International Encyclopedia of the Social and Behavioral Sciences, 2nd edition -

CONTRIBUTORS' INSTRUCTIONS

PROOFREADING

The text content for your contribution is in final form when you receive proofs. Read proofs for accuracy and clarity, as well as for typographical errors, but please DO NOT REWRITE.

Titles and headings should be checked carefully for spelling and capitalization. Please be sure that the correct typeface and size have been used to indicate the proper level of heading. Review numbered items for proper order – e.g., tables, figures, footnotes, and lists. Proofread the captions and credit lines of illustrations and tables. Ensure that any material requiring permissions has the required credit line and that we have the relevant permission letters.

Your name and affiliation will appear at the beginning of the article and also in a List of Contributors. Your full postal address appears on the non-print items page and will be used to keep our records up-to-date (it will not appear in the published work). Please check that they are both correct.

Keywords are shown for indexing purposes ONLY and will not appear in the published work.

Any copy-editor questions are presented in an accompanying Author Query list at the beginning of the proof document. Please address these questions as necessary. While it is appreciated that some articles will require updating/revising, please try to keep any alterations to a minimum. Excessive alterations may be charged to the contributors.

Note that these proofs may not resemble the image quality of the final printed version of the work, and are for content checking only. Artwork will have been redrawn/relabelled as necessary, and is represented at the final size.

DESPATCH OF CORRECTIONS

PLEASE KEEP A COPY OF ANY CORRECTIONS YOU MAKE.

Proof corrections should be returned in one communication to Claire Byrne, Gemma Tomalin or Mike Nicholls (ISB2proofs@elsevier.com), by [DATE 7 days after proofs sent out by typesetter] using one of the following methods:

1. **PREFERRED:** Corrections should be annotated to the PDF and sent attached to an email to the Elsevier MRW Department at **ISB2proofs@elsevier.com**

2. Listed in an e-mail and sent to the Elsevier MRW Department at **ISB2proofs@elsevier.com**.

The e-mail should state the article code number in the subject line. Corrections should be consecutively numbered and should state the paragraph number, line number within that paragraph, and the correction to be made.

3. If corrections are substantial, send the amended hardcopy by courier to Claire Byrne, Gemma Tomalin or Mike Nicholls, **Elsevier MRW Department, The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK.**

Note that a delay in the return of proofs could mean a delay in publication. Should we not receive corrected proofs within 7 days, Elsevier may proceed without your corrections.

CHECKLIST

Author queries addressed/answered?	☐
Affiliations, names and addresses checked and verified?	☐
Permissions details checked and completed?	☐
Outstanding permissions letters attached/enclosed?	☐
Figures and tables checked?	☐

If you have any questions regarding these proofs please contact the Elsevier MRW Department at: **ISB2proofs@elsevier.com**

ISB2 56035

Non Print Items

Author and Co-author Contact Information

Will Penny,
12 Queen Square,
London WC1N 3BG,
UK.
Tel.: +44 203 448 4365;
fax: +44 207 813 1420.
E-mail: w.penny@ucl.ac.uk

Keywords

Bayesian inference; Behavioral neuroscience; Cognitive neuroscience; Collective decision making; Computational neuroscience; Decision making; Dynamical models; Normative model; Probabilistic inference; Sensorimotor integration; Sensory integration; Social neuroscience; Visual processing

Author Query Form

Title: ISB2
Article Title/Article ID: Bayesian Models in Neuroscience/56035

Dear Author,

During the preparation of your manuscript for typesetting some questions have arisen. These are listed below. (AU indicates author queries; ED indicates editor queries; and TS/TY indicates typesetter queries.) Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list. Your responses should then be returned with your marked proof/list of corrections to Claire Byrne, Gemma Tomalin or Mike Nicholls at Elsevier via isb2proofs@elsevier.com

Queries and/or remarks

[AU1]	Please provide institute/university name for the author "Will Penny."
[AU2]	Please confirm the edits to the sentence "Kersten et al. (2004)…"
[ED1]	Can you please verify the article titles "Bayesian Statistics; Bayesian Decision Theory; Bayesian Graphical Models and Networks; Vision, High Level Theory of" listed in "See also" section and provide appropriate ones so that these appear in the contents of this publication. Please refer to the Table of Contents at the project website: http://mrw.elsevier.com/isb2/menu.htm for the Ms-code and article titles.

