

Understanding the Behaviour of Complex Cellular Processes

Systems Biology Meets Bayesian Inference

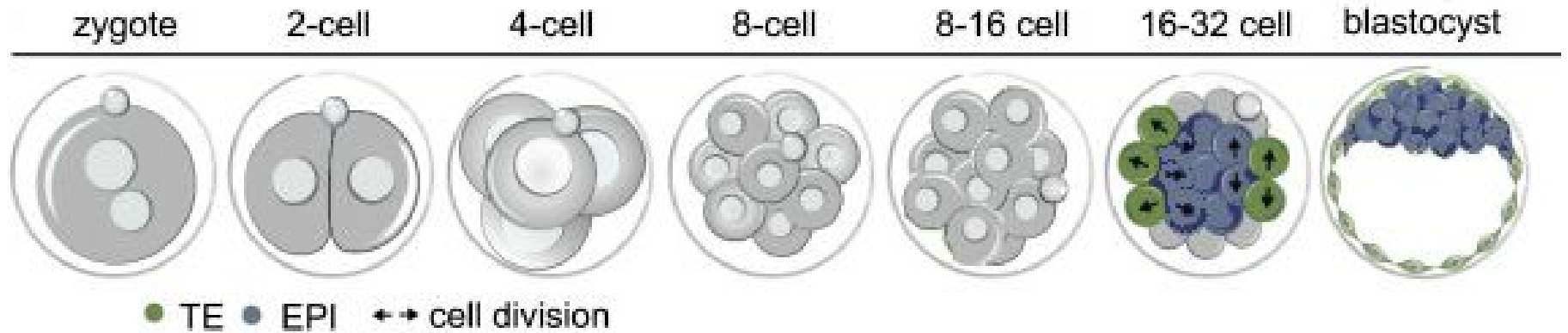
Giles Harper-Donnelly

Department of Genetics, University of Cambridge.

Bayes on the Beach, 13th Nov, 2017

Early Mammalian Development

Fertilisation to Blastocyst.



Adapted from Goulam et al.; Cell, 2016.

Blastocyst

- Formation begins 5 days after fertilization in humans.
- 200-300 cells
- 2x cell types:
 - Inner cell mass
 - Trophoblast

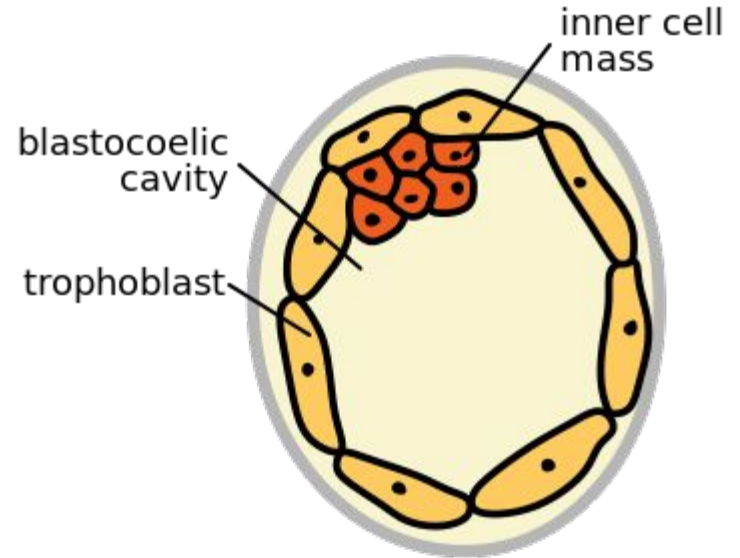
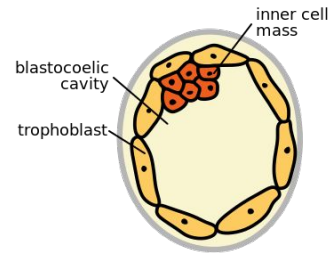
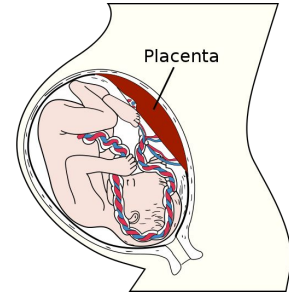
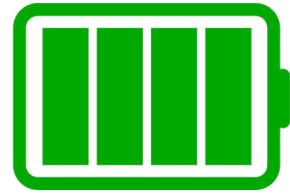
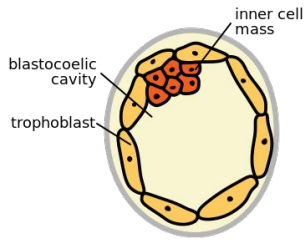
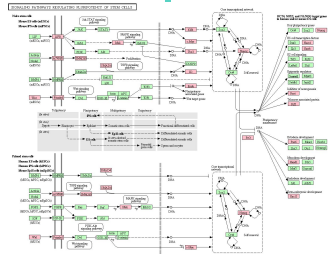
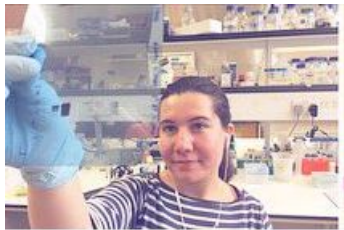


