A Bayesian perspective on magnitude estimation

Frederike H. Petzschner[1], Stefan Glasauer[2,3,4], and Klaas E. Stephan[1,5]

[1] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zürich & ETH Zürich, Switzerland
[2] Center for Sensorimotor Research and Department of Neurology, Ludwig-Maximilian University Munich, Munich, Germany
[3] German Center for Vertigo and Balance Disorders (DSGZ), Ludwig-Maximilian University Munich, Munich, Germany
[4] Bernstein Center for Computational Neuroscience, Ludwig-Maximilian University Munich, Munich, Germany
[5] Wellcome Trust Centre for Neuroimaging, University College London, London, UK

Our representation of the physical world requires judgments of magnitudes, such as loudness, distance, or time. Interestingly, magnitude estimates are often not veridical but subject to characteristic biases. These biases are strikingly similar across different sensory modalities, suggesting common processing mechanisms that are shared by different sensory systems. However, the search for universal neurobiological principles of magnitude judgments requires guidance by formal theories. Here, we discuss a unifying Bayesian framework for understanding biases in magnitude estimation. This Bayesian perspective enables a re-interpretation of a range of established psychophysical findings, reconciles seemingly incompatible classical views on magnitude estimation, and can guide future investigations of magnitude estimation and its neurobiological mechanisms in health and in psychiatric diseases, such as schizophrenia.

Theories of magnitude estimation

Our ability to judge duration, distance, or size is crucial for a mental representation of, and interaction with, the physical world, such as building a cognitive map, performing accurate movements, playing an instrument, or doing sports [1,2]. It has long been known that humans show strikingly similar behavioral signatures (and biases) in magnitude estimation across different sensory modalities, such as proprioception, vision, or audition [3–9]. Along with imaging studies, the universal expression of these behavioral effects has supported the idea of a generalized magnitude estimation system [10–14]. However, at the same time, each physical quantity might also have a specialized representation that is related to the sensory organs with which it is typically associated and the computational problems in whose treatment it has a role [15]. Therefore, previous work has called for computational models as a way to disentangle common and distinct processes in magnitude representation and estimation [16].

So far, however, attempts to model magnitude estimation have often led to modality-specific or effect-specific explanations [17]. By contrast, recently proposed Bayesian accounts of magnitude estimation have the potential to provide a more general explanation that covers a wide set of behavioral characteristics and transcends any specific modality [18–20]. This Bayesian framework suggests that behavioral phenomena of magnitude estimation, such as characteristic biases observed across sensory domains, are the result of integrating noisy sensory information with prior experience. From this perspective, estimation errors are neither due to limitations of the sensory channels nor result from erroneous cortical representations. Instead, on average, they optimize behavioral outcomes by accounting for noise and are the natural consequence of general principles underlying perceptual inference (i.e., the deployment of a predictive model that takes the learned statistics of the environment into account) [21]. This perspective derives from long-standing theories of perception in general and provides a formal foundation to examine aberrations of magnitude estimation in psychiatric diseases, such as schizophrenia [22].

In this review, we discuss how a Bayesian framework can: (i) provide a unifying perspective that explains a variety of behavioral features of magnitude estimation; (ii) shed new light on classical psychophysical laws by reconciling the work of Weber-Fechner and Stevens and

---

Glossary

Discrimination task: requires binary decisions about the difference between two consecutively or simultaneously presented stimuli (e.g., whether tone A is louder than tone B).

Generative model: specifies a joint probability distribution of hidden states and/or parameters and the observed data; this requires specification of likelihood and prior.

Kalman filter: a statistical technique, which infers the current (hidden) state of a state space model based on the previous observations. It can be used to model an online Bayesian estimation process that is updated on a trial-by-trial basis.

Matching task: requires that the magnitude of a new stimulus is actively adjusted to a previously experienced one. Matching tasks can be used within the same stimulus dimension ('within-modality matching'), such as reproducing a walked distance, or across different modalities ('cross-modality matching'), such as matching a number to the brightness of a light bulb.

Stevens' power law: proposes a power law relation between physical magnitudes and the representation by sensory systems. The power law exponent is characteristic for the respective sensory modality.

Weber-Fechner law: proposes a logarithmic relation between physical magnitudes and the representation by sensory systems.

providing a re-interpretation of their laws; and finally (iii) guide the exploration of the neurobiological underpinnings of magnitude estimation in health and disease.

A Bayesian framework for magnitude estimation

Regardless of whether we examine the estimation of time, distances, length, or loudness, certain behavioral phenomena reoccur across studies (Figure 1A) [23]. The most common ones are depicted in Figure 1B: (i) A tendency of subjective estimates to be biased towards the center of the distribution (regression effect); (ii) an increase of this bias for larger sample ranges (range effect); (iii) a linear increase in standard deviation of estimates with mean magnitude (scalar variability); and (iv) correlations between subsequent magnitude judgments (sequential or order effects) (see Box 1 for a detailed description). Although scalar variability seems to be the consequence of a general logarithmic representation of magnitudes according to the Weber-Fechner law [24] (see Box 2 and Glossary), the remaining effects have often only been explained by modality-specific theories [17].
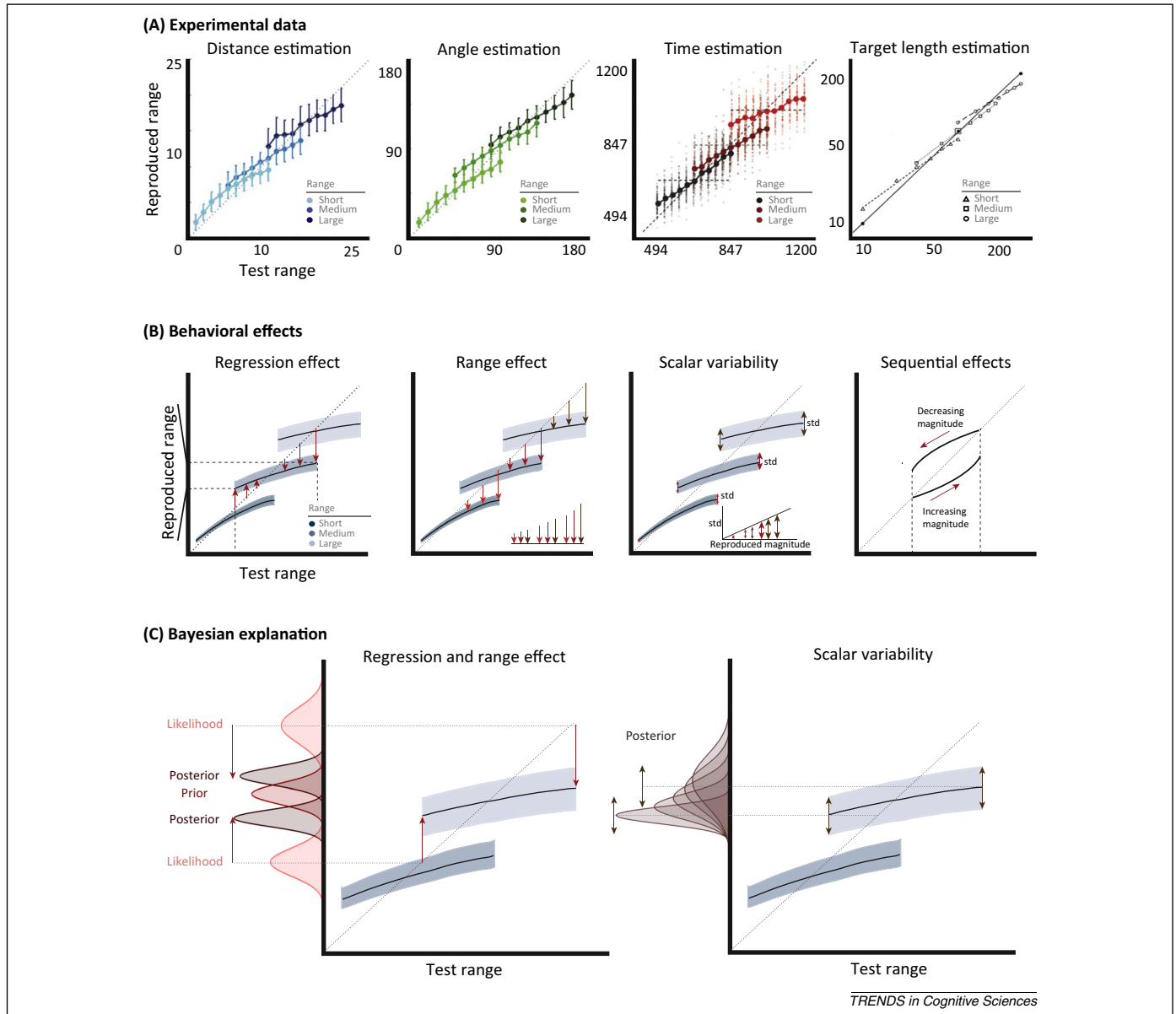