a0005
Bayesian Models in Neuroscience

AU1
Will Penny, London, UK

Abstract

abspara0010 The last two decades have seen a revolution in the statistical sciences with classical inferential approaches becoming slowly replaced by Bayesian ones. This has impacted on many fields in science and engineering, including neuroscience. A Bayesian perspective on neuroscience includes the idea that human behavior in many domains is close to statistical optimality and that the structure and machinery of the brain itself is reflected in message-passing algorithms that implement Bayesian inference.

p0010 This article presents a review of Bayesian models in Neuroscience. The work in this field splits broadly into two categories: (1) Bayesian modeling of behavior, which develops normative behavioral models based on statistical optimality principles, and (2) Bayesian models of the brain, which describe the internal operations of the brain that would give rise to such behaviors. In the latter category, neuroscientific data from a variety of sources are used including invasive electrophysiological recordings from mice, rats, and monkeys. It also focuses on detailed neurophysiological mechanisms over a range of spatial scales, from plasticity in dendrites to changes in connectivity among brain regions. Given the nature of this encyclopedia, our review focuses most strongly on the former category (behavior), as this is more closely related to psychology and sociology.

p0015 The review is structured in two main sections. The first section provides a mathematical description of Bayesian inference, introducing the basic concepts that will be referred to later on. We use the example of a game of tennis to make the main ideas more concrete. The second section reviews Bayesian models of behavior, focusing on a range of topics. Under 'sensory integration,' we review how, for example, information from visual and tactile senses are optimally combined. Under 'visual processing,' we review models of visual search and the processing of ambiguous or conflicting visual information. Under 'sensorimotor integration,' we describe how dynamic Bayesian models accurately describe how the consequences of movements can be predicted and perceptually downweighted. Under 'collective decision making,' we review how Bayesian inference has been used to describe how groups of people interact to make decisions.

s0010
Bayesian Inference

p0020 Consider some quantity, x. Our beliefs about the likely values of x can be described by the probability distribution $p(x)$. If we make a new observation y that is related to x, then we can update our belief about x using Bayesian inference. In statistics the optimal way of updating your beliefs is via Bayes rule.

p0025 First we need to specify the likelihood of observing y given x. This is specified by a probability distribution called the likelihood, $p(y|x)$. It tells us, if we know x, what are the likely values of y. Our updated belief about x, that is, after observing the new data point y is given by the posterior distribution $p(x|y)$.This can be computed via Bayes rule:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \qquad [1]$$

The denominator ensures that $p(x|y)$ sums to 1 over all possible values of x, i.e., that it is a probability distribution. It can be written as p0030

$$p(y) = \int p(y|x')p(x')\mathrm{d}x' \qquad [2]$$

Equations [1] and [2] describe the basic computations underlying Bayes rule. These are multiplication and normalization (eqn [1]) and marginalization (eqn [2]). Following Wolpert and Ghahramani (2004), we will use the game of tennis to illustrate key points. Imagine that you are receiving serve. One computation you need to make before returning serve is to estimate x, the position of the ball when it first hits the ground, as depicted in **Figure 1**. p0035

It is possible to make an estimate solely on the basis of the balls trajectory, i.e., via the data y. We can find the value of x which maximizes the likelihood, $p(y|x)$. This is known as Maximum Likelihood (ML) estimation. It is also possible to p0040

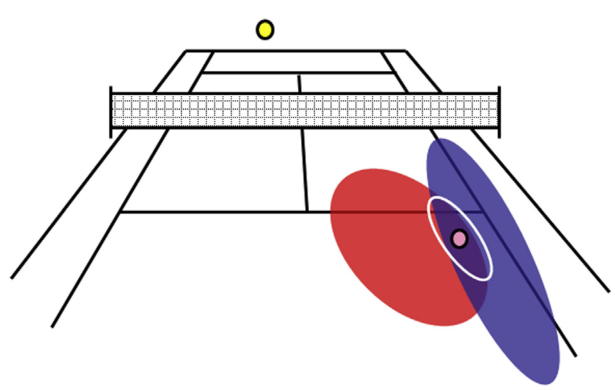

Figure 1 Estimating the position of the ball when it first lands. The prior is shown in blue, the likelihood distribution in red, and the posterior distribution with the white ellipse. The maximum posterior estimate is shown by the magenta ball. This estimate can be updated in light of new information about the ball's trajectory (yellow). Adapted from Wolpert, D., Ghahramani, Z., 2004. Bayes rule in perception, action and cognition. In: Oxford Companion to the Mind. Oxford Univ. f0010

estimate the uncertainty in this estimate. The ML estimate and the uncertainty in it together give rise to the likelihood distribution shown in Figure 1.

