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 inbesondere 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:

$h_1, ..., h_n$ 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 beschrieben 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.

Vereinfaches Bayes Theorem

(Standardvariante folgt direkt aus der Definition von bedingter Wahrscheinlichkeit)

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

$$\text{Posterior Probability} = \text{Likelihood} \times \text{Prior Probability}$$

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

- Bayes Theorem eigentlich

$$P(h_i|e) = \frac{P(e|h_i)P(h_i)}{P(e)}$$

- Wieso wird der Nenner $P(e)$ weggelassen?
- Für das reine Ranking der Hypothesen irrelevant, da konstanter Faktor.
- Mathematisch: Normalisierungskonstante, so dass $\sum_i P(h_i|e) = 1$ (und nicht weniger).

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 Baysianistischen 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 Probalistische Inferenz vom Effekt zur Ursache. Aber beschreibt das auch *Wahrnehmung* adäquat? Hohwy nennt vier Einwände zu dieser Interpretation.

Einwand 1 Probalistische 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 Kategosierung von Input?

Einwand 4 Arbitrarität der *Prior Beliefs*: Wie zeichnet dieses Modell probabilistisch konsistente (d.h. keine Verletzung der Kolmogorov-Axiome), aber völlig realitätsferne Wahrscheinlichkeitsfuntionen als schlecht/falsch aus?

1. Intellectualist Objection

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

Argument:

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.

Hohwys Strategie:

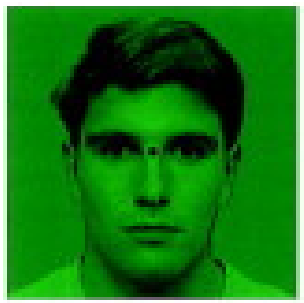
- Fallstudien, deren *beste Erklärung* eine automatische Inferenzleistung des Gehirns ist.
- Mindestens eine der beiden Prämissen ist dann nicht haltbar.
- Studien basieren auf dem Phänomen *Binocular Rivalry*.
- Hohwy: Phänomen lässt sich am besten mit Bayesian Inference erklären.

Binocular Rivalry

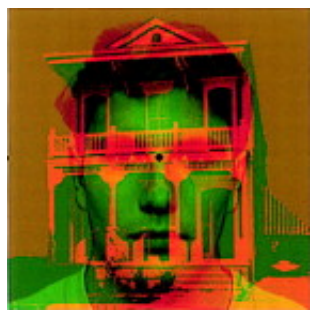
> *If a picture of a house is shown to one eye and a picture of a face is shown to the other, then one should surely just see a face-house. But this is not what happens, as Porta and Wheatstone and many others have described. The brain somehow seems to decide that there are two distinct things out there, a face and a house—and perception duly alternates between seeing one or the other every few seconds, sometimes with periods of patchy rivalry in between.*

- Varianten von Diaz-Caneja (1928), Logothetis (1996), Denison, Piazza et al. (2011), Zhou, Jiang et al. (2010).

Input linkes Auge / Rechtes Auge




'Intuitiv' erwartetes Perzept



Stattdessen stabile sequentielle Wahrnehmung:

Hohwy: Für dieses Phänomen lässt sich eine plausible Bayesianistische Geschichte erzählen.

Input	e	Gesicht/Haus linkes Auge/rechtes Auge
Hypothesen	h_{f+h}	Es ist ein Gesichts-Haus-Mix.
	h_f	Es ist ein Gesicht.
	h_h	Es ist ein Haus.

Likelihoods $P(e|h_f) \approx P(e|h_h) < P(e|h_{f+h})$.

Priors $P(h_f) > P(h_h) \gg P(h_{f+h})$.

Posteriors $P(h_f|e) > P(h_h|e) > P(h_{f+h}|e)$.

Bayesianistisches Modell erklärt, warum ein konstantes Bild gesehen wird.

- Alternierende Wahrnehmung wird in dieser simplen Version *nicht* erklärt.
- Trotzdem sieht Hohwy das Erklärungspotentiel als starke Evidenz für probabilistisches Wahrnehmungsschließen.
- Weitere Evidenz: Probanden Priming-Effekt unterzogen, so dass sie mehrheitlich ein bestimmtes Bild sehen (Zhou, Jiang et al., 2010)
  - Input Bild Text Marker/Rose. Priming durch Rosen*geruch*. Effekt durch Modell vorhergesagt.

2. Anthromorphist Objection

- Ist das Gehirn tatsächlich mit Probabilistischem Schließen beschäftigt, oder ist das unzulässige 'Anthromorphisierung' des Gehirns?
- Howhy argumentiert (a) funktionalistisch und (b) strukturell.
- 'Bayesian coding hypothesis': The brain represents sensory information probabilistically, in the form of probability distributions. [1]
- Fragestellung in Computational Cognitive Neuroscience.

[1] Knill/Pourget (2004)

3. Phenomenologist Objection

- Selbst wenn die Kategorisierung von Wahrnehmungsinhalten Bayesianistisch funktioniert, heißt das nicht, dass das Wahrnehmungserlebnis erklärbar wird.
- "It is not just that we see a car but that we see it, as a car, from our own perspective." (p. 26)

4. Skepticist Objection

- Das bisher beschriebene Framework ist anfällig für *pathological Priors*:
- Völlig absurde, realitätsfremde Priors zeichnen bestimmte Hypothesen als wahrscheinlich aus, gegeben die Sinneseindrücke, und können so bestärkt werden.
- Bootstrapping-Problem (ähnlich epistemischem Internalismus).
- Hohwys Antwort: Bisher nur allgemeines Bayesianisches Framework, *Predictive Coding* erfordert mehr: 'Reality Check'
- *Prediction Error Minimization*.

References

- Tong, Nakayama et al. Binocular Rivalry and Visual Awareness in Human Extrastriate Cortex. Neuron (21), 1998.
- Knill and Pourget. The Bayesian brain: the role of uncertainty in neural coding and computation. TRENDS in Neurosciences (27), 2004.
- Hohwy. The Predictive Mind. Oxford University Press, 2013.