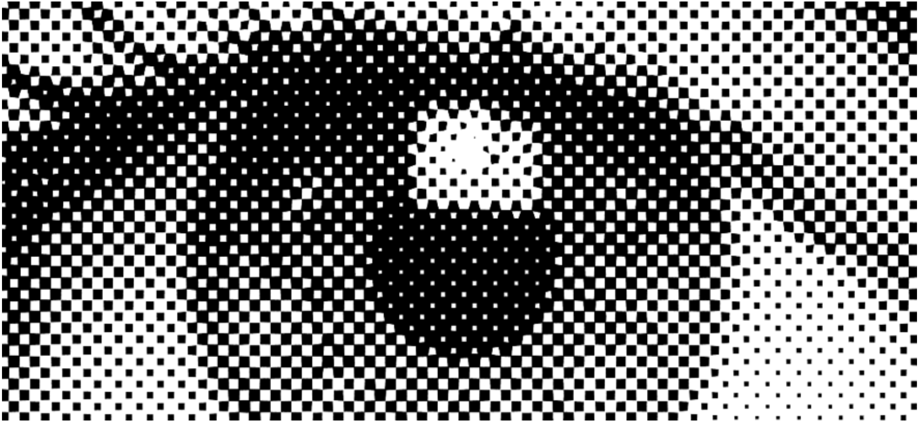
Bayesian Reverse-Engineering of Perception and Cognition

Frank Jäkel
jaekel@psychologie.tu-darmstadt.de

TECHNISCHE UNIVERSITÄT DARMSTADT



Categorization Example: What is an Ideal Observer?

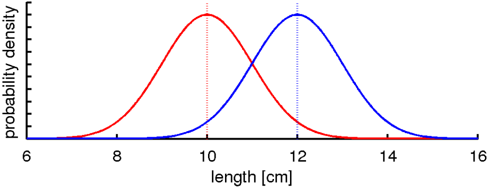
TECHNISCHE UNIVERSITÄT DARMSTADT

Task

White bars of varying length are presented on an otherwise black screen. The bars come from two two equally probable categories: The short and the long bars, named category 1 and category 2 respectively.

1. Stimuli from category 1 are drawn from a normal distributions with mean 10 cm and standard deviation 1 cm.
2. Stimuli from category 2 are drawn from a normal distributions with mean 12 cm and standard deviation 1 cm.

Categorization Example: Task



Categorization Example: Ideal Observer

Task

1. $p(X = x \mid C = 1) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_1)^2\right)$
2. $p(X = x \mid C = 2) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_2)^2\right)$
3. $p(C = 1) = p(C = 2) = \frac{1}{2}$

Question

How would an ideal subject solve the task? What is the best strategy?

Categorization Example: Ideal Observer

TECHNISCHE UNIVERSITÄT DARMSTADT

Task

1. $p(X = x \mid C = 1) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_1)^2\right)$
2. $p(X = x \mid C = 2) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_2)^2\right)$
3. $p(C = 1) = p(C = 2) = \frac{1}{2}$

Question

How would an ideal subject solve the task? What is the best strategy?

Categorization Example: Bayesian Posterior

Task

1. $p(X = x \mid C = 1) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_1)^2\right)$
2. $p(X = x \mid C = 2) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_2)^2\right)$
3. $p(C = 1) = p(C = 2) = \frac{1}{2}$

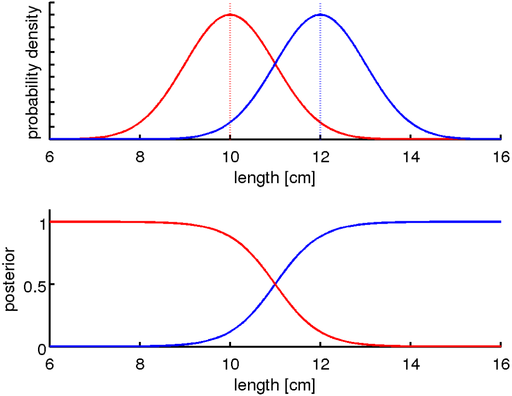
$$p(C = 1 \mid X = x) = \frac{p(X = x \mid C = 1)p(C = 1)}{p(X = x \mid C = 1)p(C = 1) + p(X = x \mid C = 2)p(C = 2)}$$

