Event2Mind: Commonsense Inference on Events, Intents, and Reactions

Hannah Rashkin*, Maarten Sap*, Emily Allaway, Noah A. Smith, and Yejin Choi

Paul G. Allen School for Computer Science & Engineering
University of Washington, Seattle, USA

*: equal contribution

uwnlp.github.io/event2mind

Takeaways

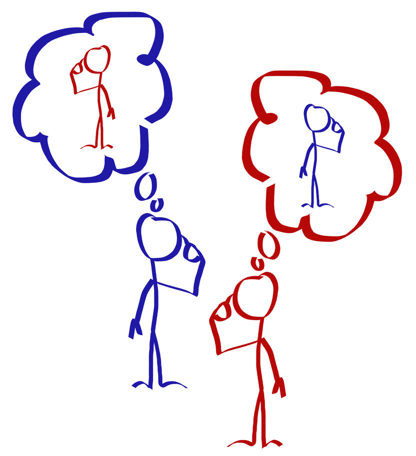
In this work, we:

- introduce the **new task** of reasoning about and generating people's **intents and reactions** in relation to events
- create **a new large knowledge graph of 77K nodes** that supports this type of commonsense inference
- provide a **novel way to analyze gender bias** in movies using commonsense inference

Commonsense Inference

Pragmatic reasoning about the mental states of people in relation to events

- Numerous AI systems need to anticipate people's *intents* and emotional *reactions* (e.g., conversational AI, ad ranking, narrative understanding)
- This type of social commonsense reasoning goes far beyond the widely studied entailment tasks (e.g., SNLI).



