A central challenge in developing a theory of how the mind processes language is how the meanings of ambiguous words are resolved. Approximately 85% of the words in the world's languages are ambiguous, yet humans resolve most ambiguities correctly and effortlessly (Klein and Murphy, 2001). For example, in the phrase "A prime minister wields significant power" a reader has no difficulty evoking a "political power" interpretation as opposed to an "electrical power" interpretation. However, how ambiguity is resolved remains open, despite having received attention from interdisciplinary researchers (including philosophers such as myself, as well as linguists, and psychologists), and despite the game-changing role that a theory of ambiguity resolution could have for artificial intelligence.

My proposal builds upon the latest insights from several disciplines to develop and test a novel account of ambiguity resolution. My philosophy background makes me appreciate the fundamental weakness of a cognitive system without grounded and referential information connected to the outer world (Searle, 1980). Meanwhile, from the psycholinguistics and neuroscience literatures (including work by my proposed advisor, Professor Blair Armstrong), I will draw on recent findings demonstrating that building in additional neurobiological plausibility into simulations are essential for simulating ambiguity resolution (Armstrong and Plaut, 2016).

RESEARCH PLAN: My research plan starts from a (relatively) simple neural network model and gradually increasing the complexity of this model to achieve better performance. It quantifies how each new level of complexity improves over simpler models. Such complexity is not achievable until the recent development in training methods and hardwares. These models will be trained on databases of images and their associated captions, such as the Common Objects in Context dataset, so that they will learn to see objects while simultaneously processing their verbal description. And each model will be tested against published data in the psycholinguistics literature which will serve as the "gold standard" for successfully modeling human ambiguity resolution abilities (Eddington and Tokowicz, 2015).

1. Language-only model. The first model will be trained only on the verbal caption in the COCO dataset, and provides a baseline for how much disambiguating information can be extracted from text alone (Beekhuizen et al., 2018).

2. Integrated vision-and-language model. The second model will add a visual processing system to the model developed in (1), and will serve as a first validation for how visual information can provide additional constraint on word meaning than the prior model.

3. More biologically plausible integrated model. The third model will focus on how several key principles from cognitive and systems neuroscience could further enhance performance in the model: 3(a) Addition of an attention mechanism. Humans do not process all of the information their eyes take in equally well, rather, they focus attention on one part of the visual field. 3(b) More realistic cortical connectivity. In human neural networks, neurons send excitatory OR inhibitory information, NOT both as in the artificial neural networks; and there are fewer connections between different brain regions compared to artificial networks. Prior work (Laszlo and Armstrong, 2014; Armstrong and Plaut, 2016) has established that these principles critically shape what type of information shapes network performance, including word disambiguation.

EXPECTED OUTCOME AND IMPLICATIONS: I will revolutionize how ambiguity is thought about in many fields by merging insights from several fields to develop an account that would not emerge from any one discipline in isolation. The understanding of how ambi-

guity is resolved will have a huge impact on studying how language interacts with contexts, both the narrower linguistic and broader social-cultural. Finally, the project can be potentially patented, if successful, to allow its free use to academics as well as its incorporation into industry, like Microsoft, Facebook, Google, etc, including building better "bots" to communicate with customers online, and better automated screening of user content for remarks that are inappropriate or discriminatory.
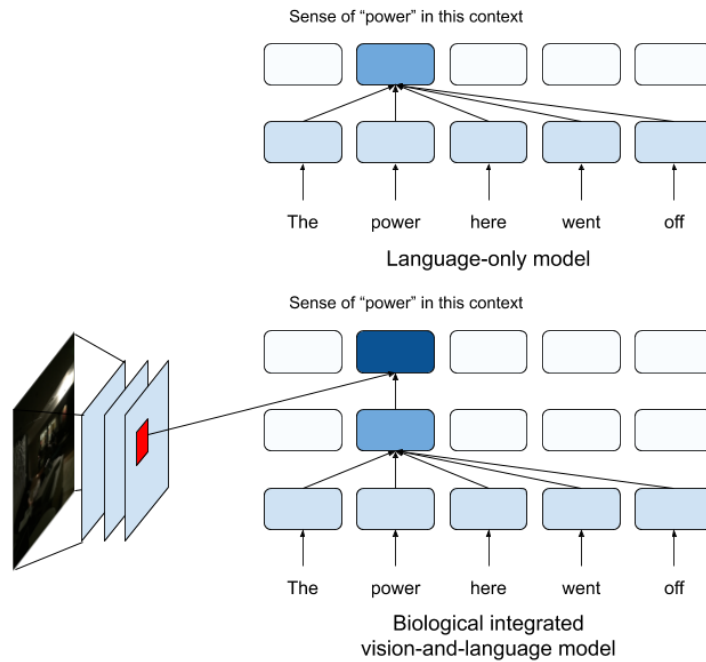

Figure 1: Computational models that are going to be build in this project.

References

Armstrong, B. C. and Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, 31(7):940–966.

Beekhuizen, B., Milic, S., Armstrong, B. C., and Stevenson, S. (2018). What Company Do Semantically Ambiguous Words Keep? Insights from Distributional Word Vectors. In *CogSci*.

Eddington, C. M. and Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: The current state of the literature. *Psychonomic bulletin & review*, 22(1):13–37.

Klein, D. E. and Murphy, G. L. (2001). The Representation of Polysemous Words. *Journal of Memory and Language*, 45(2):259–282.

Laszlo, S. and Armstrong, B. C. (2014). PSPs and ERPs: Applying the dynamics of post-synaptic potentials to individual units in simulation of temporally extended Event-Related Potential reading data. *Brain and Language*, 132:22–27.

Searle, J. R. (1980). Minds, brains, and programs. *& &*, pages 417–457.