

Multisensorische Wahrnehmung des Menschen

Human Multisensory Perception

Ernst, Marc O.; Bühlhoff, Heinrich H.

Max-Planck-Institut für biologische Kybernetik, Tübingen

Korrespondierender Autor

E-Mail: marc.ernst@tuebingen.mpg.de

Zusammenfassung

Um die Umwelt wahrzunehmen, nutzt unser Gehirn eine Vielzahl von Sinnesreizen, die es von verschiedenen Sinnesmodalitäten, wie Sehen, Hören oder Fühlen bekommt. Teilweise geben Informationsquellen aus verschiedenen Modalitäten Aufschluss über dieselbe Eigenschaft eines Objekts oder eines Events in der Umwelt. Zum Beispiel kann die Größe eines Objekts sowohl mit den Augen gesehen, wie auch mit der Hand gefühlt werden. Dies nennt man redundante Informationsquellen. In diesem Bericht verdeutlichen wir, wie redundante Information vom menschlichen Gehirn genutzt wird, um zielgerichtet mit der Umgebung interagieren zu können. Des Weiteren zeigen wir auf, welche Rolle Vorwissen über die statistischen Regelmäßigkeiten der Umwelt spielt und wie es unsere Wahrnehmung beeinflussen kann. Als Beschreibungsmodell dieser sensomotorischen Interaktionen dient hier die Bayes'sche Entscheidungstheorie.

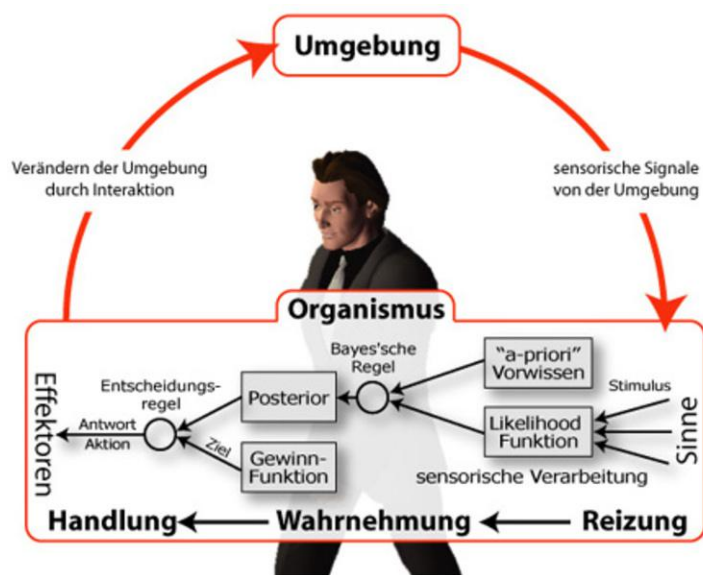
Summary

For perceiving the environment our brain uses multiple sources of sensory information derived from several different modalities, including vision, touch and audition. Some sources of sensory information derived from different modalities provide information about the same object property or event. For example, the size of an object can both be seen with the eyes and felt with the hands. This is called redundant sources of sensory information. In this report we will show how such redundant sources of sensory information are used by the human brain in order to interact with the environment in a purposive fashion. Further, we describe which role prior knowledge plays concerning the statistical regularities in the world and how this can affect the process of perception. As a model for describing such somatosensory interactions we apply the Bayesian Decision Theory (BDT).

Der sensomotorische Kreislauf

Zweck unserer reichhaltigen Wahrnehmung ist es, Handlung und Interaktion mit der Umgebung zu ermöglichen. Greifen wir beispielsweise nach einem Objekt, ist ein komplexes Zusammenspiel von visuellen, taktilen, und kinästhetischen Sinnesreizen notwendig, um das Objekt erfolgreich zu fassen. Alle Informationsquellen müssen zu einem kohärenten Gesamteindruck des Objekts bzw. der Handlung kombiniert werden, um diese Interaktion mit der Umgebung zu ermöglichen. Im Forschungsbereich der sensomotorischen

Integration stehen wir auf dem Standpunkt, dass motorische Handlung und sensorische Verarbeitung untrennbar miteinander verknüpft sind. Daher werden diese auch hier gesamtheitlich betrachtet und zusammen untersucht.