Collecting Inference Data

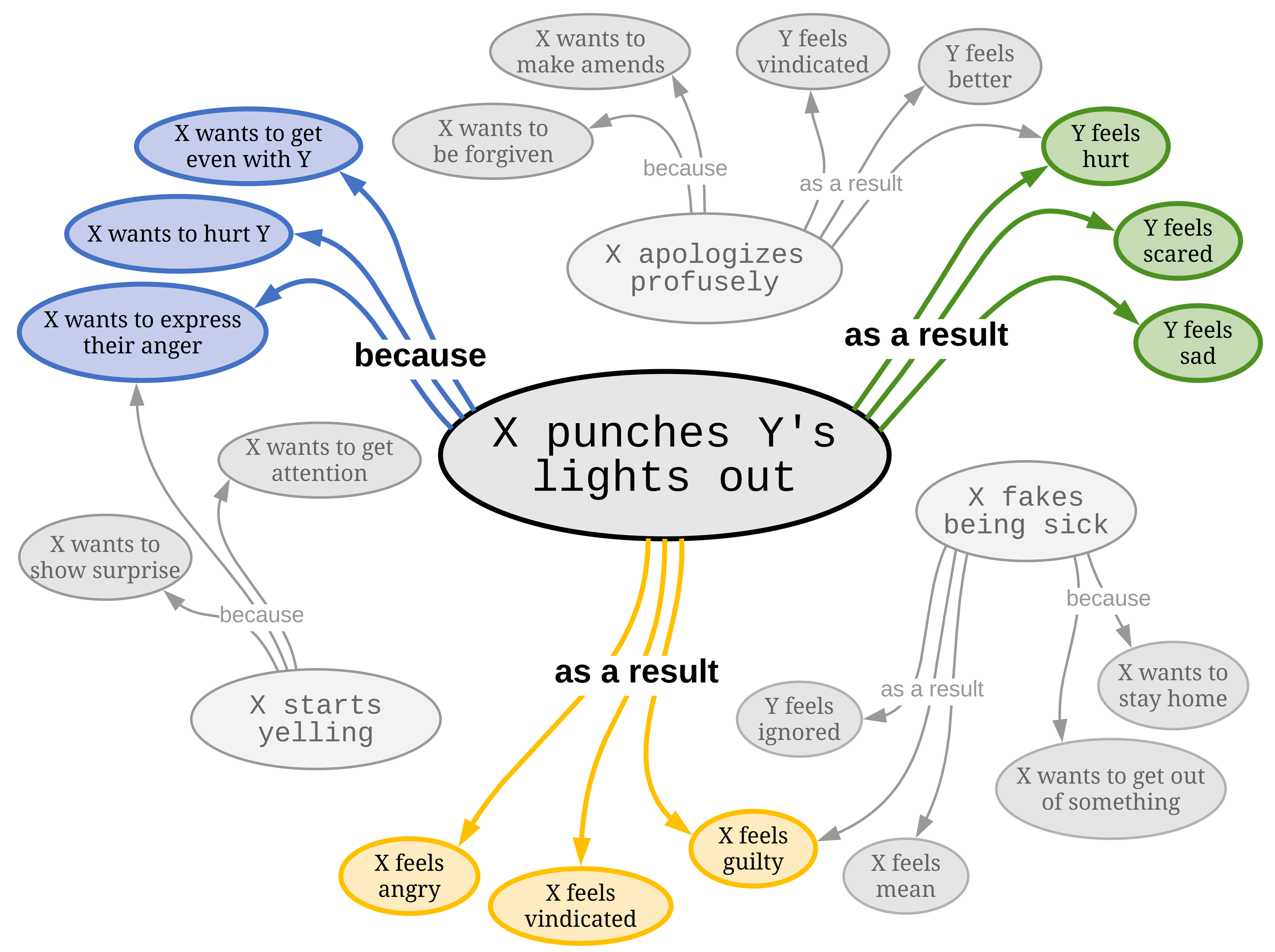
Annotated dimensions

PersonX's Intent	Why does X cause the event?
PersonX's Reaction	How does X feel after the event?
Others' Reaction	How do others feel after the event?

- Automatically extracted events from: ROC Stories, Google NGrams, Spinn3r, Wiktionary Idioms[1]
- Annotated by 3 Amazon MTurkers
- Moderate agreement $\kappa = .45$