p0045 But before our opponent hits the ball we may have a good idea as to where they will serve. It may be the case, for example, that when they serve from the right the ball tends to go down the line. We can summarize this belief by the prior distribution $p(x)$ (shown in blue in Figure 1). We can then use Bayes rule to estimate the posterior distribution. This is the optimal combination of prior knowledge ('down the line') and new data (visual information from the ball's trajectory). Our final single best estimate of where the ball will land is then given by the maximum of the posterior density. This is known as MAP estimation (from 'Maximum a Posteriori').

p0050 As we continue to see the ball coming toward us we can refine our belief as to where we think the ball will land. This can be implemented by applying Bayes rule recursively such that our belief at time point n depends only on our belief at the previous time point, $n-1$. That is,

$$p(x_n|Y_n) = \frac{p(y_n|x_n)p(x_n|Y_{n-1})}{p(Y_n)} \qquad [3]$$

where $Y_n = \{y_1, y_2, \ldots, y_n\}$ denotes all observations up to time n. Our prior belief, that is, prior to observing data point y_n is simply the posterior belief after observing all data points up to time $n-1$, $p(x_n|Y_{n-1})$. Colloquially, we say that "today's prior is yesterday's posterior."

p0055 The variable x is also referred to as a hidden variable or hidden state because it is not directly observed. If the hidden state were a discrete variable, such as whether the ball landed in or out of the service box, one can form a likelihood ratio:

$$\mathrm{LR} = \frac{p(x_n = \mathrm{IN}|Y_n)}{p(x_n = \mathrm{OUT}|Y_n)} \qquad [4]$$

p0060 Decisions based on the likelihood ratio are statistically optimal in the sense of having maximum sensitivity for any given level of specificity. In contexts where LR is recursively updated, these decisions correspond to a sequential likelihood ratio test (Bogacz et al., 2006). There is a good deal of evidence showing that the firing rate of single neurons in the brain report evolving log LR values (Gold and Shadlen, 2001).

s0015 Gaussians

p0065 If our random variables x and y are normally distributed, then Bayesian inference can be implemented exactly using simple formulae. These are most easily expressed in terms of precisions, where the precision of a random variable is its inverse variance; a precision of 10 corresponds to a variance of 0.1. We first look at inference for a single variable (e.g., distance from side of tennis court).

p0070 For a Gaussian prior with mean m_0 and precision λ_0 and a Gaussian likelihood with mean m_d and precision λ_d, the posterior distribution is Gaussian with mean m and precision λ where

$$m = \frac{\lambda_0}{\lambda} m_0 + \frac{\lambda_d}{\lambda} m_d \qquad [5]$$

p0075 So, precisions add and the posterior mean is the sum of the prior and data means, but each weighted by their relative precision. This relationship is illustrated in Figure 2. Though fairly simple, eqn [5] shows how to optimally combine two sources of information. As we shall see in the following section, various aspects of human behavior from sensory integration to instances of collective decision making have been shown to conform to this 'normative model.' Similar formulae exist for multivariate (instead of univariate) Gaussians (Bishop, 2006) where we have multidimensional hidden states and observations, e.g., three-dimensional position of the ball, and two-dimensional landing position on court surface.

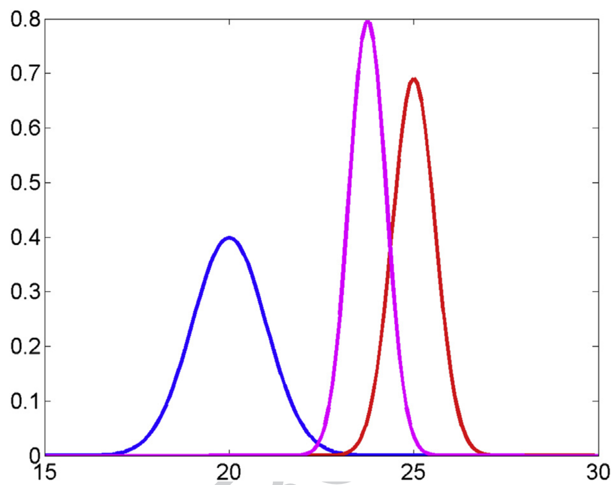

Figure 2 Bayes rule for Gaussians. For the prior $p(x)$ (blue) $m_0 = 20$, $\lambda_0 = 1$ and the likelihood $p(y|x)$ (red) $m_d = 25$, $\lambda_d = 3$, the posterior $p(x|y)$ (magenta) shows the posterior distribution with $m = 23.75$, $\lambda = 4$. The posterior is closer to the likelihood than the prior because the likelihood has higher precision. Bayes rule for Gaussians has been used to explain many behaviors from sensory integration to collective decision making. f0015