Figure 1. Overview of the behavioral signatures in magnitude estimation and their Bayesian explanation. **(A)** Similar behavioral characteristics observed in data from four different magnitude estimation experiments over more than 100 years on distance estimation, turning angle estimation, time estimation, and target length of a guided movement [18,19,103,104]. Each study used three test ranges (short, medium, and large range). Note that length estimation is plotted on logarithmic scales; therefore, there is no characteristic curvature in the reproduction data. **(B)** Detailed depiction of the observed behavioral characteristics in magnitude estimation from (A). The regression effect refers to a characteristic bias towards the center of each test range, leading to a smaller reproduced range compared to the physical test range. The range effect refers to an increase of the regression effect for larger sample ranges. Scalar variability refers to a linear increase in standard deviation with the mean of the reproduced magnitude. Sequential effects refer to a bias in magnitude estimates towards the recent history of stimuli experienced (see also Box 1). **(C)** A Bayesian framework can explain the characteristics effects shown in (B). A prior around the center of the test distribution would bias posterior estimates towards the center of the respective test range, causing the range and regression effect. Scalar variability predicting an increase in standard deviation with the mean of the likelihood would cause the bias to be stronger for larger magnitudes (larger sample ranges). On all plots: the tested sampled magnitudes are on the X-axis and estimated reproduced magnitudes are on the Y-axis. Diagonal lines reflect the location of a nonbiased, veridical magnitude estimate. Adapted from [18,19,103] (A).

Box 1. Characteristic effects in magnitude estimation

Regression and range effect

The regression effect or 'central tendency of judgment' describes the tendency of magnitude estimates to be systematically biased towards the center of the tested distribution. It causes an underestimation of large magnitudes and an overestimation of small ones within a fixed range of test samples [7,50] (see Figures 1B and 2 in main text). The regression effect occurs independently of the size of the test range, but becomes more pronounced for larger test magnitudes. This increase of the bias towards the center with increasing magnitudes is called the 'range effect' (see Figure 1B in main text) [105].

Scalar variability

The increase in estimation bias for larger magnitudes is accompanied by a linear increase in the standard deviation of the estimated magnitude with its mean. That is, estimates of larger magnitudes are noisier than the estimates of smaller magnitudes (see Figure 1B in main text). The effect was described in the context of time estimation as 'coefficient of variation' or 'scalar variability' [8], but is also related to the 'size effect' in number estimation [106]. Scalar variability is equally well explained by a linear increase in noise with increasing magnitudes on linear scales or by a logarithmic representation of magnitude with fixed variability [16,107]. Support for the logarithmic coding hypothesis comes from animal studies that found evidence for a nonlinear compression of numerical information and velocity in relation to behavioral readouts of magnitude estimation [108,109].

Sequential effects

Magnitude judgments are dependent not only on the test range and magnitude of the stimulus, but also on the recent history of stimuli. This sequential effect, also known as 'order effect', involves correlations between subsequent responses, such that estimates after a large previous stimulus tend to be larger, while estimates after a small previous stimulus tend to be smaller. The sequential dependencies are responsible for a perceptual 'hysteresis effect', which means that approaching a stimulus from larger stimulus magnitudes (decreasing) will lead to a different estimate compared with approaching the same stimulus from smaller stimulus magnitudes (increasing) [110] (see also Figure 1B in main text). Sequential effects are important because they highlight a role of learning and prior beliefs and are one of the motivations for considering Bayesian perspectives on magnitude perception.

Box 2. The psychophysical laws

Around 1850, Weber discovered that the just-noticeable difference (JND) between two magnitudes $\Delta\pi$ in a discrimination experiment is directly proportional to the absolute physical magnitude π [111] (Equation I):

$$k = \frac{\Delta\pi}{\pi} \tag{I}$$

Assuming that the constant 'Weber fraction' k might in fact reflect a JND difference in sensation $\Delta s \propto k$, Fechner derived a logarithmic relation between absolute physical and sensed magnitudes, the Weber-Fechner law [5] (Equation II):

$$s \propto \ln\frac{\pi}{\pi_0} \tag{II}$$

where the integration constant π_0 is often referred to as the detection threshold.

Nearly 100 years later, Stevens showed empirically that the relation between reproduced and physical magnitudes was best fit by a power-function across various sensory modalities [6] (Equation III):

$$r \propto \pi^n \tag{III}$$

This observation was based on a series of matching tasks where the magnitude of a new stimulus is actively adjusted to a previously experienced one. Stevens concluded that Fechner's assumption was wrong and viewed the estimated power law exponent n as a fixed property of the respective sensory system (e.g., $n_{loudness} = 0.67$).