Image credit, Richard Wheeler.

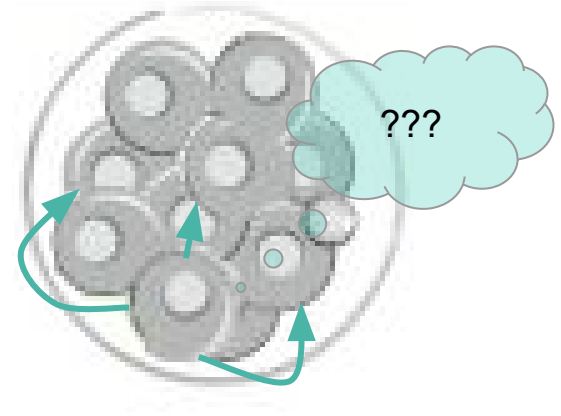
Bottom Up



Top Down

Cells adapt to their environment

- Mammalian development more regulative than deterministic.
- Organism faces simultaneous challenges:
 - Produce a huge variety of cell types in the **correct proportions**.
 - Ensure that cells are **distributed appropriately** in space.
- These challenges require individual cells to **interactively adapt their behaviour** in response to the state of their environment.



Cells need to do inference!

Structure of the Cell

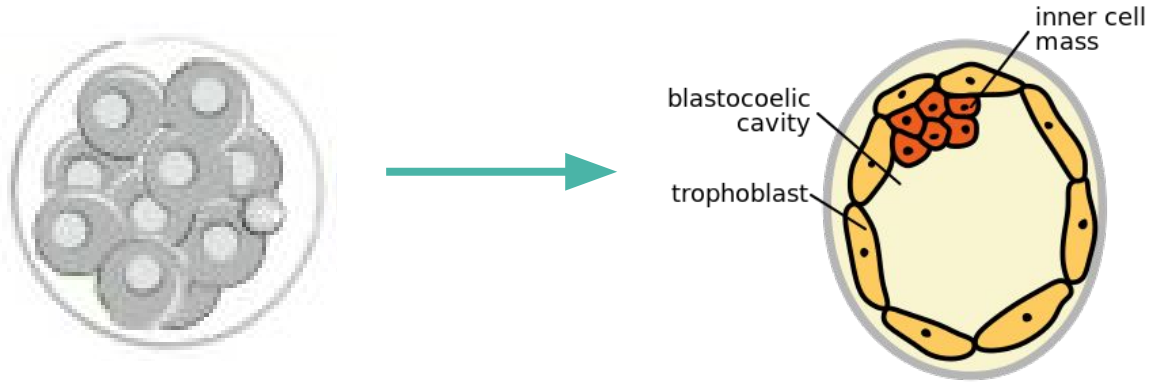
Internal workings of the cell do not have direct access to the external world.

External causes can only be experienced vicariously via the effect they have on a some small subset of cell.

e.g. Signalling/Mechano-receptors.

This gives structure to our inference problem, whereby the activity of some set of hidden states (missing data) must be inferred from data available from visible states (observed data).

Example: Patterning in the Blastocyst.