Categorization Example: Bayesian Posterior

$$p(C = 1 \mid X = x) = \frac{1}{1 + \exp\left(-\frac{1}{2\sigma^2}\left((x - \mu_2)^2 - (x - \mu_1)^2\right)\right)}$$

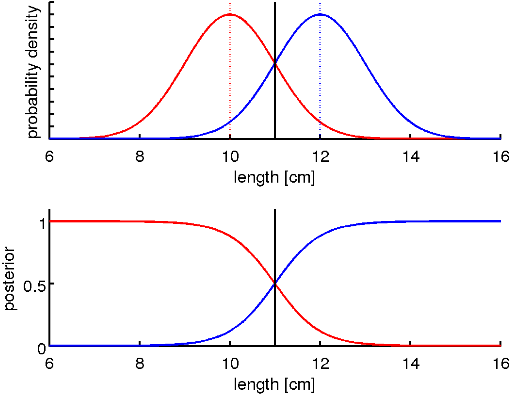
$$p(C = 2 \mid X = x) = 1 - p(C = 1 \mid X = x)$$

Categorization Example: Bayesian Posterior

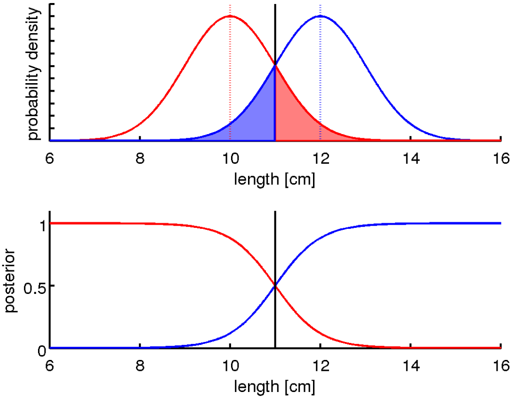


Categorization Example: Decision Rule

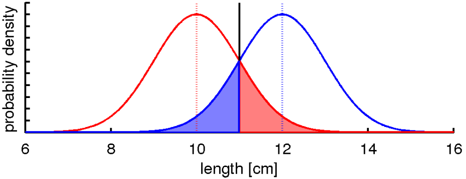
TECHNISCHE UNIVERSITÄT DARMSTADT

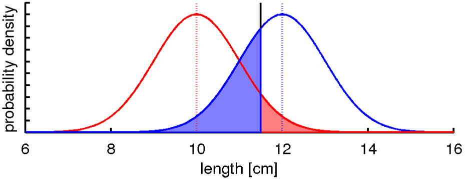


Categorization Example: Minimum Errors



Categorization Example: Minimum Errors





Categorization Example: Ideal Observer

TECHNISCHE UNIVERSITÄT DARMSTADT

Optimal Response Strategy for Categorization Task

Respond "1" if and only if bar length $x \leq 11$cm. Otherwise respond "2". This strategy will minimize the expected number of errors.

Note

The ideal observer for this categorization task does not compute probabilities, does not use Bayes' rule and it does not optimize. It only checks the stimulus against a decision criterion. Probability theory was merely used as a tool to derive the optimal strategy.

Categorization Example: Ideal Observer

TECHNISCHE UNIVERSITÄT DARMSTADT

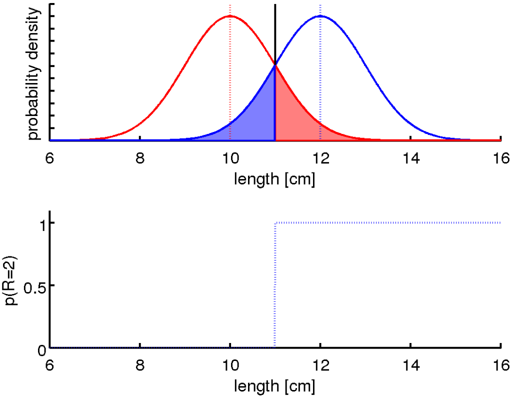
Optimal Response Strategy for Categorization Task

Respond "1" if and only if bar length $x \leq 11\text{cm}$. Otherwise respond "2". This strategy will minimize the expected number of errors.