Notably, Stevens and Fechner had not only different conceptualizations of how sensation can be measured, but also based those on different experimental designs (discrimination versus matching tasks). Subsequent work proposed that in fact, both laws might be valid, but simply reflect different processing stages in magnitude estimation [41,42,45] (see main text).

The strongest criticism of Stevens work came from Poulton, one of his students who found contextual variations in the fitted exponent depending on the range of stimuli tested, the task instructions, and various other components that were not directly related to the sensory system [23]. Poulton argued that a full understanding of magnitude estimation needs to take these characteristics into account. As discussed in this paper, a Bayesian perspective might be capable of doing so.

The existence of different theoretical accounts for similar perceptual biases across sensory domains is an unsatisfactory state of affairs, and a modality-independent principle providing a more parsimonious explanation of the above effects would be desirable. Several recent studies have suggested that the incorporation of prior experience in the estimate of magnitudes could provide such an overarching explanation [18,19,25,26]. A suitable mathematical framework refers to a Bayesian observer using a so-called 'generative model' to infer on the causes of his or her sensory inputs (Box 3). Although Bayesian models have been successfully applied to multiple perceptual phenomena over the past two decades [27–30], their foray to magnitude estimation happened more recently [18,25,31]. In this framework, a generative model combines *a priori* information (prior) with noisy sensory input (likelihood), weighing the two information sources by their relative uncertainty [32–35]. Notably, this statistically optimal combination of prior knowledge and sensory input produces biased magnitude judgments whenever the prior differs from the current physical stimulus magnitude.

This bias increases with the strength (precision) of the prior belief and decreases with the signal to noise ratio of the sensory input (Figure 2).

A simple fixed prior at the center of the test range, which encodes mean and variance of previously encountered stimuli, can explain range and regression effects. Such a 'center-prior' would be learned for naturally occurring classes of stimuli (or over the course of an experiment) and would bias estimates of a newly experienced stimulus away from the far ends of its distribution, thus causing a 'regression to the mean' (Figure 1C) [19,25,36]. Furthermore, larger stimulus magnitudes linearly increase the standard deviation of the likelihood (scalar variability), hence the influence of the center-prior on the estimates grows stronger for larger test stimuli and so does the bias towards the mean. This explains a stronger regression effect for larger sample ranges or the range effect. Empirically, a Bayesian model with center-prior has proven useful at explaining time estimation behavior, outperforming alternative models that did not implement prior knowledge [18].

The caveat with a fixed center-prior is that it cannot account for sequential dependencies between consecutive stimuli (Box 1). However, a straightforward explanation

Box 3. The Bayesian observer

The Bayesian perspective dates back to the work of Helmholtz on perception as unconscious inference [112]. Over the past two decades, numerous formal treatments of this idea have appeared (e.g., [85,113,114]; collectively, these are often referred to as the 'Bayesian brain' theory (for overviews, see [87,115]). A key theme of these theories is that the brain constructs and continuously updates a generative model of the sensory inputs it receives.

Technically, a Bayesian observer learns the statistics of the environment (e.g., the statistics of experimental stimuli), and combines this prior knowledge $P(\pi)$ with noisy sensory inputs (represented by a likelihood function $P(s|\pi)$), according to Bayes' rule (Equation I):

$$P(\pi|s) \propto P(s|\pi) \cdot P(\pi) \qquad [I]$$

This results in a posterior estimate $P(\pi|s)$ that is more accurate than either of the two information sources alone. Importantly, the combination of likelihood and prior in Bayes' rule is driven by their respective uncertainty (variance). When both distributions are Gaussian [i.e., $P(\pi) \sim N(\mu_\pi, \sigma_\pi)$ and $P(s|\pi) \sim N(\mu_s, \sigma_s)$], then the posterior is also a Gaussian distribution $P(\pi|s) \sim N(\mu_{\pi|s}, \sigma_{\pi|s})$, and the posterior mean $\mu_{\pi|s}$ is a simple uncertainty-weighted average of the 'data mean' μ_s (sensory input about physical magnitude) and the prior mean μ_π (Equation II):

$$\mu_{\pi|s} = w_s \cdot \mu_s + w_\pi \cdot \mu_\pi \qquad [II]$$

The respective weights (w_s and w_π) are also known as 'precision' and are inversely proportional to uncertainty (variance) of sensory input and prior (Equation III).

$$w_s = 1 - w_\pi = \frac{\frac{1}{\sigma_s^2}}{\frac{1}{\sigma_s^2} + \frac{1}{\sigma_\pi^2}} \qquad [III]$$

In cases where the sensory input is noisy (i.e., the variance of the likelihood, σ_s^2, is high), the weight of the likelihood is small and posterior estimates (perceptual judgments) are dominated by the prior. Conversely, for small measurement noise, posterior estimates would be driven by the measurement and would be less prone to influences by the prior.

The link between the posterior estimate of magnitude and the ensuing motor response is captured by a response model, resulting in a probability of observing a response given both prior and sensory input $P(r|s,\pi)$ (see also Figure 3B in main text).

for this type of bias can be obtained when considering the acquisition of the prior over the course of the experiment [21,37]. In contrast to a fixed center-prior, an 'online-prior' would be learned and updated on a trial-by-trial basis according to Bayes' Rule. A simple implementation of this
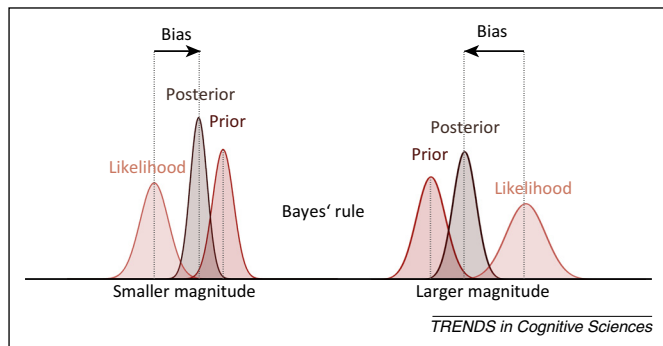

Figure 2. Schematic example for Bayes rule for two different magnitudes. The posterior is proportional to the product of the likelihood and the prior (Bayes' theorem). If the prior is closer to the center of the test range, the posterior will be biased towards the center of the distribution. The strength of the bias depends on the relative uncertainty of likelihood and prior. As the standard deviation of the likelihood for larger magnitudes increases, the bias towards the prior also increases.

learning mechanism would rest on Kalman filtering. The influence of previous stimuli on updating the online-prior decays with time, and the strongest influence is exerted by the most recent stimulus. This explains sequential dependencies between consecutive stimuli and captures both regression and range effects: because the learned prior will never lie outside the sample distribution, any estimates of stimuli at the extreme ends of the distribution would be biased towards previous ones (which are likely closer to the mean of the distribution) (Figure 2). Online-priors of this sort have successfully explained biases in human distance and angle estimation [19], number estimation [38], or reaching movements [39].

