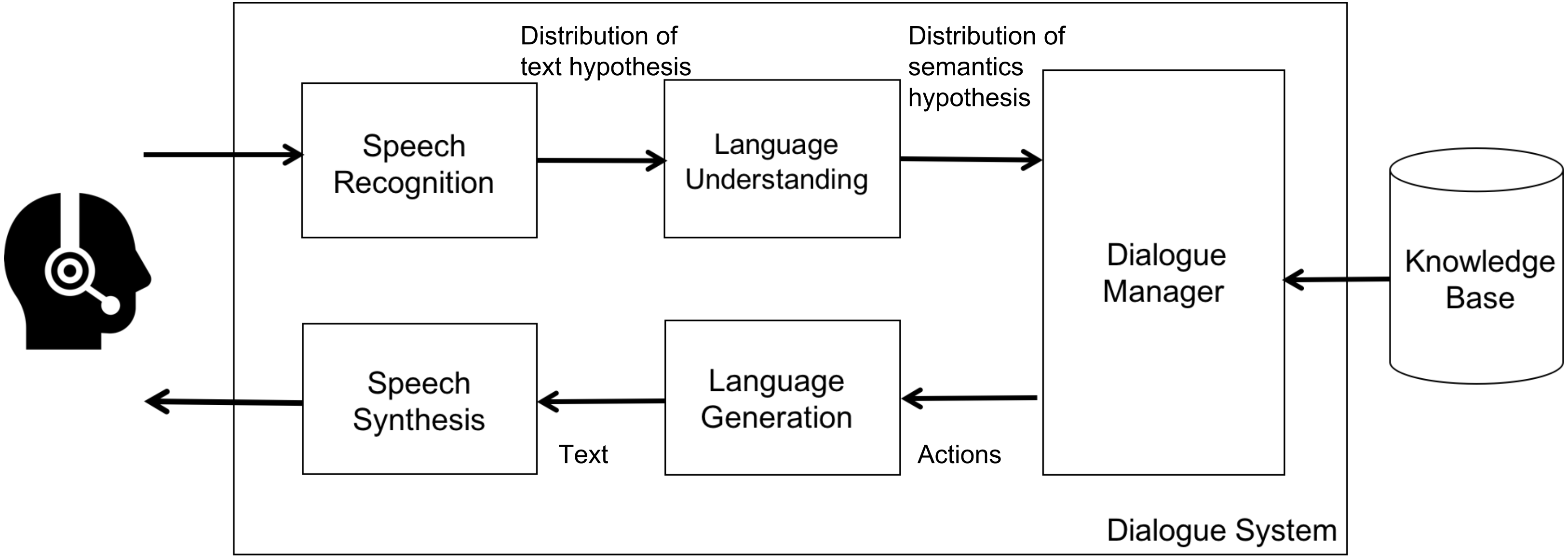
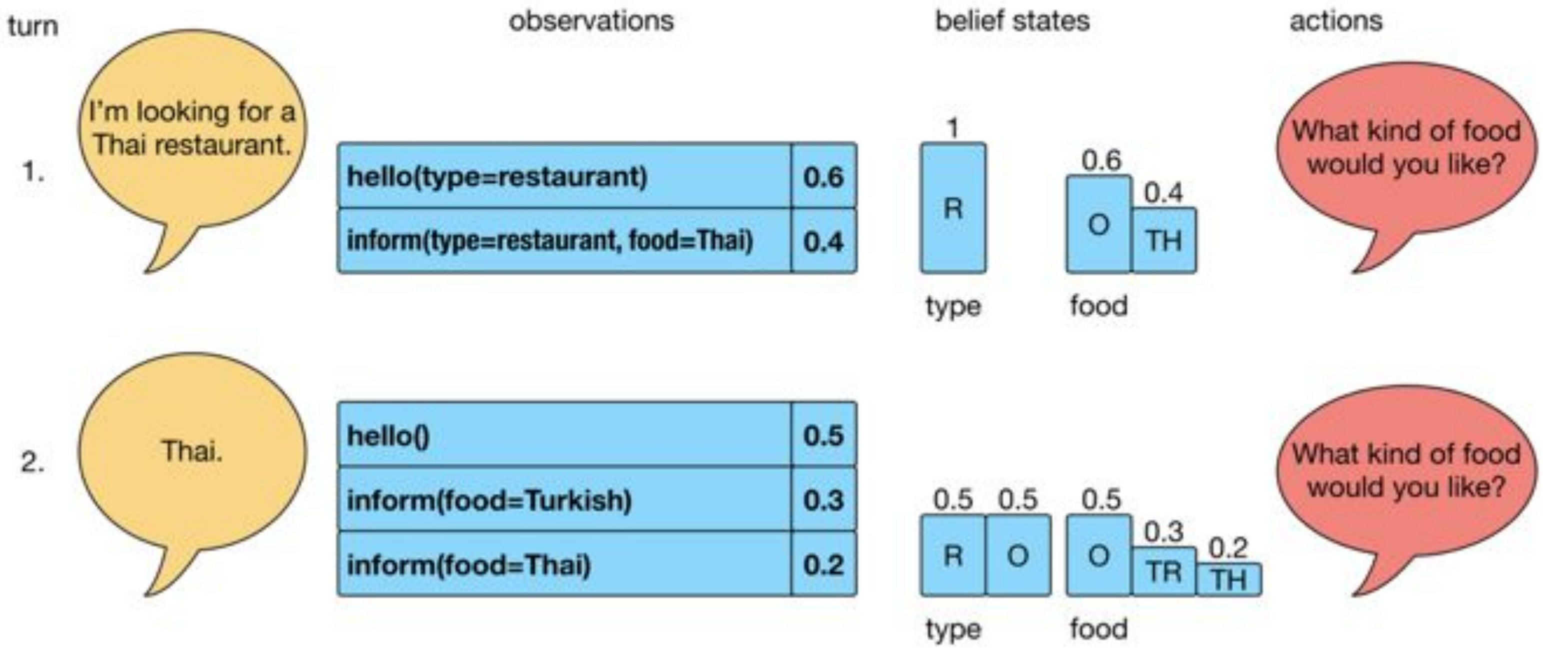
# Benchmarking Uncertainty Estimates with Deep Reinforcement Learning for Dialogue Policy Optimisation

