真ん中の文字は何に見えますか？

(c) Hideyoshi Yanagisawa, The University of Tokyo

講義の内容

1. 状態と観測
2. 生成モデル：事前信念、尤度
3. ベイズ推論：因果反転、信念更新、バイアス、精度
4. エビデンス、自由エネルギー、予測誤差
5. KLダイバージェンス、サプライズ、近似事後分布
6. 自由エネルギー原理
7. ベイズサプライズと正確さ
8. 感情と行動
9. まとめ

まとめ　—　人の認識の数理モデル

- ベイズ推論にもとづく認識モデル
  - 観測と状態の因果の反転
  - 認知バイアス＝文脈における期待（事前分布）の影響
  - 信念のデータによる更新（ベイズ更新）

- 自由エネルギー原理
  - 自由エネルギー＝予測誤差≧サプライズ
  - 認識、行動、学習は自由エネルギーの最小化として説明
  - 不確実性（エントロピー）の上昇を解消する＝生存

Generative model

$$p(s)p(o|s) = p(s,o)$$
$$= p(s|o)p(o)$$

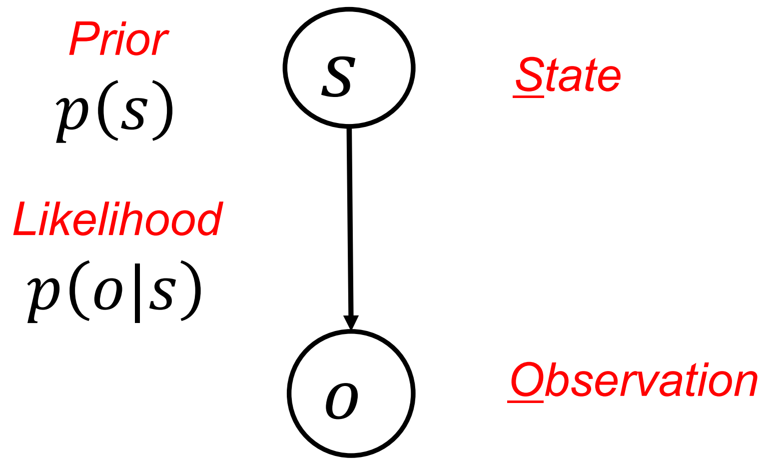
Bayes' theorem

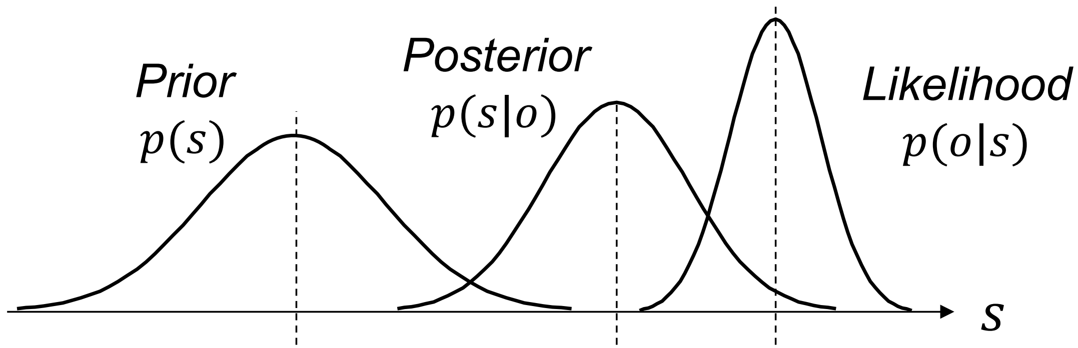