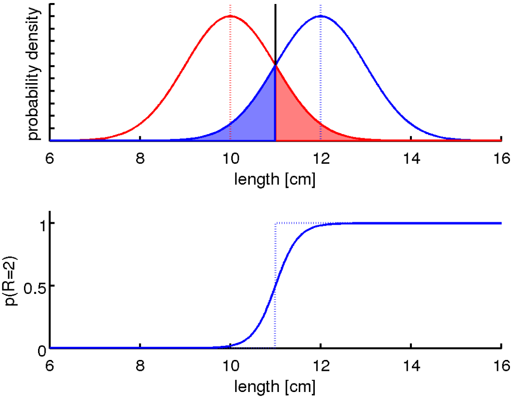
Note

The ideal observer for this categorization task does not compute probabilities, does not use Bayes' rule and it does not optimize. It only checks the stimulus against a decision criterion. Probability theory was merely used as a tool to derive the optimal strategy.

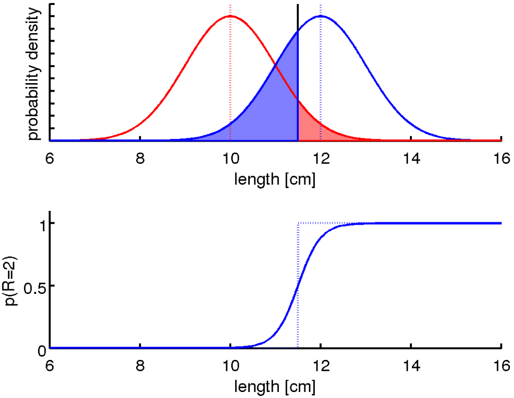
Categorization Example: Ideal Responses



Categorization Example: Actual Responses



Categorization Example: Actual Responses



Category Learning as Ideal Learning

TECHNISCHE UNIVERSITÄT DARMSTADT

Task

1. $p(X_k = x_k \mid C_k = 1, \mu_1, \sigma) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x_k - \mu_1)^2\right)$
2. $p(X_k = x_k \mid C_k = 2, \mu_2, \sigma) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x_k - \mu_2)^2\right)$
3. $p(C_k = 1) = p(C_k = 2) = \frac{1}{2}$

Question

Before we assumed that the ideal observer knows μ_1, μ_2 and σ. How would an ideal observer solve the task without this knowledge? This can be calculated by standard methods but is cumbersome.

Category Learning as Ideal Learning

TECHNISCHE UNIVERSITÄT DARMSTADT

Task

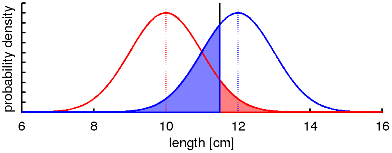
1. $p(X_k = x_k \mid C_k = 1, \mu_1, \sigma) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(x_k - \mu_1\right)^2\right)$
2. $p(X_k = x_k \mid C_k = 2, \mu_2, \sigma) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\left(x_k - \mu_2\right)^2\right)$
3. $p(C_k = 1) = p(C_k = 2) = \frac{1}{2}$

Question

Before we assumed that the ideal observer knows μ_1, μ_2 and σ. How would an ideal observer solve the task without this knowledge? This can be calculated by standard methods but is cumbersome.

Category Learning as Error Learning

TECHNISCHE
UNIVERSITÄT
DARMSTADT

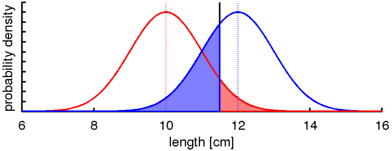


Question

As we know from the ideal observer analysis that the optimal strategy implements a decision criterion *c*, can we come up with a learning mechanism that iteratively updates the criterion and converges to the optimal criterion?