Christopher Tegho*, Pawel Budzianowski*, Milica Gasic
Dialogue Systems Group, Department of Engineering,
Cambridge University

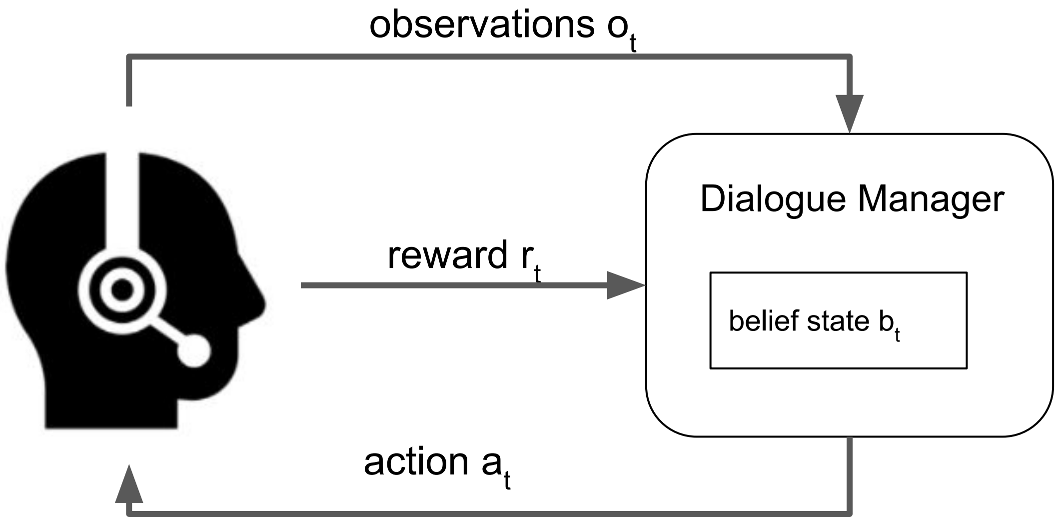
# Statistical Dialogue Management Architecture