Behavioral Models s0020

An attractive feature of Bayesian models of behavior is that they provide descriptions of what would be optimal for a given task. They are often referred to as 'ideal observer' models because they quantify how much to update our beliefs in light of new evidence. Departures from these normative models can then be explained in terms of other constraints such as computational complexity or individual differences. p0080

One way to address individual differences is to use an empirical Bayesian approach in which parameters of priors and their parametric forms are estimated from data. See Stocker and Simoncelli (2006) for an example of this approach in modeling visual motion processing. p0085

What follows is a review of Bayesian models of sensory integration, visual processing, sensorimotor integration, and collective decision making. As we shall see, the priors that we have about, for example, our visual world most readily show themselves in situations of stimulus ambiguity or at low signal-to-noise ratios. Much of the phenomenology of these perceptual illusions is long established (Gregory, 1998) but Bayesian modeling provides new quantitative explanations and predictions. A more introductory review of much of this p0090

ISB2 56035

material is available in Frith's outstanding book on mind and brain (Frith, 2007).

s0025 Sensory Integration

p0095 Ernst and Banks (2002) considered the problem of integrating information from visual and tactile (haptic) modalities. If vision v and touch t information are independent given observation of an object x, then Bayesian fusion of sensory information produces a posterior density:

$$p(x|v,t) = \frac{p(v|x)p(t|x)p(x)}{p(v,t)} \qquad [6]$$

p0100 For a uniform prior $p(x)$ and for Gaussian likelihoods, the posterior will also be a Gaussian with precision λ_{vt}. From Bayes rule for Gaussians (eqn [5]), we know that precisions add:

$$\lambda_{vt} = \lambda_v + \lambda_t \qquad [7]$$

where λ_v and λ_t are the precision of visual and haptic senses alone, and the posterior mean is a relative-precision weighted combination:

$$m_{vt} = w_v m_v + w_t m_t \qquad [8]$$

where the weights are $w_v = \lambda_v/\lambda_{vt}$ and $w_t = \lambda_t/\lambda_{vt}$.

p0105 Ernst and Banks (2002) asked subjects which of two sequentially presented blocks was the taller. Subjects used either vision alone, touch alone or a combination of the two. They recorded the accuracy with which discrimination could be made and plotted this as a function of difference in block height. This was repeated for each modality alone and then both together. They also used various levels of noise on the visual images. From the single modality discrimination curves they then fitted cumulative Gaussian density functions, which provided estimates of the precisions λ_t and $\lambda_v(i)$ where i indexes visual noise levels.

p0110 In the dual modality experiment, the weighting of visual information predicted by Bayes' rule for the ith level of visual noise is

$$w_v(i) = \frac{\lambda_v(i)}{\lambda_v(i) + \lambda_t} \qquad [9]$$

p0115 This was found to match well with the empirically observed weighting of visual information. They observed visual capture (strong weighting) at low levels of visual noise and haptic capture at high levels. Inference in this simple Bayesian model is consistent with standard signal detection theory. However, Bayesian inference is more general as it can accommodate, for example, nonuniform priors over block height.

p0120 There have been numerous studies of the potential role of Bayesian inference for integration of other senses. For example, object localization using visual and auditory cues (Alais and Burr, 2004) has supported a Bayesian integration model with vision dominating audition in most ecologically valid contexts. This visual capture is the basis of the 'ventriloquism' effect, but is rapidly degraded with visual noise. This literature has considered only simple inferences about single variables such as block height or spatial location. Nevertheless these studies have demonstrated a fundamental concept; that sensory integration is near Bayes-optimal.