Links between classical and generative models of magnitude estimation

At the core of understanding magnitude estimation is building a concept of how physical causes in the world π (actual distances, sound, etc.) are linked to the observed behavioral responses r.

From a classical view, this relation $f_{\pi \to r}$ may be decomposed into at least two parts (Figure 3A): a perceptual function $f_{\pi \to s}$ that transforms the physical magnitude into a stimulus representation s by the sensory system, and a response function $f_{s \to r}$ that maps this internal representation onto a motor act (reproduction) [40]. Early psychophysics aimed at finding a lawful relation for the perceptual function $f_{\pi \to s}$ between the physical quantities and subjective sensory representations. On the basis of discrimination tasks, Fechner originally proposed that this cause-sensation function has a logarithmic form (Weber-Fechner law). A different view arose from studies by Stevens on matching tasks, which focused on the overall relation between physical magnitude and observed responses, proposing that $f_{\pi \to s}$ is a power-function (Stevens' power law) (Box 2). Various studies have suggested that Weber-Fechner's and Stevens' law could be mathematically reconciled [41–44]. In particular, it has been pointed out that Stevens' power-function only holds if the response function $f_{s \to r}$ is linear [42,45]. It is important to note that although these two laws constitute seminal findings in psychophysics, neither of them provides an explanation for context-dependent variations in behavior, such as the range effect or sequential dependencies [23].

The Bayesian perspective extends the classical psychophysical concept of magnitude estimation by formally incorporating the influence of prior experience (Figure 2 and 3B). This not only offers an explanation for subjective biases in magnitude estimation, but, as explained in detail below, also allows for reconciling the two classical laws of psychophysics. In the examples above, we have addressed one of the simplest instantiations of a Bayesian perspective, that is, sequential belief updating under Gaussian assumptions about likelihood and prior. Notably, this views perception as an inferential process, as opposed to sequential stimulus-driven processing steps [30,46]. In brief, 'Bayesian brain' theories postulate that the brain constructs and continuously updates a so-called 'generative model', which provides the joint probability of hidden physical causes and sensory inputs (Box 3 and references therein). By inverting this generative model, the brain can
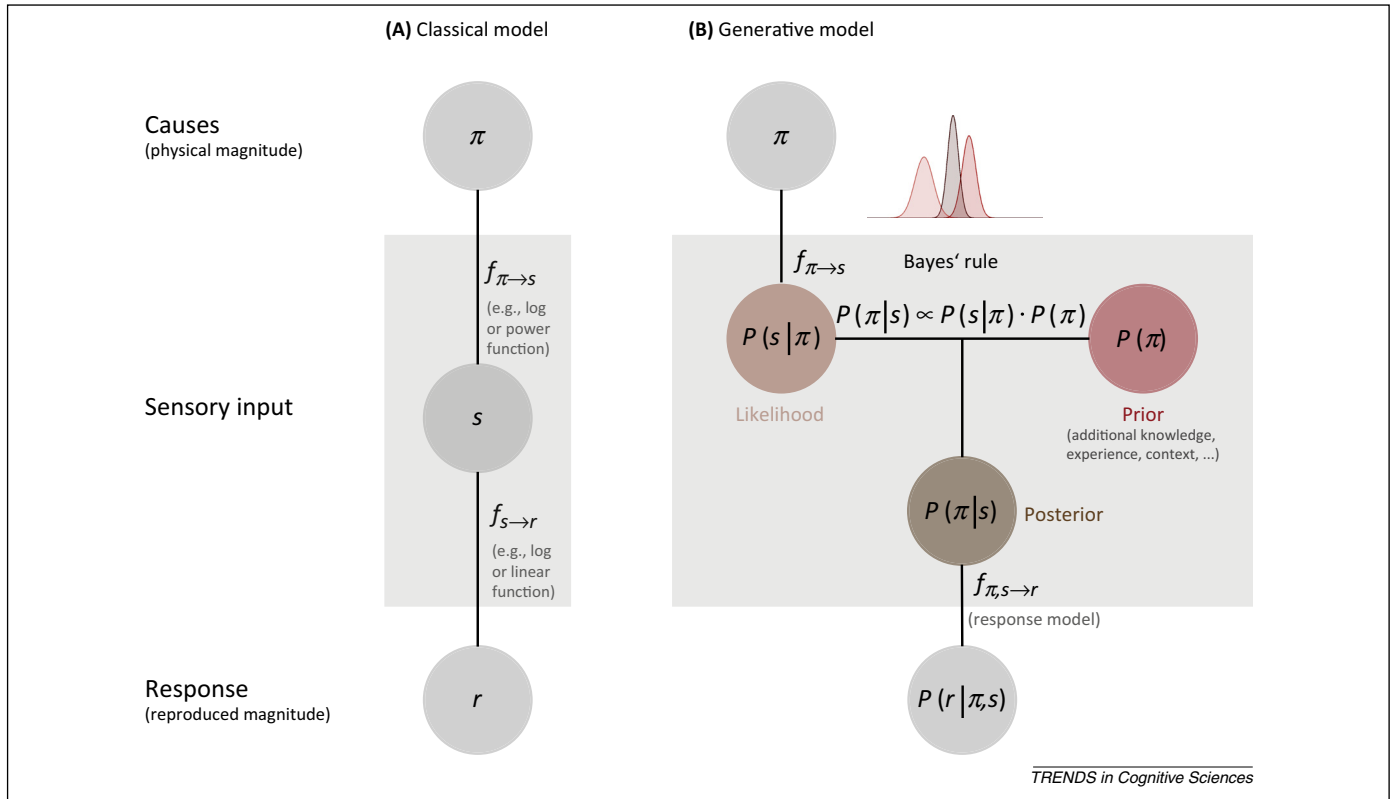
Figure 3. Classical and generative model of magnitude estimation. **(A)** A classical model relates causes such as physical stimulus magnitude to sensory input, which is then translated into a (motor) response, such as reproduced magnitude. The function relating the causes to a sensory representation as well as the function translating those into motor outputs could be of any form (logarithmic, power, etc.). **(B)** A generative model computes the joint probability of hidden causes and the observed data by taking the product of likelihood and prior. By inverting this generative model (i.e., computing the posterior probability of the states), the brain can solve the inverse problem of inferring the causes in the world based on current sensory input and a priori beliefs (additional knowledge, experience, context, etc.). The posterior estimate is then translated into a motor output via an appropriate response model.

compute the posterior probability and, thus, solve the inverse problem of inferring physical causes from sensory input and *a priori* beliefs (Figure 3B) [47].