# No Belief State Tracking



# Reinforcement Learning



# Reinforcement Learning

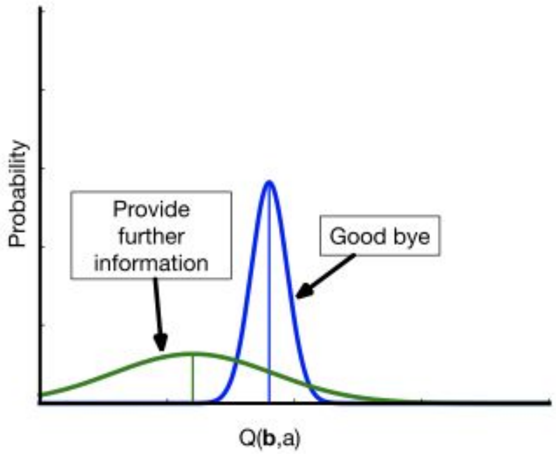
We aim at maximizing a reward obtained along the dialogue:

$$R = \sum_{n=1}^{T} \gamma^t r_t$$

by modelling Q-value function:

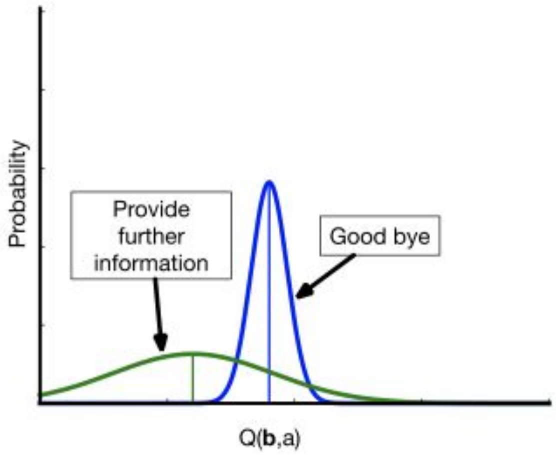
$$Q^\pi(b, a) = \mathbb{E}_\pi\{r_t + \gamma r_{t+1} + \ldots \ \ |b_t = b, a_t = a\}$$

# Uncertainty Estimates



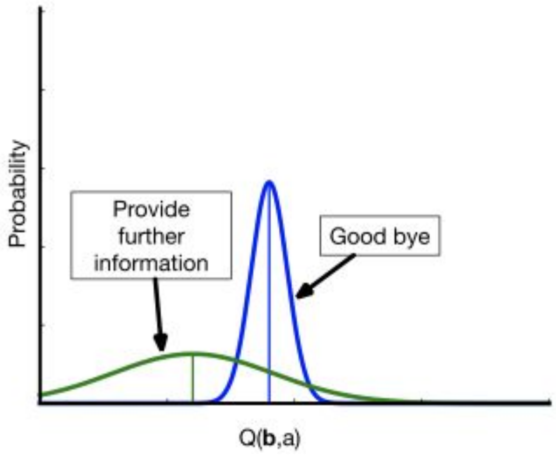

1. We know how sure the agent is about the action

# Uncertainty Estimates




1. We know how sure the agent is about the action
2. We can introduce clever exploration than epsilon-greedy
3. Choosing next action through Thomson Sampling

# Uncertainty Estimates




1. We know how sure the agent is about the action
2. We can introduce clever exploration than epsilon-greedy
3. Choosing next action through Thomson Sampling
4. Faster learning, better user experience

# Uncertainty Estimates in Neural Networks

- GP SARSA provides an **explicit estimate of uncertainty**, however, the computational complexity is **cubical**.

# Uncertainty Estimates in Neural Networks

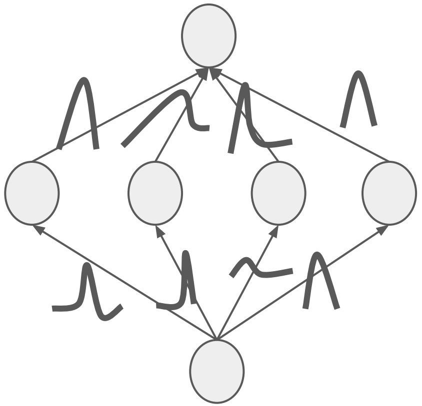
- GP SARSA provides an **explicit estimate of uncertainty**, however,  the computational complexity is **cubical**.
- Deep neural network models **scale nicely** with data, but do not provide **an explicit estimate of uncertainty**

