

Jakob Hohwy: The Predictive Mind

Chapters 1&2

Conrad Friedrich

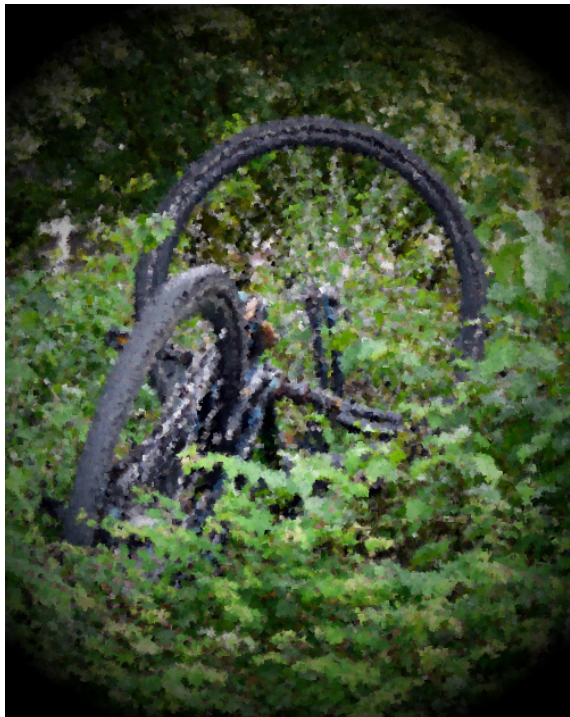
Universität zu Köln

January 29, 2017

Perception and Bayesian Inference

- Vereinheitlichende Theorie des Geistes: Wahrnehmung, Handlung, und “alles mentale dazwischen” (Auch Bewusstsein?).
- Zentrale Idee: Das Gehirn lässt sich als Hypothesen-Prüf-Mechanismus betrachten, der durchgehend damit beschäftigt ist, die Abweichung seiner Vorhersagen/Erwartungen (predictions) von seinen Sinneseindrücken zu *minimieren*.
- Hier: Fokus auf Wahrnehmung.
- Die Sinneseindrücke formen Wahrnehmung nicht direkt, sondern sind Feedback zu den Erwartungen und Anfragen des Geistes “an die Welt”.

- Wahrnehmung besteht in (lässt sich am besten beschreiben als) unbewusster Inferenz auf die wahrscheinlichste *Ursache* meiner *rohen Sinneseindrücke* (und deren Vorhersage).
- Direkter, unmittelbarer Zugang nur zu den Sinneseindrücken, nicht zu den Dingen “in der Welt”.
- *Gewusst* (in einem starken Sinn, Gewissheit) werden nur die Effekte, d.h. Sinneseindrücke. Um etwas über die “versteckten” Ursachen zu erfahren, ist Inferenz nötig.
- Inferenz weniger stark als Gewissheit und insbesondere nicht-monoton.
- Denn: Zwischen Ursachen und Effekten besteht keine 1:1 Relation (sondern $n:m$).
- D.h. Verschiedene Ursachen können denselben Effekt haben, und eine Ursache verschiedene Effekte.



- Verschiedene Objekte, Zustände könnten diesen Sinneseindruck verursachen:
 - Ein Fahrrad, das im Gebüsch liegt
 - Einzelne Fahrradteile, die irgendwie im Gebüsch hängen geblieben sind
 - Ein ungewöhnlich genau koordinierter Schwarm Bienen
- Wie kommen wir vom Sinneseindruck zum (offensichtlichen) Ergebnis, dass hier ein Fahrrad im Gebüsch liegt?

- Also: Wir brauchen eine *Inferenz* auf die 'beste' Ursache.
- Not any old inference will do: Besondere zusätzliche Beschränkungen auf die Art der Inferenz.
 - Hintergrundwissen des Agenten.
 - Möglichkeit, gute/richtige Inferenz von schlechter/falscher abzugrenzen.

- Gegeben Sinneseindruck und Hintergrundwissen soll auf die 'richtige' Ursache des Sinneseindrucks geschlossen werden.
- Ranking der möglichen Ursachen nach Wahrscheinlichkeit.
- Die 'beste' Ursache scheint guter Kandidat für den Wahrnehmungsinhalt zu sein.

Zutaten für die Bayesianistische Inferenz:

h_1, \dots, h_n Möglichen Hypothesen als Ursache eines Sinneseindrucks

e Gegebener Sinneseindruck

$P(h_i)$ *Prior*: Wahrscheinlichkeit einer Hypothese h_i , unabhängig davon, ob es die Ursache ist. Abhängig vom Hintergrundwissen.

