonsense cannot be extracted

Text is subject to **reporting bias**  
(Gordon & Van Durme, 2013)

- Idioms & figurative usage  
Black sheep problem
- Noteworthy events  
Murdering 4x more common than exhaling

Commonsense is not often written  
-> *Grice's maxim of quantity*



found when extracting commonsense knowledge on four large corpora using Knext (Gordon & Van Durme, 2013)

# Eliciting commonsense from humans

# Eliciting commonsense from humans

**Experts create knowledge base**

*OpenCyc 4.0*  
(Lenat, 2012)

*WordNet*  
(Miller et al., 1990)

# Eliciting commonsense from humans

## Experts create knowledge base

- Advantages:
  - Quality guaranteed
  - Can use complex representations  
(e.g., CycL, LISP)
- Drawbacks:
  - Time cost
  - Training users

*OpenCyc 4.0*  
(Lenat, 2012)

*WordNet*  
(Miller et al., 1990)

# Eliciting commonsense from humans

## Experts create knowledge base

- Advantages:
  - Quality guaranteed
  - Can use complex representations  
(e.g., CycL, LISP)
- Drawbacks:
  - Time cost
  - Training users

*OpenCyc 4.0*  
(Lenat, 2012)

*WordNet*  
(Miller et al., 1990)

# Eliciting commonsense from humans

## Experts create knowledge base

- Advantages:
  - Quality guaranteed
  - Can use complex representations (e.g., CycL, LISP)
- Drawbacks:
  - Time cost
  - Training users

*OpenCyc 4.0*  
(Lenat, 2012)

*WordNet*  
(Miller et al., 1990)

## Non-experts write knowledge in natural language phrases

*ATOMIC*  
(Sap et al., 2019)

*ConceptNet 5.5*  
(Speer et al., 2017)

# Eliciting commonsense from humans

## Experts create knowledge base

- Advantages:
  - Quality guaranteed
  - Can use complex representations (e.g., CycL, LISP)
- Drawbacks:
  - Time cost
  - Training users

*OpenCyc 4.0*  
(Lenat, 2012)

*WordNet*  
(Miller et al., 1990)

## Non-experts write knowledge in natural language phrases

- Natural language
  - Accessible to non-experts
  - Different phrasings allow for more nuanced knowledge
- Fast and scalable collection
  - Crowdsourcing
  - Games with a purpose

*ATOMIC*  
(Sap et al., 2019)

*ConceptNet 5.5*  
(Speer et al., 2017)

# Eliciting commonsense from humans

## Experts create knowledge base

- Advantages:
  - Quality guaranteed
  - Can use complex representations (e.g., CycL, LISP)
- Drawbacks:
  - Time cost
  - Training users

*OpenCyc 4.0*  
(Lenat, 2012)

*WordNet*  
(Miller et al., 1990)

## Non-experts write knowledge in natural language phrases

- Natural language
  - Accessible to non-experts
  - Different phrasings allow for more nuanced knowledge
- Fast and scalable collection
  - Crowdsourcing
  - Games with a purpose

*ATOMIC*  
(Sap et al., 2019)

*ConceptNet 5.5*  
(Speer et al., 2017)



# 880,000 if-Event-then-\* knowledge triples