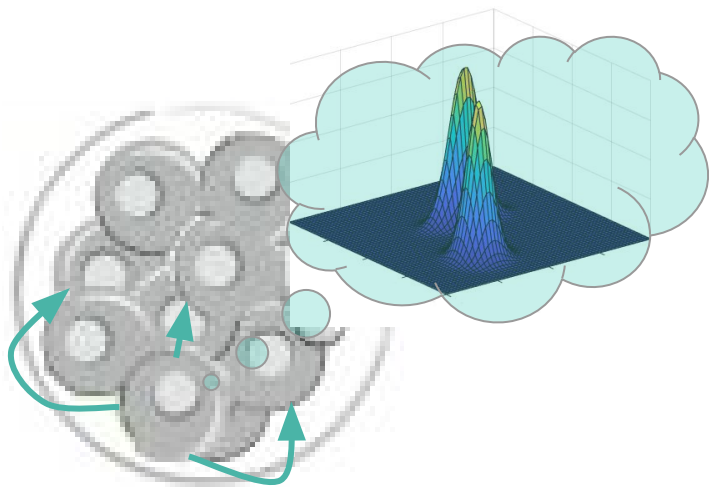


Example: Patterning by maximising model evidence.

Cells equipped with a generative model of its environment.

This model gives **high probability to observed data consistent with the right numbers of cells being in each of two states** (either internal cells or external cells).

Cells then **act to move to states with maximum probability under the model**.



Observed Data

Unobserved Data

$$P(D|m) = \int P(D|x, m)P(x|m)dx$$

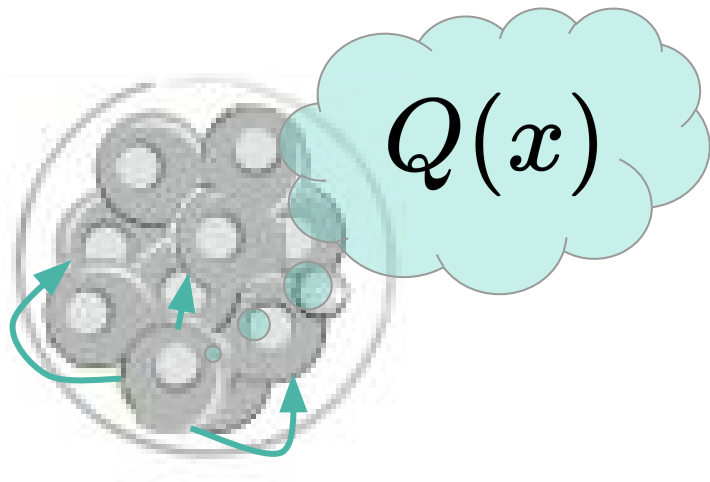
Example: Patterning through minimisation of a Variational Free Energy

Calculating the marginal likelihood is hard.

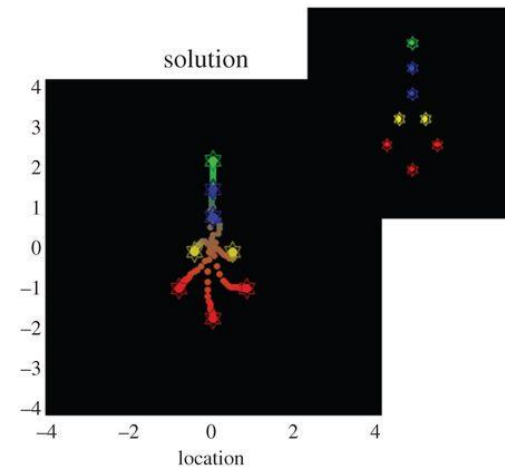
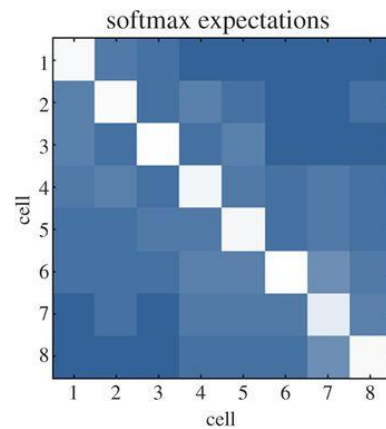
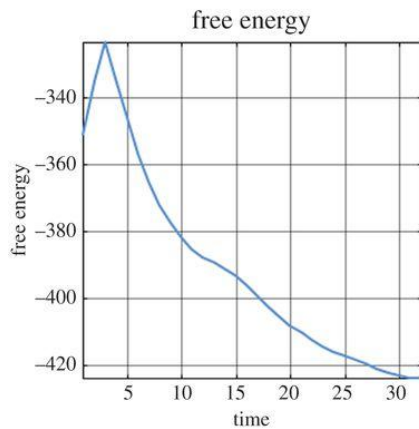
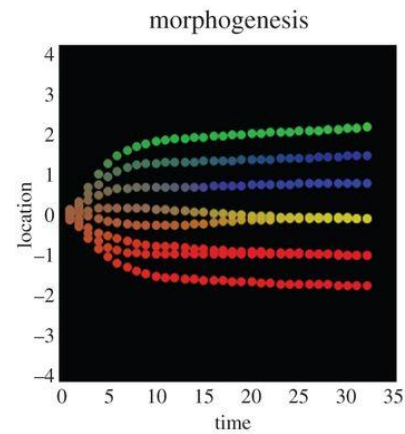
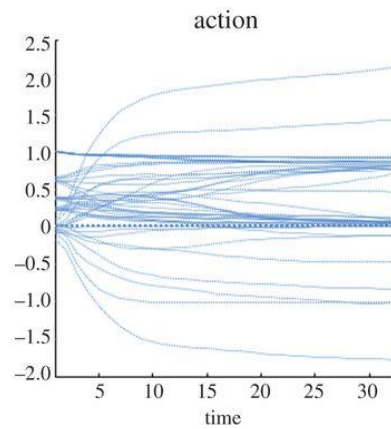
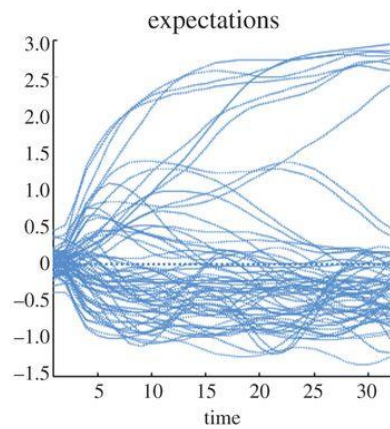
Instead minimise a lower bound on the negative log evidence.