Category Learning as Error Learning

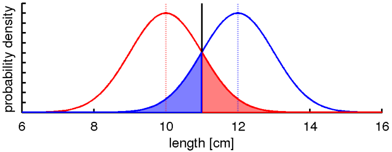
TECHNISCHE UNIVERSITÄT DARMSTADT



$$p(R=1 \mid C=2) = \Phi(c;\mu_2,\sigma)$$
$$p(R=2 \mid C=1) = 1-\Phi(c;\mu_1,\sigma)$$
$$c_{t+1} = \begin{cases} c_t - \delta & \text{if } R=1 \text{ and } C=2 \\ c_t + \delta & \text{if } R=2 \text{ and } C=1 \end{cases}$$

[Kac, 1969]

Category Learning as Error Learning



$$p(R = 1 \mid C = 2) = p(R = 2 \mid C = 1)$$

$$\Phi(c; \mu_2, \sigma) = 1 - \Phi(c; \mu_1, \sigma)$$

$$c_{t+1} = \begin{cases} c_t - \delta & \text{if } R = 1 \text{ and } C = 2 \\ c_t + \delta & \text{if } R = 2 \text{ and } C = 1 \end{cases}$$

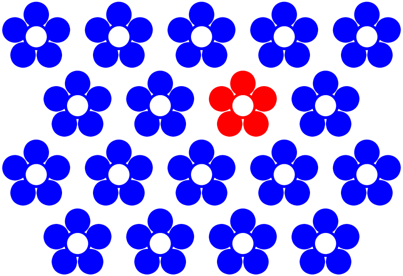
[Kac, 1969]

Bayesian Models Everywhere

- Categorization
- Memory
- Reasoning
- Cue Combination
- Motor Control
- etc.

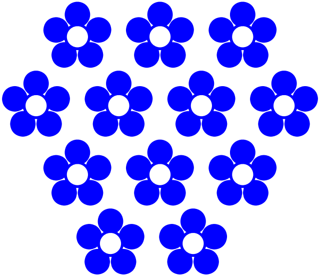
[Ashby and Perrin, 1988, Anderson and Milson, 1989, Oaksford and Chater, 2001, Ernst and Banks, 2002, Körding and Wolpert, 2004]

Judgment Example: Task



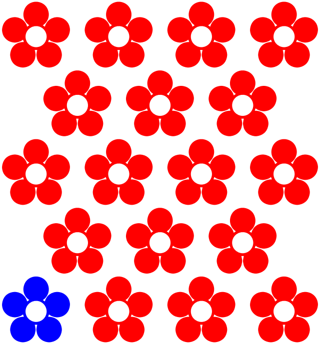
[Nisbett et al., 1983, Kemp et al., 2007]

Judgment Example: Task



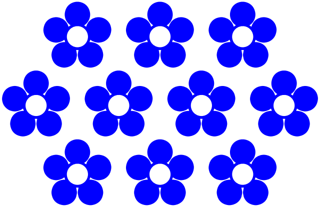
[Nisbett et al., 1983, Kemp et al., 2007]

Judgment Example: Task



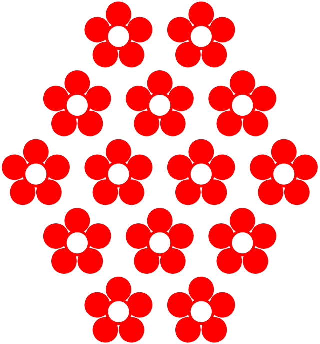
[Nisbett et al., 1983, Kemp et al., 2007]

Judgment Example: Task



[Nisbett et al., 1983, Kemp et al., 2007]

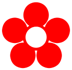
Judgment Example: Task



[Nisbett et al., 1983, Kemp et al., 2007]