Der Wahrnehmungs-/Handlungszyklus des Menschen und eine schematische Darstellung der sensomotorischen Informationsverarbeitung nach der Bayes'schen Entscheidungstheorie [4, 8].

© Max-Planck-Institut für biologische Kybernetik/Ernst

Der Wahrnehmungs-/Handlungszyklus in **Abbildung 1** dient der Illustration des sensomotorischen Verarbeitungsweges, von der sensorischen Reizung über die Wahrnehmung bis hin zur Handlung. Als Ausgangsmodell zur Untersuchung der sensomotorischen Informationsverarbeitung sehen wir in unserer Arbeitsgruppe die Bayes'sche Entscheidungstheorie. Demnach wird die Information, die aus den verschiedenen sensorischen Kanälen gewonnen wird, zuerst verarbeitet und integriert. Nach dem Bayes'schen Modell wird die sensorische Information, die von der Umwelt gewonnen wird, in der so genannten „Likelihood“-Funktion repräsentiert.

Unsere Wahrnehmung beruht aber sicherlich nicht nur auf der Auswertung der sensorischen Daten, die über das visuelle, haptische oder auditive System in unseren Körper gelangen. Sie wird wesentlich auch vom Vorwissen geprägt, das entweder aus Vorerfahrungen und der Interaktion mit der Umgebung stammen kann oder das aus früheren Generationen in unseren Genen gespeichert ist. Nach der Bayes'schen Entscheidungstheorie wird dieses Vorwissen „a-priori“-Wissen genannt und es repräsentiert Gesetzmäßigkeiten der Umgebung in Form von Wahrscheinlichkeitsverteilungen. Wie sich nun „a-priori“-Wissen mit der sensorischen Information zur Wahrnehmung zusammensetzen, wird in der Bayes'schen Entscheidungstheorie mithilfe der Bayes'schen Regel beschrieben, die besagt, dass das normierte Produkt der „a-priori“-Wahrscheinlichkeitsfunktion mit der Likelihood-Funktion die Posterior-Wahrscheinlichkeitsfunktion ergibt. Um zu einer Handlung oder einer Entscheidung zu kommen, muss allerdings nun noch die Aufgabe oder das Ziel einer Handlung berücksichtigt werden, welches ebenfalls einen Einfluss auf unsere Wahrnehmung haben kann. In der Bayes'schen Entscheidungstheorie kann ein solches Ziel über eine so genannte Gewinnfunktion definiert werden.

Optimale Integration von multimodaler Information

Der Schlüssel zu einer robusten Wahrnehmung liegt in der Integration von mehreren, teilweise redundanten Sinnesreizen. Häufig gibt es mehr als eine einzige sensorische Informationsquelle, die es erlaubt, eine Eigenschaft der Umwelt wahrzunehmen. Klopft man beispielsweise an eine Tür, um freundlich um Einlass zu bitten, so liefern sowohl das Auge (S = sehen), wie auch das Ohr (H = hören) und die Hand (F = fühlen) Information darüber, an welcher Stelle (L) exakt geklopft wurde. Man hat also drei Informationsquellen (S, H, und F) über den Ort L, an dem geklopft wurde. Wie werden diese Informationsquellen miteinander verrechnet, um eine kohärente Wahrnehmung dieses Ortes zu bilden? Glaubt man der Hand, dem Ohr oder dem Auge? Tatsächlich gibt es keine solche Auswahl, sondern es werden alle Informationsquellen, die zur Verfügung stehen, ausgenutzt und miteinander verrechnet. Im ersten Schritt müssen dazu die verschiedenen Informationsquellen erst einmal in ein gemeinsames Datenformat und Koordinatensystem gebracht werden. So müssen die Photonen, die auf die Retina des Auges fallen, in räumliche Information umgesetzt werden, genauso wie die Schallwellen im Ohr und die Kräfte, die auf die Knöchel der Hand wirken. Um den Ort der Hand im Raum zu bestimmen, muss zur Auswertung der Information von Auge und Ohr noch die Stellung des Kopfes K berücksichtigt werden (**Abb. 2**).