The representations in the brain that are used to support the above computations are a subject of ongoing study. One proposal is that populations of neurons represent probability distributions such that higher neuronal firing rates reflect higher precisions. Moreover, if the distribution of cell activities is approximately Poisson then Bayesian inference for optimal cue integration, for example, can be implemented with simple linear combinations of neural activity (Ma et al., 2006). Recent recordings in which monkeys were trained to estimate direction of motion from visual and vestibular cues showed neuronal activity consistent with this theory (Fetsch et al., 2011). p0125

Visual Perception s0030

Kersten et al. (2004) reviewed the problem of visual object perception and argued that much of the ambiguity in visual processing, for example concerning occluded objects, can be resolved with prior knowledge. This idea is naturally embodied in a Bayesian framework (Knill and Richards, 1996) and has its origins in the work of Helmholtz who viewed perception as 'unconscious inference.' An example is how the inference of shape from shading is informed by a 'light-from-above' prior. This results in circular patches that are darker at the bottom being perceived as convex. The adaptability of this prior, and subsequent perceptual experience, has been demonstrated by Adams et al. (2004). p0130 AU2

An example of such a Bayesian modeling approach is the work of Yu et al. (2009) who propose a normative model for the Eriksen Flanker task. This simple decision-making task was designed to probe neural and behavioral responses in the context of conflicting information. On each trial, three visual stimuli are presented and subjects are required to press a button depending on the identity of the central stimulus. The flanking stimuli are either congruent or incongruent. Yu et al. (2009) proposed a discrete time ideal observer model that qualitatively captured the dynamics of the decision-making process. This used the recursive form of Bayes rule in eqn [3]. In later work, a continuum time limit of this model was derived, which produced semi-analytic predictions of reaction time and error rate that accurately predicted subject behavior. They also proposed an algorithm for how these models could be implemented in the brain. p0135

Weiss et al. (2002) proposed that many motion illusions arise from the result of Bayes-optimal processing of ecologically invalid stimuli. Their model was able to reproduce a number of psychophysical effects based on the simple assumptions that measurements are noisy and the visual system has a prior which expects slower movements to be more likely than faster ones. For example, the model could predict the direction of global motion of simple objects such as rhomboids, as a function of contrast and object shape. This model was later refined (Stocker and Simoncelli, 2006) by showing the prior to be non-Gaussian and subject specific, and that measurement noise variance was inversely proportional to visual contrast. p0140

Najemnik and Geisler (2005) developed an ideal Bayesian observer model of visual search for a known target embedded in a natural texture. Prior beliefs in target location were updated to posterior beliefs using a likelihood term that reflected the foveated mapping properties of visual cortex. When this likelihood was matched to individual subjects discrimination p0145 ability, the resulting visual searches were nearly optimal in terms of the median number of saccades. Later work showed that fixation statistics were also similar to the ideal observer.

p0150 An important idea to emerge in recent years is that Bayesian inference operates at the level of cortical macrocircuits (Rao and Ballard, 1999). These circuits are arranged in a hierarchy that reflects the hierarchical structure of the world around us, e.g., scenes are comprised of objects that are in turn comprised of features that are in turn comprised of simpler elements. The connections among regions have also been proposed to have a rather specific anatomy. For example, top-down connections originate from cortical pyramidal cells lying deep in the gray matter (i.e., close to white matter) whereas bottom-up connections originate more superficially (Mumford, 1992). The signals sent up and down these cortical hierarchies then map onto the message-passing algorithms that implement Bayesian inference in dynamic, hierarchical models (Friston, 2003).

p0155 If the world we perceive is the result of hierarchical processing in cortical networks then, because this processing may take some time (of the order of 100 ms), what is perceived to be the present could actually be the past. As this would be disadvantageous for the species, it has been argued that our perceptions are based on predictive models. A 50 ms delay in processing could be accommodated by estimating the state of the world 50 ms in the future. There is much experimental evidence for this view (Nijhawan, 1994). However, a purely 'predictive' account fails to accommodate recent findings in visual psychophysics. The flash-lag effect, for example, is a robust visual illusion whereby a flash and a moving object that are located in the same position are perceived to be displaced from one other. If the object stops moving at the time of the flash no such displacement is perceived. This indicates that the position of the object after the flash affects our perception of where the flash occurred. This 'postdictive' account explains the phenomenon (Eagleman and Sejnowski, 2000) and related data where the object reverses its direction at the flash time. A simple Bayesian model has been proposed to account for the activity of V4 neurons in this task (Sundberg et al., 2006) and later experimental work found evidence for a linear combination of both predictive and postdictive mechanisms.