$$p(s|o) = \frac{p(s)p(o|s)}{p(o)}$$

$$\underline{p(s|o)} \propto \underline{p(s)}\,\underline{p(o|s)}$$

Update

Reversal



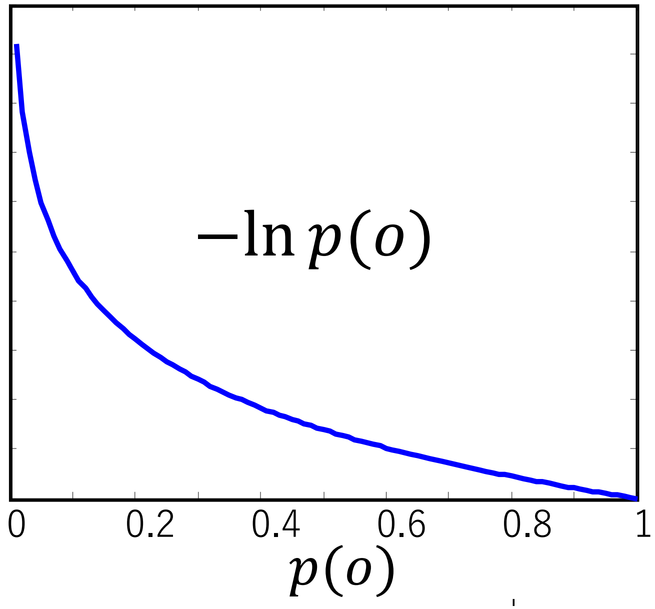


(c) Hideyoshi Yanagisawa, The University of Tokyo

$$p(s|o) = p(s)p(o|s)/p(o)$$

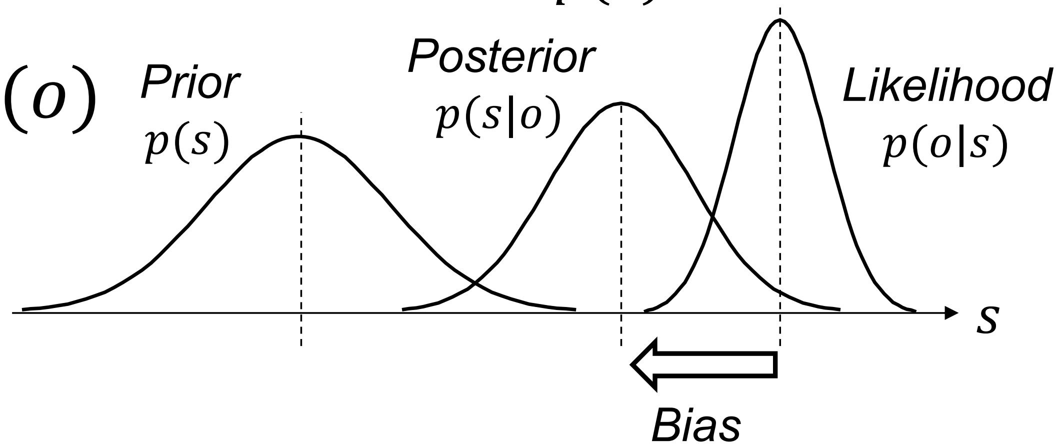
Evidence: $p(o) = \sum_s p(s)p(o|s)$

Surprise: $-\ln p(o)$



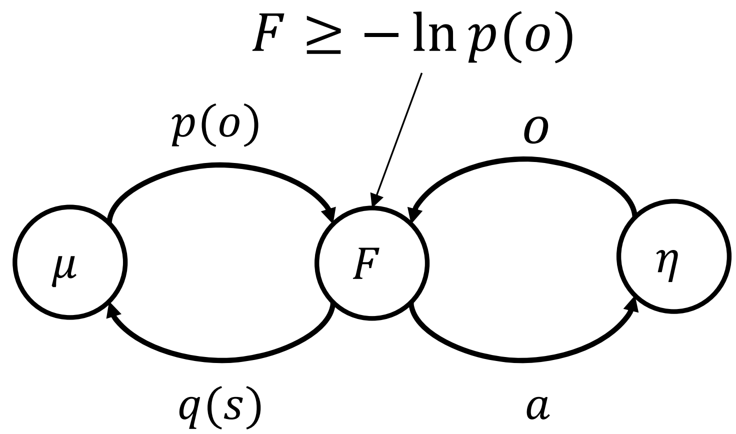
Free energy

$$F = \langle \ln q(s) - \ln p(s,o) \rangle_{q(s)}$$
$$= \langle \ln q(s) - \ln p(s|o) \rangle_{q(s)} - \ln p(o)$$
$$= D_{KL}[q(s)||p(s|o)] - \ln p(o)$$
$$q(s) \approx p(s|o), F \approx -\ln p(o)$$



