nature neuroscience

REVIEW ARTICLE
https://doi.org/10.1038/s41593-019-0340-4

Optimality and heuristics in perceptual neuroscience

Justin L. Gardner

The foundation for modern understanding of how we make perceptual decisions about what we see or where to look comes from considering the optimal way to perform these behaviors. While statistical computation is useful for deriving the optimal solution to a perceptual problem, optimality requires perfect knowledge of priors and often complex computation. Accumulating evidence, however, suggests that optimal perceptual goals can be achieved or approximated more simply by human observers using heuristic approaches. Perceptual neuroscientists captivated by optimal explanations of sensory behaviors will fail in their search for the neural circuits and cortical processes that implement an optimal computation whenever that behavior is actually achieved through heuristics. This article provides a cross-disciplinary review of decision-making with the aim of building perceptual theory that uses optimality to set the computational goals for perceptual behavior but, through consideration of ecological, computational, and energetic constraints, incorporates how these optimal goals can be achieved through heuristic approximation.

Making decisions can be messy. Consider deciding whether to take a job in a new city, whether to accept an applicant to a graduate program, or whether a company should invest in a new factory. Multiple, seemingly incommensurate, predictors of outcomes must be weighed against each other, and potential benefits and costs of the choice must be considered against the likelihood of their occurrence. Fortunately, decision theory gives us clear, quantitative prescriptions for how to make optimal decisions. Unfortunately, people generally fail to make decisions that comply with optimal theory. Psychologists over the last several decades have reveled in pointing out the many ways in which people's decision-making diverge from such optimality, giving rise to cognitive illusions and substantial bias[1,2]. Many of the lessons from this literature on judgement and decision-making have found their way into the neuroscience of economic or value-based decision-making, where biases are routinely captured by making probabilities subjective or by accepting that gains and losses are not symmetric. But what about perceptual decision-making that reflects what a decision-maker senses about an objective, external reality, rather than decisions based on internal, possibly idiosyncratic, preferences? Neural mechanisms of perceptual decision-making have had the opportunity to be calibrated against the statistics of natural environments over evolutionary history and over the literally thousands of daily perceptual decisions we make as to where to look and how to interpret what we see. These sorts of decisions have been quite fruitfully formulated in optimal terms in which evidence and outcomes can be quantitatively and precisely stated. Nonetheless, recent evidence suggests something messier lurking behind the curtain of optimality that must be acknowledged if we are to uncover the neural mechanisms underlying perceptual decision-making.

Optimality in perception

A crowning achievement of optimality theory in perception is the application of principles of statistical decision-making, in the form of signal detection theory, to psychophysical measurement[3–5]. Signal detection theory tamed a vexing problem for psychophysics: that detection of a signal in noise could be governed not just by sensory sensitivity, but a somewhat inconvenient and messy cognitive factor or criterion (Box 1). Detection theory was validated by showing that subjects could be induced to change their false-alarm rate by changing the probability of target presence and the relative expected value associated with hits versus correct rejects[3,6]. In contradiction to prevailing assumptions that false-alarm rates were simply guesses when sensory responses did not reach a high-enough threshold, signal detection theory asserted that false alarms occurred when occasional noise rose above criterion and thus correctly predicted a negative dependency between false-alarm and miss rates[3,4,7]. Performance on a two-alternative forced-choice task could be predicted from the area under the receiver operating characteristic curve measured from a detection task[8], which is particularly elegantly when subjects used a ratings scale from which a full receiver operating characteristic curve can be easily constructed[9]. These findings developed into the now familiar procedures for computing sensory sensitivity, or d', independently of criterion and thus, seemingly, squarely defeating the problem of uncontrolled criterion in psychophysical measurement.

Beyond this methodological contribution, signal detection theory provided a framework for testing human perceptual behavior against optimal benchmarks. The abscissa of the signal detection graph was conceived of as representing a quantity monotonically proportional to the likelihood ratio of signal presence compared to absence, with the assumption of independent equal variance Gaussian distributions. The optimal criterion then could be computed based on prior probability and expected value (Box 1) and human performance tested against these optimal expectations with reasonably good match[3,5] (but see below). The likelihood ratio for a signal detection task was modeled by an ideal observer, i.e., an optimal model, typically a matched filter for the signal that could be correlated with the noisy stimulus. Considerations of the deviations in performance from the ideal-observer models led, among other things, to the concept that human subjects may not know precisely the template of the signal they are asked to detect[5] and that this sort of uncertainty might explain various aspects of perceptual behavior[10–12]. This idea has been echoed more recently in suggestions that behavioral variability might