The utility of this perspective for understanding magnitude estimation is illustrated by a recent model [19], where the likelihood of the sensory input given its physical cause was represented by a logarithmic transformation $f_{\pi \to s}$ according to the Weber-Fechner law (Equation 1):

$$P(s|\pi) = N(\mu_s, \sigma_s) = N(\log(\pi), \sigma_s) \qquad [1]$$

This likelihood function was then combined with a prior, also on a logarithmic scale, by taking the weighted average of the two information sources (Equation II in Box 3). To reproduce magnitudes on a linear scale, the model assumed that the resulting posterior magnitude estimates $P(\pi|s) = N(\mu_{\pi|s}, \sigma_{\pi|s})$ were transformed into motor responses using an exponential response model $f_{\pi,s \to r}$, giving a mean reproduced magnitude of (Equation 2):

$$r \propto e^{\mu_{\pi|s}} \propto e^{w_s \cdot \mu_s + w_\pi \cdot \mu_\pi} \qquad [2]$$

Given that the mean of the likelihood μ_s is a log-transform of the actual physical magnitude π, Equation 2 can be rewritten as power law relation between physical magnitude and response (Equation 3):

$$r \propto \pi^{w_s} \qquad [3]$$

In other words, regarding magnitude estimation as perceptual inference is capable of reconciling the seemingly incompatible theories by Weber-Fechner and Stevens: applying Bayes rule to a logarithmic representation of magnitude, according to Weber-Fechner's law, and assuming an appropriate belief-response mapping yields a power law dependency between physical and reproduced magnitudes on matching tasks, as proposed by Stevens' power law. Notably, discrimination tasks, which require only binary decisions about stimulus differences (bigger versus smaller), would not require the exponential transformation of posterior estimates to motor responses and, thus, yield the logarithmic relation suggested by Weber.

Importantly, the exponent in the Bayesian model has a different interpretation than Stevens originally proposed. Instead of representing a fixed property of a given sensory system, it captures a variable weighting of the sensory input in relation to the prior (Equation 2). This suggests that higher sensory uncertainty (noisier or more difficult stimuli) should lead to less weighting of the likelihood (sensed magnitude) and a stronger regression effect. Indeed an increase of the regression effect with increasing task difficulty has been reported across a series of studies [48–50].

Although the interpretation of the power law exponent as reflecting a weighting by sensory uncertainty provides convincing explanations of many empirical findings, it also faces a potential limitation. The weight of the likelihood is bounded between 0 and 1, but there are empirical findings from pain studies in which the estimated exponents were

reported to be larger than unity [51]. One potential explanation for this discrepancy could be cross-modality matching. Given that perceived pain intensity is not reproducible by motor acts but is reported on a numerical scale instead, magnitude reproduction requires the transformation from one modality to another [43,52]. A rescaling factor between the modalities would enter the exponent as a denominator and could explain power law exponents larger than one in a cross-modality matching setting.

A generative model not only serves to infer hidden states of the world from noisy sensory inputs, but, as implied by its name, to generate cognitive representations. Such synthesis of cognitive states on the basis of a generative model has been proposed, in particular, as a mechanism for imagery as well as hallucinations (e.g., [53,54]). Assuming a generative model as the basis for imagery would explain why perception and imagery of physical magnitudes share many behavioral effects, with the notable exception of sequential dependencies [55]. This absence of sequential dependency effects is immediately plausible under a generative model perspective: when magnitudes are purely imagined (by sampling from the prior to synthesize 'sensory inputs'), there is no prediction error and, thus, no need to update the prior.

The utility of a Bayesian framework for future psychophysical and neuroimaging studies

An ongoing debate concerns the question of whether different types of magnitude share a fundamental processing or representation system from which specialized metrics, such as time, space, or size, have evolved. This notion has been supported by behavioral studies showing analogous biases across modalities (discussed above) and from imaging studies that point towards a potentially universal representation of magnitude in the parietal cortex ([14], reviewed in [11,12,15]). However, it is possible that parietal encoding of magnitude is a downstream effect, reflecting a convergence of inputs from systems with specialized representations of different physical magnitudes (e.g., distance encoding by grid cells in the entorhinal cortex [1], or time encoding by cortico-thalamic-basal ganglia circuits [2]). In other words, parietal activation may reflect a multisensory combination of various information sources, leading to magnitude estimates that are not only a function of the stimulus itself, but also influenced by experience, context, or abstract knowledge [56–59]. Thus, understanding the nature of potentially universal indices of magnitude estimation systems faces similar conceptual problems as deciphering magnitude estimation within a given sensory system. Disentangling different aspects of magnitude representation, whether for a specific physical quantity or as a modality-independent phenomenon, will greatly benefit from computational modeling.

A Bayesian framework may prove useful for this endeavor, both conceptually and practically. Conceptually, it provides a framework for different processing stages, in which representation of a stimulus (likelihood), context (prior), and estimation or inference (posterior) of a physical magnitude may recruit different neuronal circuits. Selective experimental variations of these stages in neuroimaging experiments could reveal their unique or joint implementation in

terms of regional circuitry. A similar approach was adopted by a recent fMRI study on visual decision-making, which varied the uncertainties of prior and likelihood. This study found that likelihood uncertainty co-varied with activity in early sensory-processing areas, whereas prior uncertainty was reflected by activity in higher cognitive areas, such as putamen, insula, and orbitofrontal cortex [60]. Another study showed that the estimation stage in roughness judgments recruits areas different from those involved in pure representation [61].

In addition, the Bayesian concept may provide a simple explanation for universal features of magnitude estimation, which reoccur across modalities. This may be due to 'global' priors [29,62] that, in addition to 'local' (modality-specific) online-priors discussed above, reflect aspects of general environmental statistics, or cognitive contexts that elicit an additional influence on the individuals' weighting of prior and likelihood. Other domains of cognitive science have shown that stimulus judgments are biased not only by prior experience, but also by their affiliation to more global stimulus categories. These effects are observed in areas such as speech perception or episodic memory and could be explained by similar Bayesian accounts as discussed above [56,63,64]. For example, more complex contextual information could be captured by hierarchical Bayesian models in which higher levels determine estimates at lower levels (e.g., the influence of priors about environmental volatility on associative learning processes [65] or of abstract knowledge on object category judgments [66,67]).

Concerning magnitude estimation, the existence of global priors may explain intriguing findings such as the significant correlation of the subject-specific weighting across two independent tasks, distance estimation and turning angle estimation [19]. This finding is difficult to explain without reference to overarching contextual influences, given that there are distinct and specialized neuronal processing mechanisms for translational and rotational optic flow [68]. Similarly, other studies have reported within-subject correlations in power law exponents for circle size, number, and line length or weight, taste, and smell [69,70]. Finally, there are not only correlations across estimates of different magnitude types, but also direct intermodal influences on the estimation process. Quantities in the non-temporal domain, such as numerosity, size, or luminance, influence the estimation of time, such that larger stimuli are associated with longer durations [71,72]. Judgments of orientation, but not of color or shape, were influenced by the co-occurrence of task-irrelevant digits [73]. One explanation of the coupling between some magnitudes is that this might convey useful heuristics about the physical world [11,74], for example, longer distances usually take more time and larger objects are often heavier. In other words, the statistical co-occurrence of different magnitudes reliably reflects properties of the physical world that link different magnitudes such as time, space, or size, and a global prior reflecting these probabilistic relations might lead to coupled estimates of different physical magnitudes. In this case, it should be possible to induce novel coupling between the magnitude of quantities that are not systematically correlated in the physical world; *cf*. [75,76]. Some evidence for learning of

such an 'artificial' coupling between magnitude estimates comes from infant studies. When larger objects were presented with a different color-pattern than smaller ones, 9-month-old infants expected the same magnitude-color mapping to hold for numerosity and duration. This cross-dimensional transfer occurred bidirectionally for size, numerosity, and duration [77].