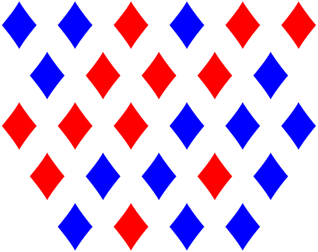
Judgment Example: Task

TECHNISCHE
UNIVERSITÄT
DARMSTADT



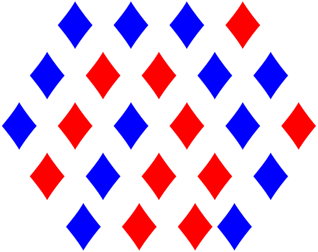
[Nisbett et al., 1983, Kemp et al., 2007]

Judgment Example: Task



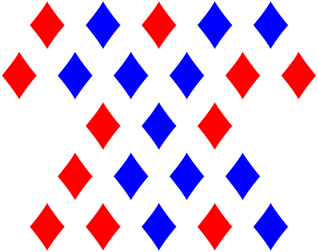
[Nisbett et al., 1983, Kemp et al., 2007]

Judgment Example: Task



[Nisbett et al., 1983, Kemp et al., 2007]

Judgment Example: Task



[Nisbett et al., 1983, Kemp et al., 2007]

Judgment Example: Task

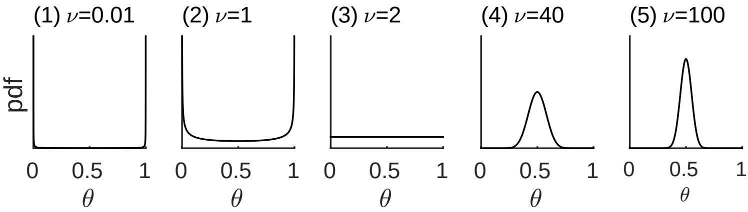
TECHNISCHE UNIVERSITÄT DARMSTADT

♦ ♦ ♦ ♦

[Nisbett et al., 1983, Kemp et al., 2007]

Judgment Example: Ideal Observer

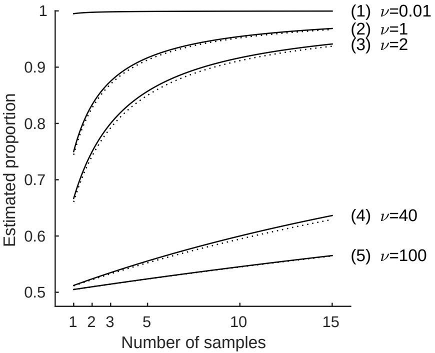
TECHNISCHE UNIVERSITÄT DARMSTADT



[Jäkel and Grusdt, 2019]

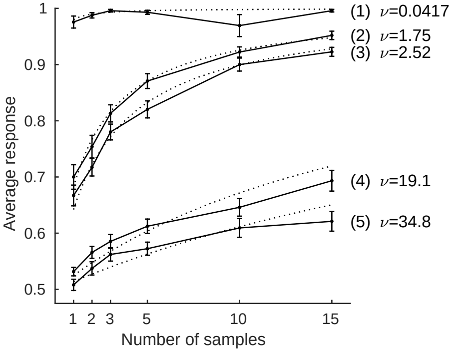
Judgment Example: Ideal Responses

TECHNISCHE
UNIVERSITÄT
DARMSTADT



[Jäkel and Grusdt, 2019]

Judgment Example: Actual Responses



[Jäkel and Grusdt, 2019]

Criticism of Bayesian Models

- “Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition”
- “Bayesian just-so stories in psychology and neuroscience”
- “How Robust Are Probabilistic Models of Higher-Level Cognition?”

[Jones and Love, 2011, Bowers and Davis, 2012, Marcus and Davis, 2013]

Criticism of Bayesian Models

- Neglect of mechanisms
- Obsession with optimality
- Post-hoc rationalizations

[Jones and Love, 2011, Bowers and Davis, 2012, Marcus and Davis, 2013]

Bayesian Fundamentalism

"[...] the primary goal of much Bayesian cognitive modeling has been to demonstrate that human behavior in some task is rational with respect to a particular choice of Bayesian model. We refer to this school of thought as Bayesian Fundamentalism, because it strictly adheres to the tenet that human behavior can be explained through rational analysis – once the correct probabilistic interpretation of the task environment has been identified – without recourse to process, representation, resource limitations, or physiological or developmental data."