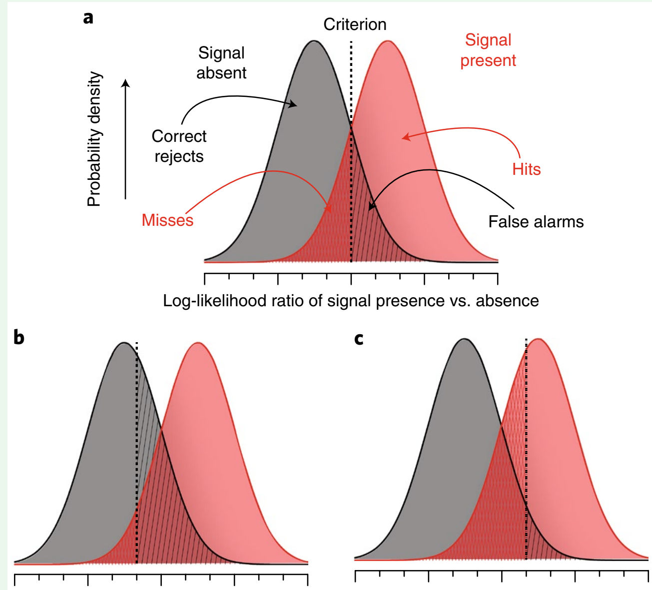
Department of Psychology, Stanford University, Stanford, California, USA. e-mail: jlg@stanford.edu

Box 1 | Signal detection theory as an optimal theory of visual perception

When asked to detect the presence or absence of a signal, such as a dim light, a spatial pattern, or the presence of a bomb or a tumor in an X-ray image, one can consider the evidence or likelihood ratio of presence versus absence (**a**, abscissa is proportional to log likelihood ratio) where actual signal presence (red curve) and absence (black curve) are associated with different probability distributions. An arbitrary criterion (dotted line) for deciding the threshold likelihood ratio for which to report presence results in different proportions of hits (signal correctly defined as present) or correct rejects (signal correctly defined as absent) and corresponding errors (misses and false alarms). Criterion is a key cognitive factor because if it is set too conservatively, the presence of a bomb could be missed, but if it is set too liberally, a patient may suffer from undue worrying about cancer when none is present. Signal detection theory provides an optimal formulation[3] for setting criterion β

$$\beta = \frac{p\,(signal\ \ absent)}{p\,(signal\ \ present)} \cdot \frac{V\,(correct\ \ reject) + K\,(false\ \ alarms)}{V\,(hit) + K\,(miss)}$$

Where p is the prior probability of signal presence or absence, V is the value of correct answers, and K is the cost of errors. By setting the criterion to β, the observer can optimize their expected gain. Thus, for an ideal observer maximizing expected gain, conditions in which prior probability of signal presence, value of hits, and cost of misses are high all promote shifting criterion to the left (**b**) and conditions where prior probability of signal absence, value of correct rejects and cost of false alarms are high promote shifting criterion to the right (**c**).



Signal detection theory and optimal criterion. Signal detection distributions (**a**) with optimal criterion for high (**b**) or low (**c**) prior probability for signal presence, value of hits and cost of misses.

be less the outcome of noisy neural representation and more suboptimal inference, akin to a mismatch of the perceptual template with the stimulus to be detected[13]. A straight line can be drawn from the development of the optimality constructs of signal detection theory to many key developments in perception, including those in visual search[12,14], mechanisms for attention[15–19], the optimal combination of different sensory cues[20], off-channel looking[21–23], signal integration[24,25], the neural basis of perceptual decision-making[26,27], and optimal accumulation of sensory evidence[28–30].

Different considerations of optimality, namely how to maximize the efficiency of information representation and transmission in neural systems, led to considerable progress in explaining different aspects of neural representation of perceptual stimuli. Barlow[31] highlighted that if firing an action potential is a cost to the nervous system, an efficient representation is one that assigns the smallest number of spikes to the most common stimulus. This optimal coding hypothesis firmly placed environmental statistics at the forefront for understanding neural codes. Indeed, imposing sparsity constraints on representing natural scenes was found to produce simple cell-like representations[32]. Measurements of natural scene depth helped to explain space perception[33]. Orientation distributions in natural scenes explained perceptual biases and early sensory representations[34]. Perception of a variety of motion illusions could be elegantly explained by a prior that biases speed perception toward slow[35–38]. Prior probability of a number of image statistics could be shown to be implicitly coded in neural representations[39]. Retinal representations could be viewed as whitening the redundancies of natural stimuli[40]. The distributions of different illuminants and surfaces provided explanations for color constancy performance[41]. The development of rods and cones in the retina could be explained by statistics of environmental image ensembles[42]. Contour grouping was predicted by edge co-occurrence in natural images[43]. The fundamental insight of considering optimal representations is the governing idea behind predictive coding models of visual processing[44] in which, for example, end-stopping in primary visual cortex is interpreted as a mechanism to suppress the unnecessary transmission of redundant sensory information[45]. Environmental statistics are the basis for priors that guide perceptual inference in Bayesian accounts of sensory perception[46,47] (Fig. 1).