# Uncertainty Estimates in Neural Networks

- GP SARSA provides an **explicit estimate of uncertainty**, however, the computational complexity is **cubical**.
- Deep neural network models **scale nicely** with data, but do not provide **an explicit estimate of uncertainty**
- Uncertainty estimates with NN can be obtained **by approximation**
- Number of approached explored - 4 casted in the **variational inference** framework

# Bayes By Backprop

- All weights are represented by probability distributions over possible values given observed dialogues
- We use sampling-based variational inference. The intractable posterior is approximated with variational posterior:
- Loss to minimize:
  KL( posterior | prior) - log likelihood of data



Sources:

- Blundell, Charles, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. "Weight uncertainty in neural networks." (2015).
- Lipton, Zachary, Xiujun Li, Jianfeng Gao, Lihong Li, Faisal Ahmed, and Li Deng. "BBQ-Networks: Efficient Exploration in Deep Reinforcement Learning for Task-Oriented Dialogue Systems."

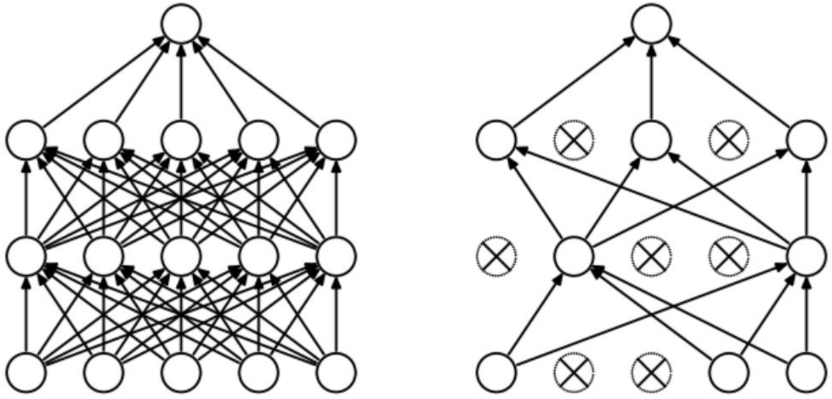
# Uncertainty Estimates in NN

- **Alpha-divergence:** The α-divergence measures the similarity between two distributions.
- It's a generalization over KL divergence
- Bayes By Backrop uses KL divergence, equivalent to Alpha-divergence with alpha = 0

Source: Hernández-Lobato, José Miguel, Yingzhen Li, Mark Rowland, Daniel Hernández-Lobato, Thang Bui, and Richard Eric Turner. "Black-box α-divergence minimization." (2016).

# Uncertainty Estimates in NN

- **Dropout**: Multiply the weight matrix in a given layer by some random noise.
- **Concrete dropout**: Continuously relax the dropout’s discrete masks and optimize the dropout probability using gradient methods.

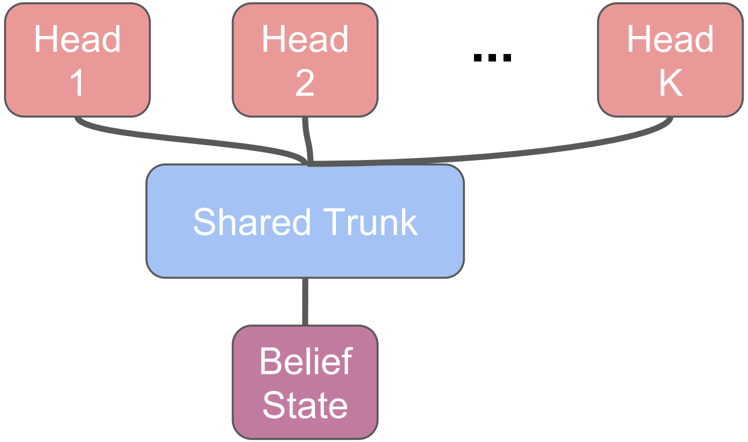


(a) Standard Neural Net (b) After applying dropout.