s0035 Sensorimotor Integration

p0160 Wolpert et al. (1995) have examined the use of dynamic Bayesian models, also referred to as forward models, for sensorimotor integration. These models describe relations among a current state variable x_n, motor commands u_n, and sensory observations y_n. The state might correspond, for example, to a multivariate vector comprising a list of joint angle positions and velocities. These dynamic models have a transition density $p(x_n|x_{n-1}, u_{n-1})$ describing how the hidden state evolves over time. The posterior distribution over hidden states can then be computed as described in standard texts (Bishop, 2006). First, the dynamical equation describing state transitions is integrated to create an estimate of the next state. This requires as input a copy of the current motor command (so-called 'efference copy') and the current state. This is referred to as a time update step. A prediction of sensory input can then be made based on the predicted next state and the mapping from x_n to y_n. Finally, a measurement update or correction step can be applied, which updates the state estimate based on current sensory input.

Wolpert et al. (1995) cite a number of key features of dynamic Bayesian models including the following. First, they allow outcomes of actions to be predicted and acted upon before sensory feedback is available. This may be important for rapid movements. Second, they use efference copy to cancel the sensory effects of movement ('re-afference'), e.g., the visual world is stable despite eye movements. Third, simulation of actions allows for mental rehearsal which can potentially lead to improvements in movement accuracy. p0165

This mathematical framework was applied to the estimation of arm position using proprioceptive feedback and a forward model based on a linear dynamical system (Wolpert et al., 1995). Inference in this model was then implemented using a Kalman filter. The resulting bias and variance in estimates of arm position were shown to closely correspond to human performance, with proprioceptive input becoming more useful later on in the movement when predictions from the forward model were less accurate. p0170

One of the core ideas behind these forward models is that, during perceptual inference, the sensory consequences of a movement are anticipated and used to attenuate the percepts related to these sensations. This mechanism reduces the predictable component of sensory input to self-generated stimuli, thereby enhancing the salience of sensations that have an external cause. This has many intriguing consequences. For example, it predicts that self-generated forces will be perceived as weaker than externally generated forces. This prediction was confirmed in a later experiment (Shergill et al., 2003), thereby providing a neuroscientific explanation for force escalation during conflict; children trading tit-for-tat blows will often assert the other hit him harder. p0175

Körding and Wolpert (2004) have investigated learning in the sensorimotor system using a visual reaching task in which subjects moved their finger to a target and received visual feedback. This feedback provided information about target position that had an experimentally controlled bias and variance. Subjects were found to be able to learn this mapping (from vision to location) and integrate it into their behavior in a Bayes-optimal way. p0180

Returning to our tennis theme, an analysis of 3 years of Wimbledon games has indicated that the outcome of the current point depends on the outcome of the previous point (Klaassen and Magnus, 2001). There are multiple potential sources of correlation here. It could be that a player intermittently enjoys a sweet parameter spot where his internal sensorimotor model accurately predicts body and ball position and is able to hit the ball cleanly, or perhaps a player finds a new pattern in his opponent's behavior such as body position, or previous serve, predicting current service direction. p0185

Collective Decision Making s0040

Sorkin et al. (2001) have applied Bayes rule for Gaussians (see eqn [5]) in their study of collective decision making. Here the optimal integration procedure involves each group members' input to the collective decision being weighted proportionally by the member's competence at the task. p0190

Statistically, 'competence' corresponds to precision. This model of group behavior was shown to be better than a different model which assumed members made individual decisions which were then combined into a majority vote. This latter model better described collective decision making when members did not interact.

p0195 Bahrami et al. (2010) investigated pairs of subjects (dyads) making collective perceptual decisions. Dyads with similarly sensitive subjects (similar precisions) were found to produce collective decisions that were close to optimal, but this was not the case for dyads with very different sensitivities. These observations were explained by a Bayes-optimal model under the assumption that subjects accurately communicated their confidence. This confidence sharing proved essential for the group decision to be better than the decision of the best subject.

ED1 *See also:* Bayesian Statistics; Bayesian Decision Theory; Bayesian Graphical Models and Networks; 43021; 43068; 43090; Vision, High Level Theory of.