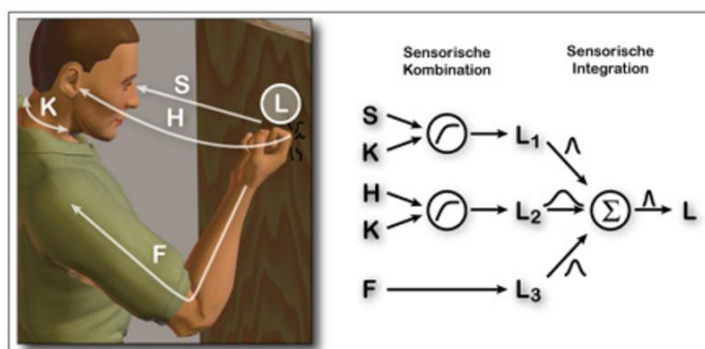


Illustration eines multisensorischen Events: Beim Klopfen an eine Tür liefern sowohl der Seh- wie auch der Hör- und Tastsinn Information über den Ort des Klopfens. Diese drei Informationsquellen müssen im ersten Schritt in ein gemeinsames Koordinatensystem überführt werden, bei dem die Stellung des Kopfes mit in Betracht gezogen werden muss. Im zweiten Schritt können die sensorischen Informationen dann in eine kohärente Repräsentation des Ortes integriert werden.

© Max-Planck-Institut für biologische Kybernetik/Ernst

Bei der Wahrnehmung sollte das Gehirn alle zur Verfügung stehenden Informationsquellen möglichst optimal nutzen. Mathematisch betrachtet würde ein statistisches Modell, die so genannte Maximum-Likelihood-Schätzmethode [7], die optimale Kombination liefern. Diese besagt, dass alle drei Informationsquellen (L_1 , L_2 , L_3) über den Ort des Klopfens, gewonnen von Auge, Ohr und Hand, gemittelt werden müssen, und zwar so, dass die Information je nach Qualität mehr oder weniger starken Einfluss hat. Das ist ein gewichtetes Mittel [siehe 4 für einen Überblickartikel]. Nur auf diese Weise könnten alle bereitstehenden Sinnesinformationen über ein Objekt so effektiv wie möglich ausgeschöpft werden. Mathematisch lässt sich das wie folgt beschreiben:

$$\hat{L} = \sum_i w_i \hat{L}_i \quad \text{mit} \quad \sum_i w_i = 1 \quad (\text{Eq. 1})$$

©

wobei w_i die einzelnen Gewichte sind. Diese sind normiert, sodass deren Summe 1 ergibt. So wie physikalische Messgeräte immer eine kleine Fehlertoleranz haben, sind auch unsere Sinnesorgane nicht perfekt: Alle dem Gehirn gemeldeten Eindrücke sind auf ihre Art mehr oder weniger „verrauscht“ – das heißt, sie haben eine

gewisse Varianz, die ein Maß für die Qualität (oder auch die Verlässlichkeit) des Signals ist. Definiert ist die Qualität r_i der sensorischen Information als invers proportional zur Varianz $r_i = 1/\sigma_i^2$ des sensorischen Signals. Je verrauschter und damit unsicherer die Sinnesdaten z.B. beim Tasten sind, desto weniger Gewicht sollten diese im Wettbewerb mit den anderen Informationsquellen (beispielsweise dem Sehen) bekommen. Somit sind die Gewichte durch die normierte Verlässlichkeit bestimmt:

$$w_j = \frac{r_j}{\sum_{i=1, \dots, j, \dots, N} r_i} \quad (\text{Eq. 2})$$

©

Der Vorteil mehrerer zur Verfügung stehender Signale ist, dass dem Gehirn eine genauere Schätzung erlaubt wird, also eine Schätzung mit geringerer Varianz (höherer Verlässlichkeit). Werden die Gewichte entsprechend Eq. 2 gewählt, kann gezeigt werden, dass dies der optimalen Ausnutzung der Information entspricht und somit das Gesamttrauschen minimiert wird. Die Verlässlichkeit r der Schätzung ist dann die Summe der Verlässlichkeiten der einzelnen Informationsquellen. Das bedeutet, dass die Gesamtverlässlichkeit r der Schätzung durch die Integration von verschiedenen Informationsquellen i stetig zunimmt.