Source: Srivastava, Nitish, et al. ”Dropout: a simple way to prevent neural networks from overfitting”, JMLR 2014

# Uncertainty Estimates in NN

**Bootstrapped DQN:** Several neural networks are randomly initialized which predict in ensemble uncertainty estimates.

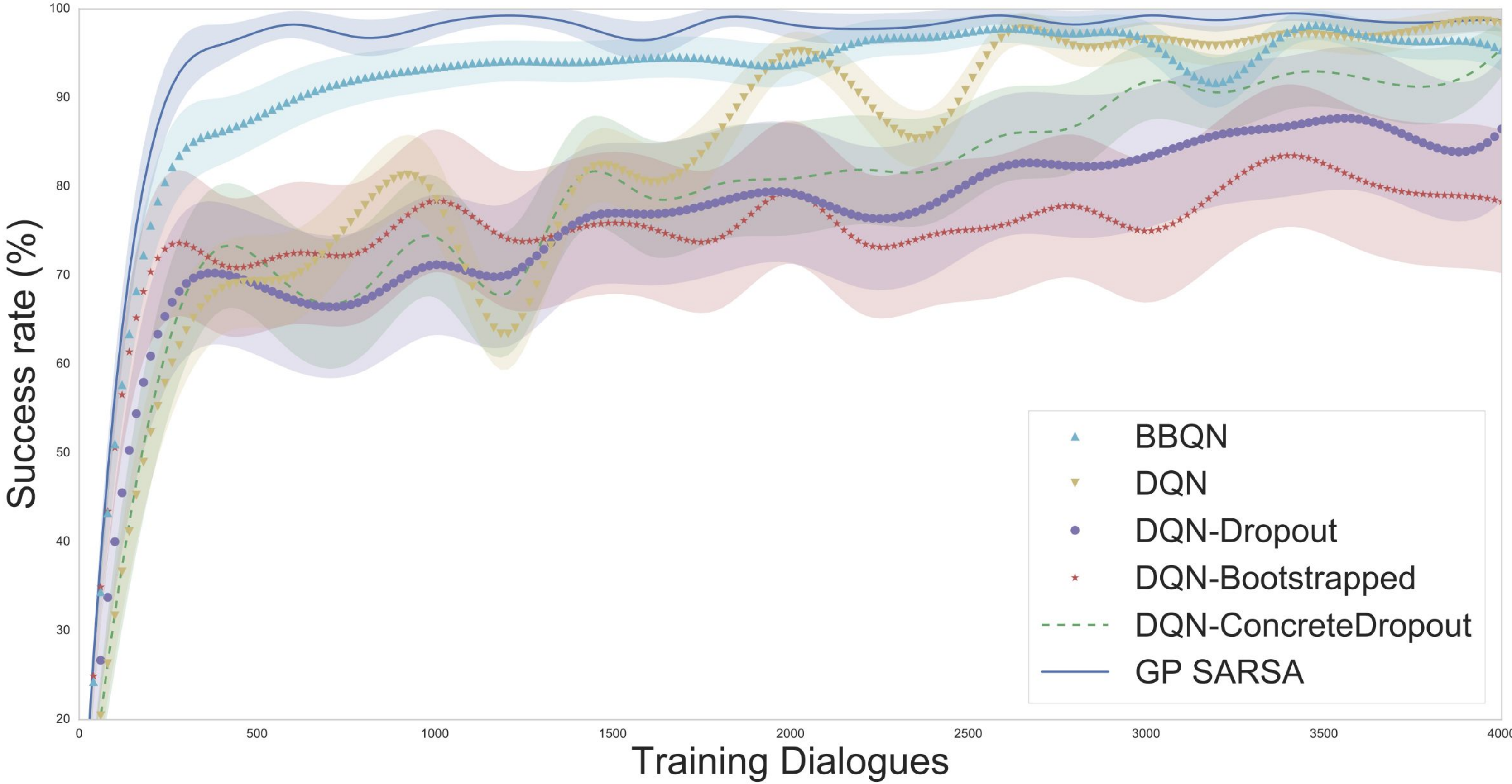



Source: Osband, Ian, Charles Blundell, Alexander Pritzel, and Benjamin Van Roy. "Deep exploration via bootstrapped DQN." In Advances in neural information processing systems, pp. 4026-4034. 2016.