[1] Mostafazadeh et al., 2016; Goldberg & Orwant, 2013; Gordon & Swanson, 2008

Commonsense Knowledge Graph



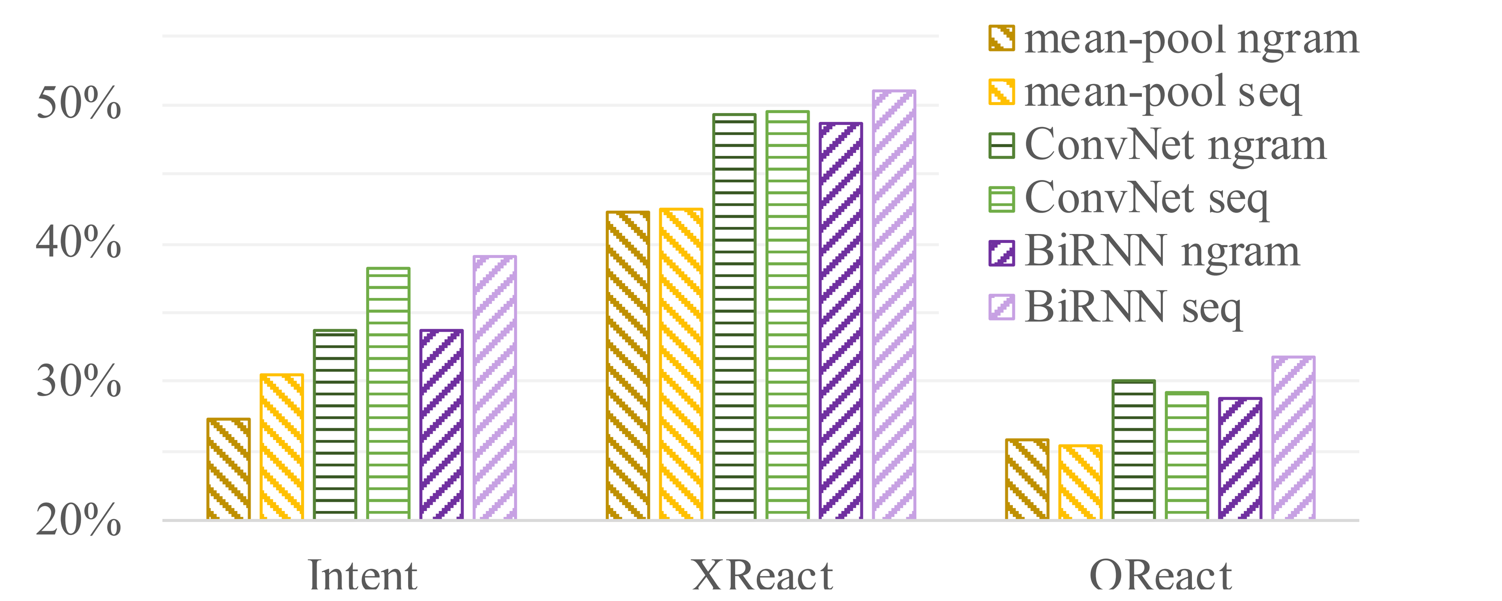
Knowledge graph statistics

# unique events	24,716
# unique intents	33,373
# unique reactions	19,083
total # nodes	**77,172**

Annotation statistics

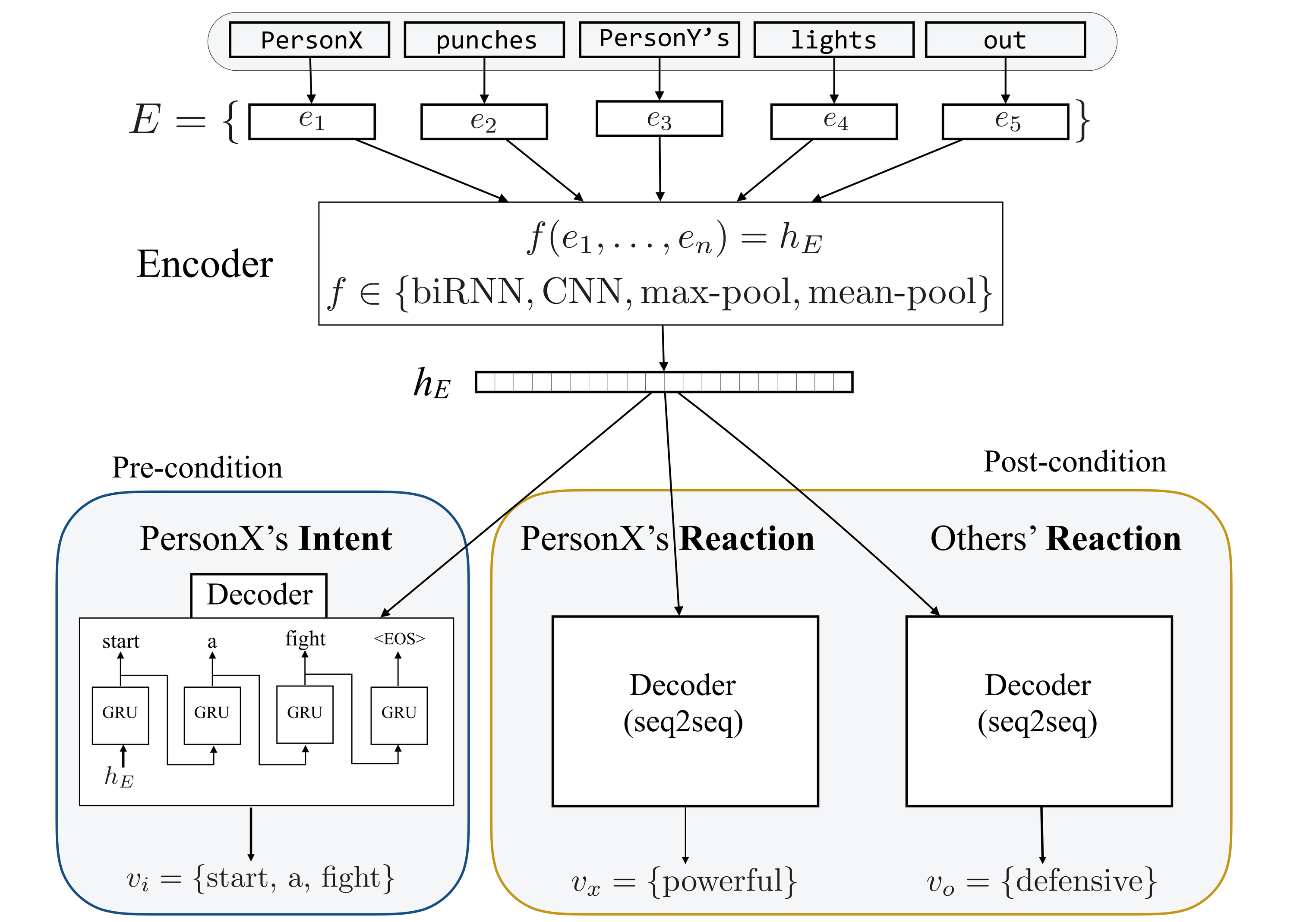
- Intents longer than reactions (3.4 vs. 1.5 words)
- 26% of events involve other participants
- 22% contain blanks