Practically, a Bayesian perspective on magnitude estimation enables new types of experiments and applications. For example, we can use subject-dependent and unit-less parameter estimates, such as the individual uncertainty-weighting of sensory inputs, and prior knowledge across different modalities for group-level analyses of neuroimaging data. This might allow one to examine more closely the source of global priors as a sign of shared representations. In future experiments, varying the uncertainty of either prior (e.g., by changing the distribution of training data) or likelihood (e.g., by changing signal to noise ratio of stimuli) could be used to parametrically modulate their individual influence and identify specific or joint circuits. Furthermore, using response models that are sensitive to contextual variations, trial-by-trial model predictions about performance could be used to assess regions sensitive to these contextual influences (see Box 4 for additional open questions).

Finally, it is worth noting that the explanatory power of Bayesian concepts of cognition has been questioned [78,79] and defended [80,81]. In the specific context of magnitude estimation, non-Bayesian approaches to incorporate prior experience have also been proposed [82,83]. At the purely behavioral level, it may be difficult to decide between different theories of cognitive representations. Eventually, this question may only be resolved fully at the level of the underlying neuronal code [84–87]. However, as discussed in this review, a Bayesian perspective on magnitude estimation is demonstrating considerable theoretical and practical utility; notably, this utility is also beginning to emerge in clinical applications.

The utility of a Bayesian framework for clinical studies

The Bayesian framework discussed in this review has important connections to empirical findings and theoretical concepts in clinical research. Empirically, aberrant magnitude estimation has been found in numerous neurological and psychiatric diseases. For example, distortions

of time estimation have been reported for Parkinson's disease, attention deficit hyperactivity disorder, autism, and schizophrenia [88,89]. Similarly, the susceptibility to visual illusions concerning size or length, such as the Ebbinghaus, Ponzo, or Müller-Lyer illusion, is altered in various diseases, including schizophrenia [22,90,91] and possibly autism [92,93].

Schizophrenia is of particular interest here, given that several contemporary theories of this disease postulate that a disturbance of hierarchical Bayesian inference represents the core computational abnormality of this disease (e.g. [94–96]), providing a direct connection to the framework discussed in this paper. Empirically, several studies have demonstrated prominent disturbances of physical magnitude estimation in schizophrenia, for example, on tasks probing force matching [97,98] or time discrimination [99]. Notably, hierarchical Bayesian models can not only explain the empirically observed force matching deficits in schizophrenia [100], but also provide several testable predictions for future studies on magnitude estimation in schizophrenia, with putative links to neurobiological mechanisms, such as the encoding of belief precision by neuromodulatory transmitters [101]. Also, Bayesian modeling of magnitude estimation combined with neuroimaging may allow for defining subgroups of patients with different impairments in perceptual inference (*cf*. [102]). Overall, given the importance of perceptual inference for pathophysiological theories of schizophrenia and the availability of suitable Bayesian models, it would be promising to translate more basic psychophysical paradigms of magnitude estimation to schizophrenia research (Box 4).

Concluding remarks

In this paper, we have revisited magnitude estimation, one of the oldest topics in psychophysics, and re-examined it under the general framework of Bayesian inference. This Bayesian perspective suggests a generic and principled mechanism for perceptual inference, providing a modality-independent explanation of a wide set of empirical findings on magnitude estimation. Furthermore, we have demonstrated how this view can help reconciling the seemingly incompatible laws by Weber-Fechner and Stevens and provide us with a re-interpretation of the Stevens power law exponent as a quantity that captures the weight of sensory uncertainty relative to a priori knowledge. We anticipate that the modeling framework reviewed here might be useful for distinguishing between shared and selective representations of magnitude in neuroimaging experiments and how they relate to cognitive distortions observed in psychiatric disorders.

> **Box 4. Outstanding questions**
>
> - Do links between magnitude estimation across modalities result from statistical co-occurrences of different magnitudes types in the world, which are captured by 'global priors'?
> - What are the exact environmental statistics that are represented by these global priors?
> - Is the weighting of priors (i.e., their precision) fixed or can it be learned depending on the reliability of previous experience?
> - Is there a way to 're-train' global priors and, thus, alter the cross-modality links in magnitude estimation?
> - Do global priors explain overlapping activations in neuroimaging studies of different types of magnitude estimation?
> - Can deficits in magnitude estimation be used as diagnostic probes in mental disease with putative abnormalities in Bayesian inference, such as schizophrenia?

References

1 McNaughton, B.L. *et al.* (2006) Path integration and the neural basis of the 'cognitive map'. *Nat. Rev. Neurosci.* 7, 663–678
2 Merchant, H. *et al.* (2013) Neural basis of the perception and estimation of time. *Annu. Rev. Neurosci.* 36, 313–336

3 Stevens, S.S. and Galanter, E.H. (1957) Ratio scales and category scales for a dozen perceptual continua. *J. Exp. Psychol. Gen.* 54, 377–411

4 Dehaene, S. and Brannon, E.M. (2010) Space, time, and number: a Kantian research program. *Trends Cogn. Sci.* 14, 517–519

5 Fechner, G.T. (1860) *Elemente der Psychophysik*, Breitkopf und Härtel

6 Stevens, S.S. (1961) To honor Fechner and repeal his law: a power function, not a log function, describes the operating characteristic of a sensory system. *Science* 133, 80–86

7 Hollingworth, H.L. (1910) The central tendency of judgment. *J. Philos. Psychol. Sci. Methods* 7, 461–469

8 Gibbon, J. (1977) Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* 84, 279–325

9 DeCarlo, L.T. and Cross, D.V. (1990) Sequential effects in magnitude scaling: models and theory. *J. Exp. Psychol. Gen.* 119, 375–396

10 Hubbard, E.M. *et al.* (2005) Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* 6, 435–448

11 Bueti, D. and Walsh, V. (2009) The parietal cortex and the representation of time, space, number and other magnitudes. *Philos. Trans. R. Soc. Biol. Sci.* 364, 1831–1840