[Jones and Love, 2011]

Obsession with Optimality

TECHNISCHE UNIVERSITÄT DARMSTADT

"[…] Together these studies demonstrate that people are adept at combining prior knowledge with new evidence in a manner predicted by Bayesian statistics. […] Recent studies have found that when combining information this way, people are also similar to optimal. […] Their performance in these cue-combination trials can be predicted using the rules of Bayesian integration, further evidencing people's ability to optimally cope with uncertain information. […] In typical cases cues are combined by the subjects in a fashion that is close to the optimum prescribed by Bayesian statistics."

[Berniker and Körding, 2011]

Neglect of Mechanisms?

- We do care about mechanisms!
- We don't necessarily want to claim that the brain is doing the same Bayesian computations that we did in our analysis of the task (remember the categorization example).
- But what role do (optimal, rational, ideal) Bayesian models play in discovering mechanisms then?

Marr's Three Levels of Explanation

TECHNISCHE UNIVERSITÄT DARMSTADT

Computational theory
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?

Representation and algorithm
How can this computational theory be implemented? In particular, what is the representation for the input and the output, and what is the algorithm for the transformation?

Hardware implementation
How can the representation and the algorithm be realized physically?

Note
Ideal observer models formalize computational level analysis.

[Marr, 1982]

Marr's Three Levels of Explanation

Computational theory
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?

Representation and algorithm
How can this computational theory be implemented? In particular, what is the representation for the input and the output, and what is the algorithm for the transformation?

Hardware implementation
How can the representation and the algorithm be realized physically?

Note

Ideal observer models formalize computational level analysis.

[Marr, 1982]

What's an Ideal Observer, Really?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Ideal given the physical limits (quantal nature of light, diffraction limits, retinal sampling, etc.)
- Ideal given the information that the experimenter has available and perfect measurement (signal known, no temporal uncertainty, etc.)
- Ideal given the information that is in principle available to the observer (minimally informed observer)
- Ideal given the statistics of the environment
- Ideal given the information that is available to the observer given processing mechanisms and limitations

[Stüttgen et al., 2011]

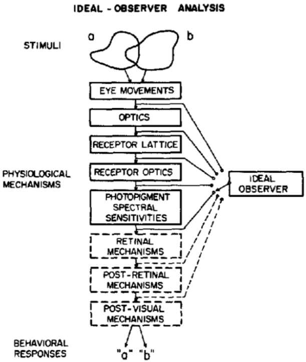
The First Ideal Observer: Quantal Fluctuations

“The purpose of this paper is, in fact, to lay out clearly the absolute limitations to the visual process that are imposed by fluctuation theory and to compare the actual performance of the eye with these limitations. The gap, if there is one, between the performance to be expected from fluctuation theory and the actual performance of the eye is a measure of the “logical space” within which one may introduce special mechanisms, other than fluctuations, to determine its performance. These special mechanisms can only contract the limits already set by fluctuation theory.“

[Rose, 1948]

Sequential Ideal Observer Analysis

TECHNISCHE UNIVERSITÄT DARMSTADT

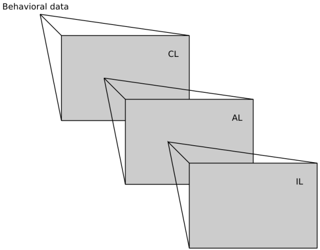


- Precise description of stimulus
- Physical limits
- Start from periphery
- Ideal observer not part of mechanism
- Ideal observer just a tool to avoid central processes
- Works well for solid boxes
- How to apply ideal observers centrally?

[Geisler, 1989]

A Heuristic Research Strategy

TECHNISCHE UNIVERSITÄT DARMSTADT



Marr-Bayes Reverse-Engineering

Bayesian ideal observer analysis is a formal framework for computational-level modeling. The computational level is key in reverse-engineering the mind/brain. But ultimately we want explanations on all three levels: Computational, algorithmic, and implementational.