$$r = \sum_i r_i$$

©

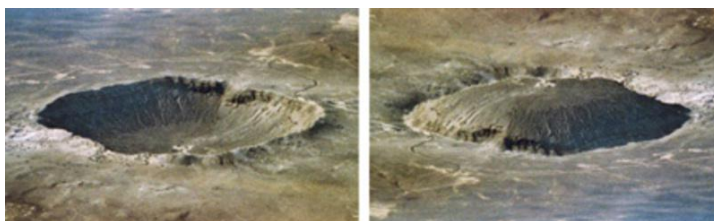
In einigen psychophysikalischen Verhaltensstudien konnten wir kürzlich zeigen, dass der Mensch sensorisch Information tatsächlich in dieser statistisch optimalen Art und Weise ausnutzt [3]. Die Versuchspersonen, die an den Verhaltensexperimenten teilgenommen haben, hatten die Aufgabe, die Größe zweier balkenförmiger Objekte zu vergleichen. Diese Balken konnten sie sowohl sehen als auch betasten. Aus den Fehlern, die die Versuchspersonen beim Vergleichen machten, konnten wir dann schließen, wie verlässlich die sensorischen Signale der verschiedenen Sinne im Gehirn verarbeitet wurden.

Mithilfe eines trickreichen Versuchsaufbaus, bei dem die Testobjekte virtuell am Computer erzeugt werden, war es uns möglich, das visuell wahrnehmbare und das zu ertastende Objekt unabhängig voneinander zu manipulieren. Wir erzeugten damit eine Diskrepanz zwischen den beiden Objekten und untersuchten dann, ob sich die Versuchspersonen mehr an dem visuellen oder mehr an dem ertasteten Eindruck orientierten. Auf diese Weise ermittelten wir die Gewichtung der Signale untereinander und stellten fest, dass das Verhalten der Versuchspersonen exakt mit den Vorhersagen des statistischen Modells übereinstimmte: War die Größe des Objekts klar und deutlich zu sehen, dominierte der visuelle Eindruck; wurde die Qualität des Bildes jedoch

herabgesetzt, indem gezielt Rauschen über das Bild gelegt wurde, dominierte die ertastete Größe [3].

Der Einfluss von Vorwissen auf unsere Wahrnehmung

Wie oben schon angedeutet, beruht unsere Wahrnehmung jedoch nicht nur auf Information, die von verschiedenen Sinnesmodalitäten wie Sehen, Hören oder Fühlen kommt. Um die permanent auf uns einwirkenden, vielfältigen und komplexen Reize interpretieren zu können, nutzt unser Gehirn Vorwissen, das wir über die Welt besitzen. Dieses Zusammenspiel zwischen momentaner Sinnesinformation einerseits und Vorwissen andererseits lässt sich an Beispielen eindrucksvoll zeigen (**Abb. 3**). So ist für die meisten Betrachter auf dem linken Teilbild dieser Abbildung ein Krater (Loch) zu sehen. Hingegen wird das rechte Teilbild zumeist als Vulkan (Hügel) wahrgenommen. Beide Interpretationen sind jedoch nicht eindeutig. Das merkt man sofort, wenn man das Bild auf den Kopf stellt - nun sollte sich die Interpretation der Bilder umkehren. Tatsächlich handelt es sich nämlich bei den beiden Abbildungen um exakt das gleiche Bild, nur dass die rechte Version um 180 Grad gedreht ist.



Krater oder Vulkan? Nimmt man an, das Licht kommt von links oben, so wird das linke Teilbild als Krater interpretiert und das rechte als Vulkan. Nimmt man hingegen an, das Licht kommt von rechts unten, dreht sich die Interpretation um, und das linke Teilbild wird als Vulkan an der Decke gesehen und das rechte Teilbild stellt dann einen Krater in der Decke dar. Doch diese zweite Interpretation ist für die meisten Betrachter viel schwieriger zu sehen, da unsere Wahrnehmung auf die Interpretation „Licht von oben“ getrimmt ist. Dreht man das Bild auf den Kopf, sieht man, wie sich die Interpretation Krater/Vulkan umkehrt.
© USGS copyright free policy