Free energy principle

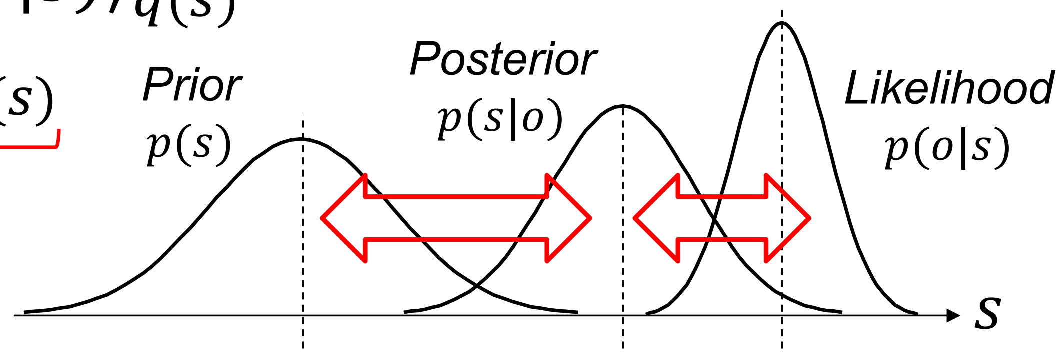
$$\mu^* = \operatorname{argmin} F(\mu, a; o)$$
$$a^* = \operatorname{argmin} F(\mu^*, a; o)$$



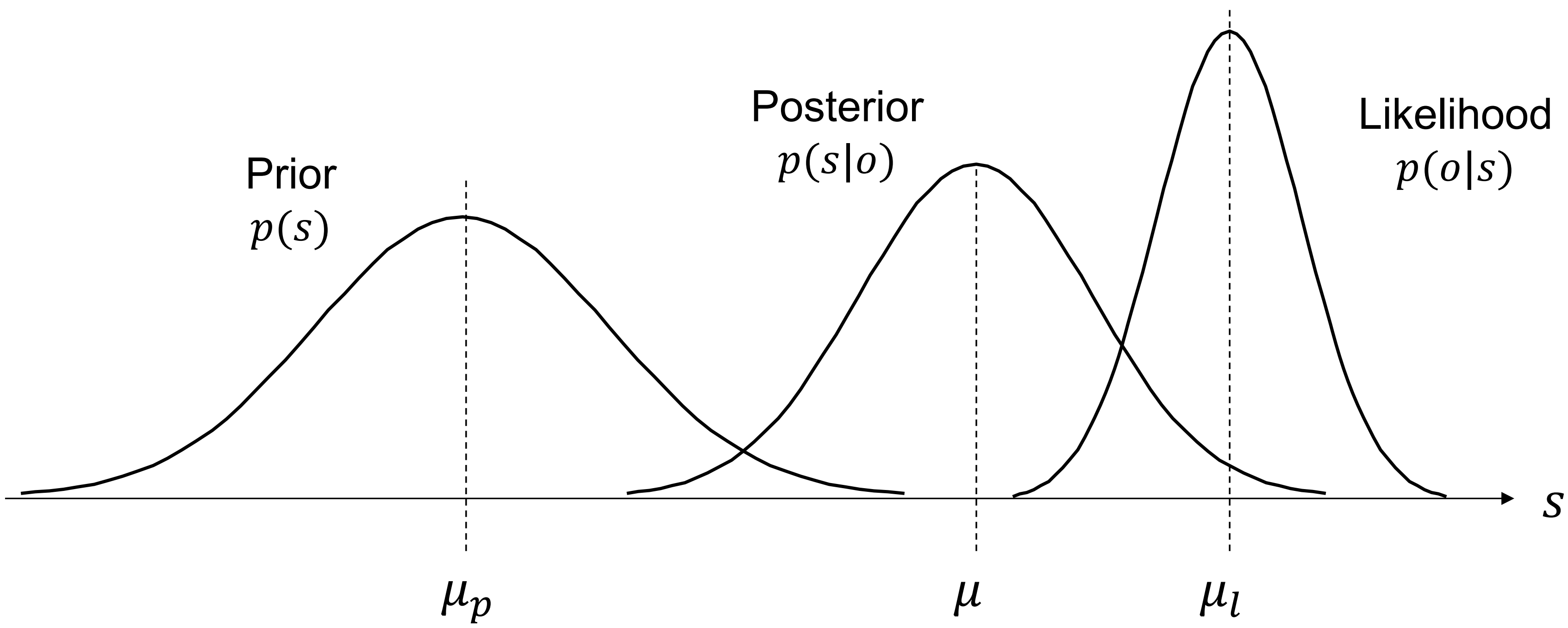
$$F = \langle \ln q(s) - \ln p(s,o) \rangle_{q(s)}$$
$$= \langle \ln q(s) - \ln p(s) \rangle_{q(s)} - \langle \ln p(o|s) \rangle_{q(s)}$$
$$= \underbrace{D_{KL}[q(s)||p(s)]}_{\text{Bayesian surprise}} - \underbrace{\langle \ln p(o|s) \rangle_{q(s)}}_{\text{Accuracy}}$$