$P(e|h_i)$ *Likelihood*: Wahrscheinlichkeit, dass die in der Hypothese beschriebene Ursache so einen Sinneseindruck hervorrufen würde. Abhängig vom Hintergrundwissen. Maß dafür, wie gut eine Hypothese den Sinneseindruck *vorhersagt*.

- Daraus lässt sich für jedes h_i die bedingte Wahrscheinlichkeit gegeben e errechnen.

Vereinfachtes Bayes Theorem

(Standardvariante folgt direkt aus der Definition von bedingter Wahrscheinlichkeit)

$$P(h_i|e) = P(e|h_i)P(h_i)$$

Posterior Probability = Likelihood \times Prior Probability

- Hypothese h_i , für die $P(h_i|e)$ maximal ist, stellt plausibelste Ursache des Sinneseindrucks e dar, gegeben das Hintergrundwissen des Agenten.

Beispiel.

e Sinneseindruck: Ein seltsames Klopfen.

h_1 Ein Specht klopft an der Wand.

h_2 Ein Einbrecher werkelt an der Tür.

...

h_{815} Ich bin ein BIV und bekomme gerade einen elektrischen Stimulus entsprechend dem Sinneseindruck.

Angenommen, ich habe zuletzt viele Spechte in der Nachbarschaft bemerkt, aber eher wenige Einbrecher. Ich habe mal ein Seminar zum Skeptizismus besucht und halte das ganze für großen Humbug. Dann gilt für die *Prior Probabilities*

$$P(h_1) > P(h_2) \gg P(h_{815}).$$

Aber:

$$P(e|h_{815}) > P(e|h_1) \approx P(e|h_2).$$

- Da: Gegeben, ich bin *tatsächlich* ein BIV, ist die Wahrscheinlichkeit für jeden Sinneseindruck, den ich habe, sehr hoch, sogar fast sicher.
- Also höher als die anderen bedingten Wahrscheinlichkeiten.

- Resultat des Rankings:

$$P(h_1|e) > P(h_2|e) > P(h_{815}).$$

- Die Spechthypothese erscheint am wahrscheinlichsten. Resultat der Bayesianistischen Inferenz - unbewusste Wahrnehmungsinferenz.

Was hat das mit Prediction (Vorhersage) zu tun? Immerhin heißt es *Predictive Coding/Processing/Mind*. In der Wahrscheinlichkeitstheorie:

- **Inference:** Reasoning from effects to causes (*latent* or *hidden*).
- **Prediction:** Reasoning from causes to (future) effects.
- Schlussfolgern wie vorgestellt Beispiel für Inferenz. Beide Vorgänge für Modell der Wahrnehmung relevant (s. später: prediction error minimization)

Objections

Beispiel für simple Probabilistische Inferenz vom Effekt zur Ursache. Aber beschreibt das auch *Wahrnehmung* adäquat? Hohwy nennt vier Einwände zu dieser Interpretation.

- Einwand 1 Probabilistische Inferenz scheint viel zu intellektualistisch zu sein, um einen automatischen Vorgang wie Wahrnehmung erklären zu können.
- Einwand 2 Unangebrachte Antromorphisierung von Prozessen im Gehirn: Wieso sollte das Gehirn etwas glauben oder Schlussfolgern, dass wir nicht bewusst machen?
- Einwand 3 Keine offensichtliche Erklärung der Phänomenologie von Wahrnehmung. Wahrnehmung fühlt sich wie etwas an, aber beschreibt die Inferenz nicht nur reine begriffliche Kategorisierung von Input?
- Einwand 4 Arbitrarität der *Prior Beliefs*: Wie zeichnet dieses Modell probabilistisch konsistente (d.h. keine Verletzung der Kolmogorov-Axiome), aber völlig realitätsferne Wahrscheinlichkeitsfunktionen als schlecht/falsch aus?

Art des Arguments:

- ① Bayesian Inference ist schwer und kognitiv anspruchsvoll.
 - Evidenz für notorisch schlechte Bayesian Inference bei Erwachsenen (Linda the Bank-Teller)
- ② Wahrnehmung ist einfach und automatisch.
 - (Kinder und auch Tiere können reliabel wahrnehmen. Außerdem
- ③ Also: Bayesian Inference und Wahrnehmung sind verschieden.

Zu intellektualistisch? - Immediate Intuition von formaler Erkenntnistheorie, in der Degree of Belief Bayesianistisch Updatet werden. Hier aber perception: Inhalte müssen keinen propositionalen Gehalt haben, nicht bewusst zugaenglich sein, usw. Ähnlichkeit zu Konditionalisierungsregel in formaler Erkenntnistheorie. Unterschied: Keine Anforderung, dass e oder h Überzeugungen sind, möglicherweise nicht-propositional.