Importantly, the utility of optimality theory in perception was not originally conceived as a test of behavior as being optimal per se. Recent debates, particularly around Bayesian theory, have argued that assumptions in optimal models—for example, the priors in an inference model—can be adjusted to make any behavior conform (or not conform) to optimality[48–50]. Conversely, many perceptual behaviors do not reach optimal standards[51]. But asking whether human perception is optimal or suboptimal was never the goal of optimal observer analysis. Indeed, Swets et al. articulate the goals as follows:

...whereas it is not expected that the human observer and the ideal detection device will behave identically, the emphasis in early studies is on similarities. If the differences are small, one may rule out entire classes of alternative models, and regard the model in question as a useful tool in further studies. Proceeding on this assumption, one may then in later studies emphasize the differences, the form and extent of the differences suggesting how the ideal model may be modified in the direction of reality. (p. 311)[3]

This notion that ideal observers are useful in ruling out alternative models and in leading to new hypotheses when discrepancies are inevitably found has been echoed by many proponents of optimality analysis since[52–55]. In this view, considerations of optimality provide a quantitative framework for expressing hypotheses that generate testable predictions.

Heuristics

At about the same time that signal detection theory and ideal observer analysis were demonstrating the great value of building a

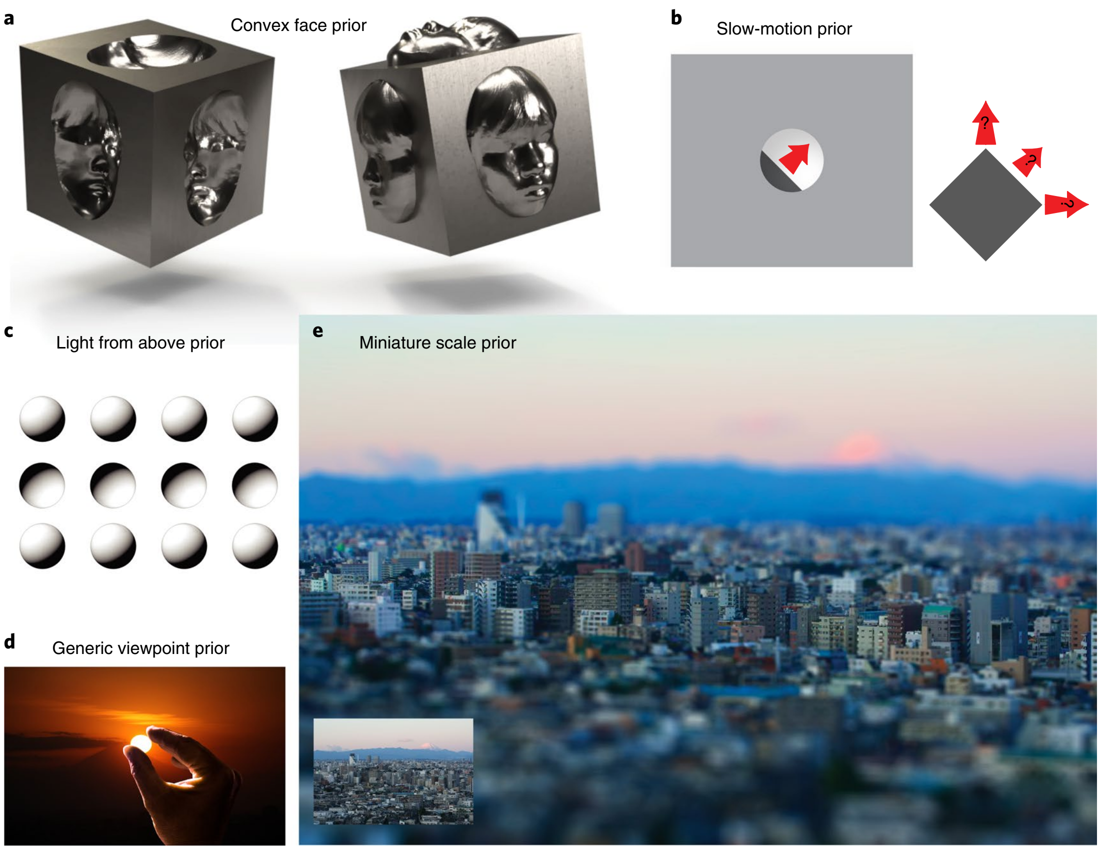


Fig. 1 | Visual illusions suggestive of optimality. The transduction of light into neural signals does not provide direct information about many aspects of the visual world, including three-dimensional (3D) structure, scale, and motion. Thus, neural mechanisms need to infer these properties, a process that Helmholtz[145] described as "unconscious inference." Visual illusions highlight this inferential nature of vision and have been championed as evidence for optimal Bayesian inference in vision[46,47]. **a**, For example, a concave (or hollow) mask appears convex (bottom two faces on left cube appear convex even though all three faces are concave, as can be seen on the top face). This hollow mask illusion[46] can be construed as interpreting the ambiguous 3D structure using a prior for convex facial forms. **b**, An object moving behind an aperture (left) is consistent with many different directions of motion (right), but observers see this motion as moving perpendicular to the viewable edge, which is the direction consistent with the slowest speed. This and other illusions of motion can be explained by the use of a prior for slow motion when motion signals are ambiguous[35–38]. **c**, Shading cues make the bumps in the middle appear as depressed into the page and the others as coming out of the page. This suggests the usage of a prior that light is coming from above[146], since the bumps are all rotated copies of each other and are equally interpretable in the opposite 3D configuration if light were coming from below. **d**, Forced perspective photographs can achieve depth and scale illusions by carefully arranging a foreground object with respect to a background object to make it look, for example, like a huge hand is grasping a tiny sun. These illusions exploit inferences based on perspective, lighting, color, and