Information gain

Epistemic value

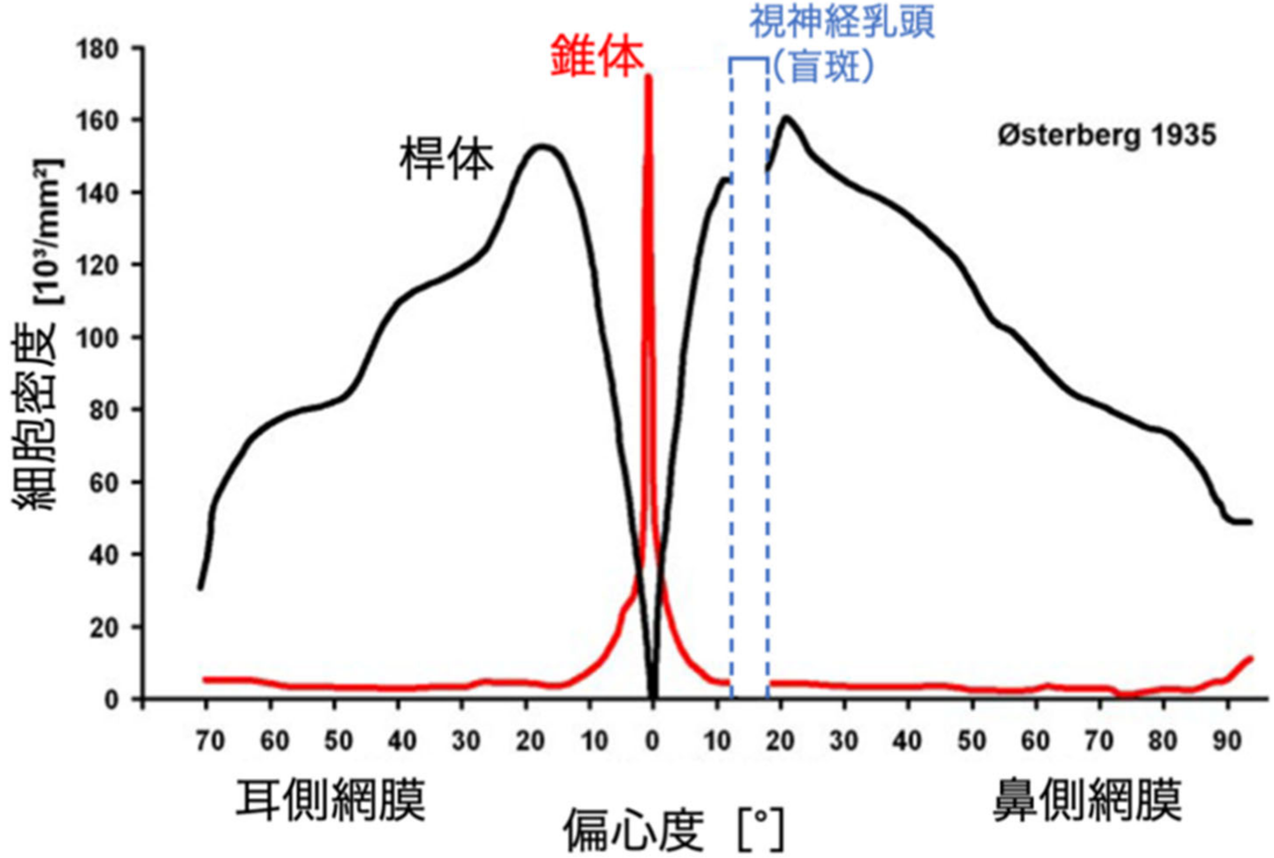


(c) Hideyoshi Yanagisawa, The University of Tokyo


Prior
$p(s)$
Posterior
$p(s|o)$
Likelihood
$p(o|s)$
s
μ_p
μ
μ_l

「見る」とはどういうことか？

色の見える錐体細胞は中心視から10度以内…



(c) Hideyoshi Yanagisawa, The University of Tokyo