Evaluation of Generated Inference



- 5 MTurk raters evaluated precision@10 of generations
- Compositional encoders (ConvNet, BiRNN) > pooling
- Sequence generation is preferred over n-gram re-ranking

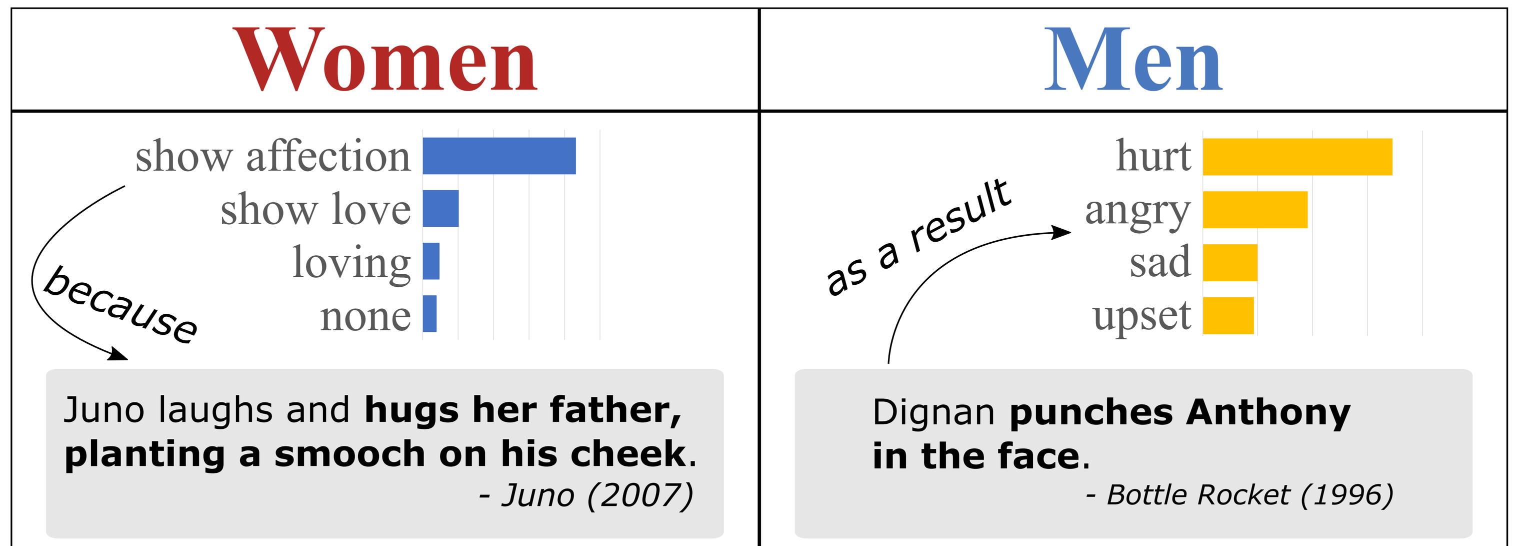
Neural Inference Model



- Encoder-decoder setup; given an event, generate intents and reactions in a multi-task way
- Two decoders: n-gram re-ranking and sequence generation

Gender Bias in Movies

- 21K characters from 772 modern english movies
- Parse narratives describing characters into events, generate intents & reactions (with BiRNN-seq model)
- Extract LIWC features from generated inference
- Logistic regression to find gender correlates



Women	Men
Intents: to be helpful, to be felt/seen	**Intents:** to accomplish, violence
Reactions: pro-social, sexual	**Reactions:** negative, achievement