[Zednik and Jäkel, 2016]

A Heuristic Research Strategy

TECHNISCHE UNIVERSITÄT DARMSTADT

"This method of attack has been found to generate useful hypotheses for further studies. Thus, whereas it is not expected that the human observer and the ideal detection device will behave identically, the emphasis in early studies is on similarities. If the differences are small, one may rule out entire classes of alternative models, and regard the model in question as a useful tool in further studies. Proceeding on this assumption, one may then in later studies emphasize the differences, the form and extent of the differences suggesting how the ideal model may be modified in the direction of reality."

[Swets et al., 1961]

Conclusion

- Bayesian models are useful for understanding tasks
- Bayesian models are useful for summarizing and organizing data
- But they are also useful in guiding a heuristic search for mechanisms
- Theoretically there are many possible algorithms and implementations, but pragmatically not so many
  - Use successful ideas from other fields (machine learning, statistics, AI)
  - Use other constraints, like cognitive limitations, known facts about the hardware, etc.

[Zednik and Jäkel, 2016]

References I

- Anderson, J. R. and Milson, R. (1989).
  Human memory: An adaptive perspective.
  Psychological Review, 96(4):703–719.
- Ashby, F. G. and Perrin, N. A. (1988).
  Toward a unified theory of similarity and recognition.
  Psychological Review, 95(1):124–150.
- Berniker, M. and Körding, K. P. (2011).
  Bayesian approaches to sensory integration for motor control.
  WIREs Cognitive Science, 2:419–428.
- Bowers, J. S. and Davis, C. J. (2012).
  Bayesian just-so stories in psychology and neuroscience.
  Psychological Bulletin, 138(3):389–414.
- Ernst, M. O. and Banks, M. S. (2002).
  Humans integrate visual and haptic information in a statistically optimal fashion.
  Nature, 415:429–433.
- Geisler, W. S. (1989).
  Sequential ideal-observer analysis of visual discrimination.
  Psychological Review, 96(2):267–314.
- Jäkel, F. and Grusdt, B. (2019).
  Combining current observations with prior experiences in everyday inductive reasoning.
  In Preparation.

References II

- Jones, M. and Love, B. C. (2011).
  Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition.
  Behavioral and Brain Sciences, 34:169–231.
- Kac, M. (1969).
  Some mathematical models in science.
  Science, 166(3906):695–699.
- Kemp, C., Perfors, A., and Tenenbaum, J. B. (2007).
  Learning overhypotheses with hierarchical Bayesian models.
  Developmental Science, 10(3):307–321.
- Körding, K. P. and Wolpert, D. M. (2004).
  Bayesian integration in sensorimotor learning.
  Nature, 427(6971):244–7.
- Marcus, G. F. and Davis, E. (2013).
  How robust are probabilistic models of higher-level cognition?
  Psychological Science, 24(12):2351–2360.
- Marr, D. (1982).
  Vision: a computational investigation into the human representation and processing of visual information.
  W. H. Freeman, San Francisco.
- Nisbett, R. E., Krantz, D. H., Jepson, C., and Kunda, Z. (1983).
  The use of statistical heuristics in everyday inductive reasoning.
  Psychological Review, 90(4):339–363.

References III

- Oaksford, M. and Chater, N. (2001).
  The probabilistic approach to human reasoning.
  Trends in Cognitive Sciences, 5(8).
- Rose, A. (1948).
  The sensitivity performance of the human eye on an absolute scale.
  Journal of the Optical Soceity of America, 38(2):196–208.
- Stüttgen, M. C., Schwarz, C., and Jäkel, F. (2011).
  Mapping spikes to sensations.
  Frontiers in Neuroscience, 5(125):1–17.
- Swets, J., Tanner, W. P., and Birdsall, T. G. (1961).
  Decision processes in perception.
  Psychological Review, 68:301–340.
- Zednik, C. and Jäkel, F. (2016).
  Bayesian reverse-engineering considered as a research strategy for cognitive science.
  Synthese, 193:3951–3985.