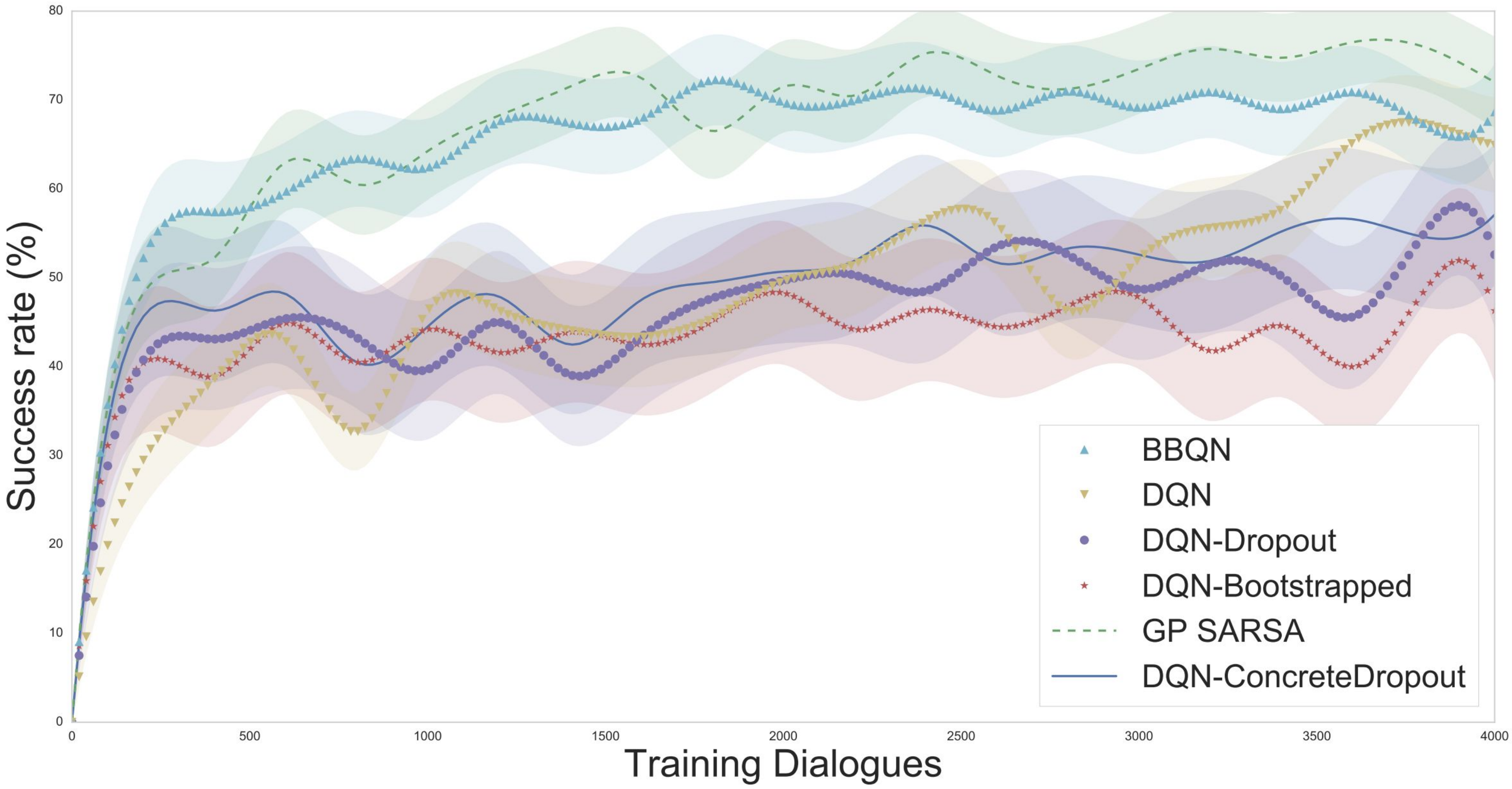
# Evaluation Setup

- Cambridge restaurant domain: 100 venues, 6 slots, 3 requestable
- Belief state input of size 268
  (last system act, distribution over user intent …)
- System summary action space of size 14 (inform, request, confirm, …)
- User simulator operating on semantic level
- Capable of simulating noise

# Results - Environment without any noise



# Results - Environment with noise



# Conclusion

We train a dialogue agent using reinforcement learning paradigm.

Vanilla Deep-RL methods proved to be unstable and sample inefficient.

# Conclusion

We train a dialogue agent using reinforcement learning paradigm.

Vanilla Deep-RL methods proved to be unstable and sample inefficient.

We tested 5 different approaches to introduce uncertainty estimates into Deep-RL agent.

BBQN achieves comparable performance to GPSARSA, especially in more noisy environments, without the cubic computational complexity.