Bibliography

Adams, W., Graf, E., Ernst, M., 2004. Experience can change the 'light-from-above' prior. Nature Neuroscience 7, 1057–1058.
Alais, D., Burr, D., February 2004. The ventriloquist effect results from near-optimal bimodal integration. Current Biology 14 (3), 257–262.
Bahrami, B., Olsen, K., Latham, P., Roepstor, A., Rees, G., Frith, C., August 2010. Optimally interacting minds. Science 329 (5995), 1081–1085.
Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.
Bogacz, R., Brown, E., Moehlis, J., Holmes, P., Cohen, J., 2006. The physics of optimal decision making: a formal analysis of models and performance in two-alternative forced-choice tasks. Psychological Review 113 (4), 700–765.
Doya, K., Ishii, S., Pouget, A., Rao, R. (Eds.), 2007. Bayesian Brain: Probabilistic Approaches to Neural Coding. MIT Press.
Eagleman, D., Sejnowski, T., March 2000. Motion integration and postdiction invisual awareness. Science 287 (5460), 2036–2038.
Ernst, M., Banks, M., 2002. Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415 (6870), 429–433.
Fetsch, C., Pouget, A., DeAngelis, G., Angelaki, D., 2011. Neural correlates of reliability-based cue weighting during multisensory integration. Nature Neuroscience 15 (1), 146–154.
Friston, K., 2003. Learning and inference in the brain. Neural Networks 16 (9), 1325–1352.
Frith, C., 2007. Making Up the Mind: How the Brain Creates Our Mental World. Wiley-Blackwell.
Gold, J., Shadlen, M., 2001. Neural computations that underlie decisions about sensory stimuli. Trends in Cognitive Sciences 4, 10–16.
Gregory, R., 1998. Eye and Brain: The Psychology of Seeing. Oxford University Press.
Kersten, D., Mamassian, P., Yuille, A., 2004. Object perception as Bayesian inference. Annual Review of Psychology 55, 271–304.
Klaassen, F., Magnus, J., 2001. Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. Journal of the American Statistical Association 96, 500–509.
Knill, D., Richards, W., 1996. Perception as Bayesian Inference. Cambridge.
Kording, K., Wolpert, D., 2004. Bayesian integration in sensorimotor learning. Nature 427, 244–247.
Ma, W., Beck, J., Latham, P., Pouget, A., November 2006. Bayesian inference with probabilistic population codes. Nature Neuroscience 9 (11), 1432–1438.
Mumford, D., 1992. On the computational architecture of the neocortex II: the role of cortico-cortical loops. Biological Cybernetics 66, 241–251.
Najemnik, J., Geisler, W., March 2005. Optimal eye movement strategies in visual search. Nature 434 (7031), 387–391.
Nijhawan, R., 1994. Motion extrapolation in catching. Nature 370, 256–257.
Rao, R.P., Ballard, D.H., January 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. Nature Neuroscience 2 (1), 79–87.
Shergill, S., Bays, P., Frith, C., Wolpert, D., July 2003. Two eyes for an eye: the neuroscience of force escalation. Science 301 (5630), 187.
Sorkin, R., Hays, C., West, R., January 2001. Signal-detection analysis of group decision making. Psychological Review 108 (1), 183–203.
Stocker, A., Simoncelli, E., April 2006. Noise characteristics and prior expectations in human visual speed perception. Nature Neuroscience 9 (4), 578–585.
Sundberg, K., Fallah, M., Reynolds, J., February 2006. A motion-dependent distortion of retinotopy in area v4. Neuron 49 (3), 447–457.
Weiss, Y., Simoncelli, E., Adelson, E., June 2002. Motion illusions as optimal percepts. Nature Neuroscience 5 (6), 598–604.
Wolpert, D.M., Ghahramani, Z., Jordan, M.I., 1995. An internal model for sensorimotor integration. Science 269 (5232), 1880–1882.
Wolpert, D., Ghahramani, Z., 2004. Bayes rule in perception, action and cognition. In: Oxford Companion to the Mind. Oxford Univ.
Yu, A., Dayan, P., Cohen, J., June 2009. Dynamics of attentional selection under conflict: toward a rational Bayesian account. Journal of Experimental Psychology Human Perception and Performance 35 (3), 700–717.

Relevant Websites

http://cnl.salk.edu/.
http://www.eecs.berkeley.edu/Faculty/Homepages/jordan.html.
http://www.fil.ion.ucl.ac.uk/~wpenny/.
http://www.gatsby.ucl.ac.uk/.
http://en.wikipedia.org/wiki/Bayesian_approaches_to_brain_function.