Wie kommt es zu dieser doppeldeutigen Interpretation? Der Grund, warum wir einmal einen Vulkan und das andere Mal einen Krater sehen, liegt daran, dass das Bild nicht eindeutig ist und wir zur Interpretation des Bildes Vorwissen benötigen. In diesem Fall ist es eine Annahme darüber, aus welcher Richtung das Licht einfällt. Da das Licht in unserer Welt meistens von oben kommt (Sonne, Deckenbeleuchtung, etc.), ist es für das Gehirn vernünftig, dieses Vorwissen bei der Interpretation der Bilder zu benutzen. Das Vorwissen, das wir hier also einsetzen, ist, dass Licht in der Regel von oben kommt. Unter dieser Vorannahme – das Licht kommt von oben – ist es die einzige „richtige“ Interpretation, das linke Teilbild als Krater zu sehen und das rechte als Vulkan. Würde unser Gehirn jedoch die Vorannahme machen, das Licht käme von unten, so dreht sich die Interpretation exakt um – das linke Teilbild würde dann als Vulkan und das rechte als Krater gesehen, jedoch dann nicht am Boden, sondern an der Decke. Wie anfänglich schon angedeutet, lässt sich dieses Zusammenspiel zwischen Sinnesinformation und Vorwissen mathematisch sehr gut mit der von Bischof Thomas Bayes 1763 entwickelten Wahrscheinlichkeitstheorie beschreiben, in der das Vorwissen mithilfe von „Priors“, der „a-priori- Wahrscheinlichkeitsverteilung“, beschrieben wird. In dem illustrierten Beispiel gäbe es also einen Prior für „Licht von oben“.

Wie kommt das Gehirn zu solchen Vorannahmen? Höchstwahrscheinlich beruhen sie auf statistischen Regelmäßigkeiten der Umwelt, und diese werden bei der Interpretation der sensorischen Information

berücksichtigt. Ob wir diese erlernen können oder ob uns dieses Vorwissen genetisch mitgegeben wurde, haben wir kürzlich untersucht [1]. Dazu benutzten wir ähnlich doppeldeutige Stimuli wie die in Abbildung 3 und untersuchten die Wahrnehmung von Versuchspersonen jeweils vor und nach einem interaktiven Training. Bei diesem Training konnten die Versuchspersonen die Stimuli nicht nur sehen, sondern auch für ca. 1,5 Stunden mit ihren Fingern abtasten. Der Trick dabei war nun, dass während des Trainings das visuelle Bild und damit der Bereich des Lichteinfalls, bei dem sich die Objekte als Vulkan anföhlten, um 30 Grad gedreht wurde. In dieser künstlich verdrehten Welt stimmte der geföhlte Eindruck nicht immer mit der visuellen Wahrnehmung überein und einige der Objekte, die vor dem Training als Vulkan gesehen wurden, föhlden sich nun wie ein Krater an und umgekehrt.

Das Ergebnis dieser Experimente war, dass sich tatsächlich aufgrund dieses interaktiven Trainings die Wahrnehmung der Probanden veränderte – einige der Krater wurden plötzlich als Vulkane gesehen. Durch das Training war die „Licht von oben“-Annahme um durchschnittlich ca. 11 Grad gedreht. Da normalerweise auch die scheinbare Helligkeit eines Objekts von seiner Orientierung zum Lichteinfall abhängt, sollte eine generelle Änderung der „Licht von oben“-Annahme auch die Wahrnehmung der Helligkeit von Objekten verändern. Tatsächlich zeigte sich, dass den Probanden Objekte nach dem Training unterschiedlich hell erschienen – in Übereinstimmung mit der Änderung der angenommenen Lichteinfallsrichtung.

Somit konnten wir zeigen, dass selbst solche fundamentalen Vorannahmen, wie die „Licht von oben“-Annahme, nicht (ausschließlich) genetisch verankert sind, sondern ständig auf der Basis der statistischen Regelmäßigkeit, mit der sie in der Umgebung auftreten, innerhalb relativ kurzer Zeit erlernt und angepasst werden. Das steht jedoch im Gegensatz zu Versuchen, die mit anderen Spezies, wie z.B. Hühnern, durchgeführt wurden, die ebenfalls eine „Licht von oben“-Annahme machen, die aber bei diesen Tieren genetisch verankert zu sein scheint. Insgesamt unterstreichen diese Ergebnisse die enorm gute Anpassungsfähigkeit des menschlichen Gehirns - und dass sich die Wahrnehmung je nach individueller Lerngeschichte und Vorwissen von Mensch zu Mensch unterscheidet.