12 Walsh, V. (2003) A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488

13 Pinel, P. *et al.* (2004) Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron* 41, 983–993

14 Eger, E. *et al.* (2003) A supramodal number representation in human intraparietal cortex. *Neuron* 37, 719–725

15 Cohen Kadosh, R. *et al.* (2008) Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Prog. Neurobiol.* 84, 132–147

16 Cantlon, J.F. *et al.* (2009) Comment on 'Log or Linear? Distinct intuitions of number scale in Western and Amazonian indigene cultures. *Science* 323, 1–2

17 Lappe, M. *et al.* (2007) Travel distance estimation from visual motion by leaky path integration. *Exp. Brain Res.* 180, 35–48

18 Jazayeri, M. and Shadlen, M.N. (2010) Temporal context calibrates interval timing. *Nat. Neurosci.* 13, 1020–1026

19 Petzschner, F.H. and Glasauer, S. (2011) Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration. *J. Neurosci.* 31, 17220–17229

20 Jürgens, R. and Becker, W. (2011) Human spatial orientation in non-stationary environments: relation between self-turning perception and detection of surround motion. *Exp. Brain Res.* 215, 327–344

21 Acerbi, L. *et al.* (2012) Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Comput. Biol.* 8, e1002771

22 Notredame, C-E. *et al.* (2014) What visual illusions teach us about schizophrenia. *Front. Integr. Neurosci.* 8, 1–16

23 Poulton, E.C. (1968) The new psychophysics: Six models for magnitude estimation. *Psychol. Bull.* 69, 1–19

24 Dehaene, S. (2003) The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends Cogn. Sci.* 7, 145–147

25 Laming, D. (1999) Prior expectations in cross-modality matching. *Math. Soc. Sci.* 38, 343–359

26 Shi, Z. *et al.* (2013) Bayesian optimization of time perception. *Trends Cogn. Sci.* 17, 556–564

27 Ernst, M.O. and Bülthoff, H.H. (2004) Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169

28 Ernst, M.O. and Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433

29 Adams, W.J. (2008) Frames of reference for the light-from-above prior in visual search and shape judgements. *Cognition* 107, 137–150

30 Knill, D.C. and Richards, W. (1996) *Perception as Bayesian Inference*, Cambridge University Press

31 Jürgens, R. and Becker, W. (2006) Perception of angular displacement without landmarks: evidence for Bayesian fusion of vestibular, optokinetic, podokinesthetic, and cognitive information. *Exp. Brain Res.* 174, 528–543

32 Vilares, I. and Kording, K. (2011) Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Ann. N. Y. Acad. Sci.* 1224, 22–39

33 Griffiths, T.L. and Tenenbaum, J.B. (2011) Predicting the future as Bayesian inference: people combine prior knowledge with observations when estimating duration and extent. *J. Exp. Psychol. Gen.* 140, 725–743

34 Lucas, C.G. and Griffiths, T.L. (2010) Learning the form of causal relationships using hierarchical bayesian models. *Cogn. Sci.* 34, 113–147

35 Colas, F. *et al.* (2010) Common Bayesian models for common cognitive issues. *Acta Biotheor.* 58, 191–216

36 Anobile, G. *et al.* (2012) Linear mapping of numbers onto space requires attention. *Cognition* 122, 454–459

37 Berniker, M. *et al.* (2010) Learning priors for Bayesian computations in the nervous system. *PLoS ONE* 5, e12686

38 Cicchini, G.M. *et al.* (2014) Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7867–7872

39 Verstynen, T. and Sabes, P.N. (2011) How each movement changes the next: an experimental and theoretical study of fast adaptive priors in reaching. *J. Neurosci.* 31, 10050–10059

40 Sun, J.Z. *et al.* (2012) A framework for Bayesian optimality of psychophysical laws. *J. Math. Psychol.* 56, 495–501

41 MacKay, D.M. (1963) Psychophysics of perceived intensity: a theoretical basis for Fechner's and Stevens' laws. *Science* 139, 1213–1216

42 Shepard, R.N. (1981) On the status of 'direct' and psychophysical scales: psychophysical measurement. *J. Math. Psychol.* 24, 21–57

43 Attneave, F. (1962) Perception and related areas. In *Psychology: A Study of a Science* (Koch, S., ed.), pp. 619–659, McGraw-Hill

44 Krueger, L.E. (1989) Reconciling Fechner and Stevens: toward a unified psychophysical law. *Behav. Brain Sci.* 12, 251–320

45 Gescheider, G.a (1988) Psychophysical scaling. *Annu. Rev. Psychol.* 39, 169–200

46 Körding, K.P. *et al.* (2007) Causal Inference in Multisensory Perception. *PLoS ONE* 2, e943

47 Pizlo, Z. (2001) Perception viewed as an inverse problem. *Vision Res.* 41, 3145–3161

48 Kappauf, W.E. (1975) Regression effect in judgments to determine equal response contours. *Percept. Psychophys.* 17, 405–410

49 Sheldon, P.E. (1973) Equal-onset contours of vibrotactile stimuli. *Percept. Psychophys.* 13, 403–407

50 Stevens, S.S. and Greenbaum, H.B. (1966) Regression effect in psychophysical judgment. *Percept. Psychophys.* 1, 439–446

51 Stevens, S.S. (1960) The psychophysics of sensory function. *Am. Sci.* 48, 226–253

52 Ekman, G. (1964) Is the power law a special case of Fecher's law? *Percept. Mot. Ski.* 19, 730

53 Burgess, N. and Becker, S. (2001) Modelling spatial recall, mental imagery and neglect. In *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference* (Leen, T.K., Dietterich, T.G. and Tresp, V., eds), pp. 96–102, MIT Press

54 Reichert, D.P. *et al.* (2013) Charles Bonnet Syndrome: evidence for a generative model in the cortex? *PLoS Comput. Biol.* 9, e1003134

55 Baird, J.C. and Harder, K.A. (2000) The psychophysics of imagery. *Percept Psychophys* 62, 113–126

56 Petzschner, F.H. *et al.* (2012) Combining symbolic cues with sensory input and prior experience in an iterative Bayesian framework. *Front. Integr. Neurosci.* 6, 1–18

57 Tcheang, L. *et al.* (2011) Visual influence on path integration in darkness indicates a multimodal representation of large-scale space. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1152–1157

58 Sereno, M.I. and Huang, R-S. (2014) Multisensory maps in parietal cortex. *Curr. Opin. Neurobiol.* 24, 39–46

59 Cisek, P. and Kalaska, J.F. (2010) Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* 33, 269–298

60 Vilares, I. *et al.* (2012) Differential representations of prior and likelihood uncertainty in the human brain. *Curr. Biol.* 22, 1641–1648

61 Kitada, R. *et al.* (2005) Tactile estimation of the roughness of gratings yields a graded response in the human brain: an fMRI study. *Neuroimage* 25, 90–100