This involves a variational distribution parameterised by internal components of the cell.

Thus cells implicitly perform a type of variational inference of their state of their environment.



$$F = -\ln P(D|m) + D_{KL} [Q(x) || P(x|D, m)]$$



Conclusions

1. There is value in **both** 'Bottom-up' and 'Top-down' approaches.
2. We can gain useful insights by modelling cells as agents faced with an inference (and decision) problem.
3. Great opportunity in investigating how this inference could be implemented by the machinery of the cell. For example through a Variational Bayes approach.



Acknowledgements

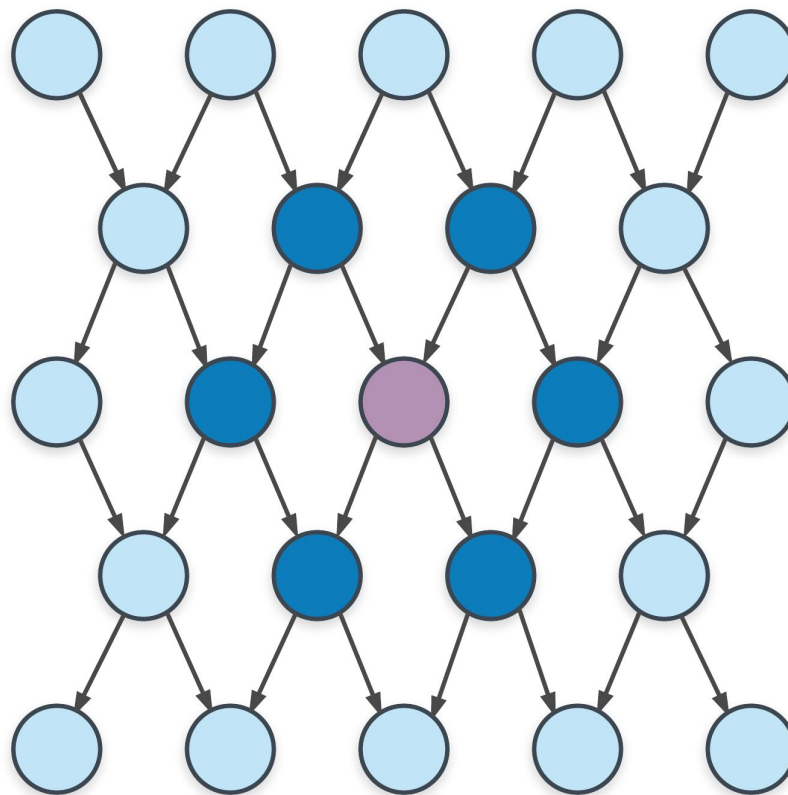
Martinez-Arias Group.

KAUST Studentship funding.



Department of Genetics, Cambridge.

Markov Structure of the Cell



Markov Structure of the Cell