Ausblick

In diesem Bericht haben wir beschrieben, wie multisensorische Information im Gehirn integriert wird und welchen Einfluss Vorwissen auf unsere Wahrnehmung haben kann. Sicherlich macht es keinen Sinn, alle Information, die dem Gehirn zur Verfügung steht, in eine gemeinsame Repräsentation zu integrieren. Obwohl sowohl die Hand wie auch das Auge dem Gehirn Größeninformation von Objekten liefern können, macht es nur Sinn, diese Information zu integrieren, wenn sie von ein und demselben Objekt kommt. Kommt die Größe, die die Hand föhlt, von einem anderen Objekt wie die Größe, die das Auge sieht, wäre es sinnlos, diese beiden Größeninformationen miteinander zu verrechnen. Das setzt aber voraus, dass das Gehirn schon vorher weiß, ob die multisensorische Information von ein und demselben Objekt stammt oder ob sie von verschiedenen Objekten herröhrt. Im Moment sind wir im Begriff, diese Grenzen multisensorischer Integration zu untersuchen, und wir haben hier schon erste Ergebnisse erzielt [5, 6]. Die Frage, die wir uns nun auch hier stellen, ist, ob das menschliche Gehirn diese Klassifizierung in „das ist ein Objekt“ oder „das sind verschiedene Objekte“ nach statistisch optimalen Kriterien vornehmen kann.

Da optimales Handeln nicht angeboren sein kann und weil sich die Umwelt und damit die Voraussetzungen dafür ständig ändern, muss das Gehirn adaptierfähig sein [vgl. auch 2]. Der Versuch, der zuletzt beschrieben wurde, verdeutlichte dies dadurch, dass sich die „Licht von oben“-Annahme durch interaktives Training an eine künstlich veränderte Umgebung angepasst hatte [1]. Die Gesetzmäßigkeiten, nach denen diese Adaptationsprozesse ablaufen, sind jedoch noch weitgehend ungeklärt. Daher sind wir gerade im Begriff,

solche sensomotorischen Adaptationsprozesse eingehend zu untersuchen. Multimodale Integrations- und Adaptationsprozesse sind zentraler Bestandteil der menschlichen Wahrnehmung, und deren Aufklärung wird daher wesentlich zum Verständnis der Funktionsweise des menschlichen Gehirns beitragen.

Originalveröffentlichungen



[Nach](#) [Erweiterungen](#) [suchen](#)[Bilderweiterung](#)[Channelticker](#)[Dateiliste](#)[HTML-Erweiterung](#)[Jobticker](#)[Kalendererweiterung](#)[Linkerweiterung](#)[MPG.PuRe-Referenz](#)[Mitarbeiter](#) [\(Employee Editor\)](#)[Personenerweiterung](#)[Publikationserweiterung](#)[Teaser](#) [mit](#) [Bild](#)[Textblockerweiterung](#)[Veranstaltungstickererweiterung](#)[Videoerweiterung](#)[Videolistenerweiterung](#)[YouTube-Erweiterung](#)

[1] **W. J. Adams, E. W. Graf & M. O. Ernst:**

Experience can change "light-from-above" prior.

Nature Neuroscience 7 (10), **2004**, 1057-1058.

[2] **M. O. Ernst, M. S. Banks, H. H. Bülthoff:**

Touch can change visual slant perception.

Nature Neuroscience 3, **2000**, 69-73.

[3] **M. O. Ernst, M. S. Banks:**

Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion.

Nature 415, **2002**, 429-433.

[4] **M. O. Ernst, H. H. Bülthoff:**

Merging the Senses into a Robust Percept.

Science 8 (4), **2004**, 162-169.

[5] **M. O. Ernst:**

A Bayesian view on multimodal cue integration.

(Eds.) Knoblich, G., Thornton, I. M., Grosjean, M., & Shiffrar, M., Perception of the human body from the inside out. Oxford University Press, New York, USA, **2005**.

[6] **J. M. Hillis, M. O. Ernst, M. S. Landy, M. S. Banks:**

Combining Sensory Information: Mandatory Fusion Within, but Not Between, Senses.

Science 298, **2002**, 1627-1630.

[7] **M. S. Landy, L. T. Maloney, E. B. Johnston, M. J. Young:**

Measurement and modeling of depth cue combination: In defense of weak fusion.

Vision Research 35, **1995**, 389-412.

[8] **A. L. Yuille, H. H. Bülthoff:**

Bayesian theory and psychophysics.

In Knill, D. and Richards, W. (eds.), Perception as Bayesian Inference, Cambridge University Press, Cambridge, **1996**, pp. 123-161.