62 Berkes, P. (2011) Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331, 83–87

63 Huttenlocher, J. *et al.* (2000) Why do categories affect stimulus judgment? *J. Exp. Psychol. Gen.* 129, 220–241

64 Feldman, N.H. *et al.* (2009) The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* 116, 752–782

65 Mathys, C. *et al.* (2011) A Bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* 5, 39

66 Hemmer, P. and Steyvers, M. (2009) A Bayesian account of reconstructive memory. *Top. Cogn. Sci.* 1, 189–202

67 Hemmer, P. and Steyvers, M. (2009) Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychon. Bull. Rev.* 16, 80–87

68 Billino, J. *et al.* (2008) Differential aging of motion processing mechanisms: evidence against general perceptual decline. *Vision Res.* 48, 1254–1261

69 Jones, F.N. and Marcus, M.J. (1961) The subject effect in judgments of subjective magnitude. *J. Exp. Psychol.* 61, 40–44

70 Robinson, G.H. (1976) Biasing power law exponents by magnitude estimation instructions. *Percept. Psychophys.* 19, 80–84

71 Xuan, B. *et al.* (2007) Larger stimuli are judged to last longer. *J. Vis.* 7, 1–5

72 Oliveri, M. *et al.* (2008) Perceiving numbers alters time perception. *Neurosci. Lett.* 438, 308–311

73 Fias, W. *et al.* (2001) Irrelevant digits affect feature-based attention depending on the overlap of neural circuits. *Brain Res. Cogn. Brain Res.* 12, 415–423

74 Glasauer, S. *et al.* (2007) Space-time relativity in self-motion reproduction. *J. Neurophysiol.* 97, 451–461

75 Adams, W.J. *et al.* (2004) Experience can change the "light-from-above" prior. *Nat. Neurosci.* 7, 1057–1058

76 Ernst, M.O. (2007) Learning to integrate arbitrary signals from vision and touch. *J. Vis.* 7, 1–14

77 Lourenco, S.F. and Longo, M.R. (2010) General magnitude representation in human infants. *Psychol. Sci.* 21, 873–881

78 Jones, M. and Love, B.C. (2011) Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behav. Brain Sci.* 34, 169–188

79 Bowers, J.S. and Davis, C.J. (2012) Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* 138, 389–414

80 Griffiths, T.L. *et al.* (2012) How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). *J. Exp. Psychol. Gen.* 138, 415–422

81 Griffiths, T.L. *et al.* (2010) Probabilistic models of cognition: exploring representations and inductive biases. *Trends Cogn. Sci.* 14, 357–364

82 Howe, C.Q. *et al.* (2006) Comparison of Bayesian and empirical ranking approaches to visual perception. *J. Theor. Biol.* 241, 866–875

83 Howe, C.Q. and Purves, D. (2002) Range image statistics can explain the anomalous perception of length. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13184–13188

84 Hinton, G.E. (2007) Learning multiple layers of representation. *Trends Cogn. Sci.* 11, 428–434

85 Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 20, 1434–1448

86 Ma, W.J. *et al.* (2008) Spiking networks for Bayesian inference and choice. *Curr. Opin. Neurobiol.* 18, 217–222

87 Friston, K. (2010) The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138

88 Allman, M.J. and Meck, W.H. (2012) Pathophysiological distortions in time perception and timed performance. *Brain* 135, 656–677

89 Smith, J.G. *et al.* (2007) The effect of Parkinson's disease on time estimation as a function of stimulus duration range and modality. *Brain Cogn.* 64, 130–143

90 Weckowicz, T. and Witney, G. (1960) The Müller-Lyer Illusion in schizophrenic patients. *Br. J. Psychiatry* 106, 1002–1007

91 Kantrowitz, J.T. *et al.* (2009) Seeing the world dimly: the impact of early visual deficits on visual experience in schizophrenia. *Schizophr. Bull.* 35, 1085–1094

92 Chouinard, P.A. *et al.* (2013) Global processing during the Müller-Lyer illusion is distinctively affected by the degree of autistic traits in the typical population. *Exp. Brain Res.* 230, 219–231

93 Walter, E. *et al.* (2009) A specific autistic trait that modulates visuospatial illusion susceptibility. *J. Autism Dev. Disord.* 39, 339–349

94 Stephan, K.E. *et al.* (2006) Synaptic plasticity and dysconnection in schizophrenia. *Biol. Psychiatry* 59, 929–939

95 Fletcher, P.C. and Frith, C.D. (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58

96 Jardri, R. and Denève, S. (2013) Circular inferences in schizophrenia. *Brain* 136, 3227–3241

97 Shergill, S.S. *et al.* (2005) Evidence for sensory prediction deficits in schizophrenia. *Am. J. Psychiatry* 162, 2384–2386

98 Shergill, S.S. *et al.* (2014) Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. *JAMA Psychiatry* 71, 28–35

99 Waters, F. and Jablensky, A. (2009) Time discrimination deficits in schizophrenia patients with first-rank (passivity) symptoms. *Psychiatry Res.* 167, 12–20

100 Brown, H. *et al.* (2013) Active inference, sensory attenuation and illusions. *Cogn. Process.* 14, 411–427

101 Adams, R.A. *et al.* (2013) The computational anatomy of psychosis. *Front. Psychiatry* 4, 1–26

102 Schmack, K. *et al.* (2013) Delusions and the role of beliefs in perceptual inference. *J. Neurosci.* 33, 13701–13712

103 Laming, D. (1997) *The Measurement of Sensation*, Oxford University Press

104 Hollingworth, H.L. (1909) The inaccuracy of movement. *Arch. Psychol.* 2, 1–87

105 Teghtsoonian, R. and Teghtsoonian, M. (1978) Range and regression effects in magnitude scaling. *Percept. Psychophys.* 24, 305–314

106 Dehaene, S. *et al.* (1998) Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* 21, 355–361

107 Dehaene, S. *et al.* (2008) Log or linear? Distinct intuitions of the number scale in Western and Amazonian Indigene cultures. *Science* 320, 1217–1220

108 Nieder, A. and Miller, E.K. (2003) Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 37, 149–157

109 Nover, H. *et al.* (2005) A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J. Neurosci.* 25, 10049–10060

110 Cross, D. V (1973) Sequential dependencies and regression in psychophysical judgments. *Percept. Psychophys.* 14, 547–552

111 Weber, E.H. (1850) Der Tastsinn und das Gemeingefühl. In *In Handwörterbuch der Physiologie* (Wagner, R., ed.), pp. 481–588, Vieweg

112 Helmholtz, H. (1867) Concerning the perceptions in general. In *Readings in the History of Psychology. Century Psychology Series*. pp. 214–230, East Norwalk, CT, US, Appleton-Century-Crofts

113 Dayan, P. *et al.* (1995) The Helmholtz machine. *Neural Comput.* 7, 889–904

114 Kersten, D. *et al.* (2004) Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304

115 Doya, K. *et al.* (2011) *Bayesian Brain: Probabilistic Approaches to Neural Coding*, MIT Press