size cues, as well as a bias for observers to reject the veridical interpretation of the scene, which requires viewing from a special vantage point rather than a generic one. This suggests a prior for a generic viewpoint and lighting angle[147,148]. **e**, The apparent scale of a photograph can be made to appear miniature (compare to original, inset) by blurring parts of the image to mimic the narrow plane of focus obtained by photographing a miniature model and increasing the color saturation and contrast. This suggests a sophisticated use of priors by the visual system to disambiguate scale, one presumably learned not over evolutionary history, but over encountering photographs of toy models. While these illusions are all suggestive of the inferential nature of vision, it should be noted that they do not necessarily require probabilistic computation. Moreover, even if Bayesian computation is employed, the computations may not be optimal if the priors used are approximate and not veridical[149].

rational choice theory in perceptual psychology, developments in economic or value-based decision-making were busy tearing down the edifices of optimality in the form of viewing humans as rational economic actors. Rational choice theory[56] assumed that humans would consider the utility (typically a concave function of wealth[57]) of potential outcomes multiplied by their probability of occurence and make optimal decisions by choosing the option which provided the highest expected utility. Yet humans make paradoxical choices that violate even simple and seemingly unobjectionable axioms of rational choice theory[58]. While probabilities of 0 or 1 are not mathematically special, humans treat sure gains as more desirable than probabilistic gambles of equal expected value, thus demonstrating risk-aversion[57,59]. Losses are typically treated oppositely; humans choose more risky gambles to try to avoid sure losses[60]. While risk attitudes of this form could be accommodated by expected utility theory, the framing or reference point[2] of gambles can dictate whether a decision-maker treats an economic choice as a potential gain or loss, radically changing preferences despite there being no overall change in what is at stake. Transitivity of choice options, an axiomatic requirement[61] for using expected utility to model choices, can be violated, particularly when decisions are made using multiple incommensurate criteria[62]. Even more fundamentally, real-world decisions often require decision-makers to seek out alternatives, for example when buying a house. Considering the cost associated with finding alternatives, Herbert Simon abandoned optimality in terms of maximizing gains in favor of a decision theory which seeks to find not the best solution, but one that "satisfices" the criterion that the decision-maker hopes to achieve:

Box 2 | Cognitive illusions suggestive of heuristics

While visual illusions have often been interpreted in an optimal framework, cognitive illusions that were meant as analogies to visual illusions point out ways in which humans behave in a heuristic fashion by neglecting basic properties of statistical decision-making. (**a**) For example, base-rate neglect was demonstrated in a series of experiments[64] in which subjects were asked whether the descriptions given were of an engineer or a lawyer and the instructions given clearly state the relevant base rates (i.e., the prior probability of being an engineer or lawyer, 30/70 or 70/30 for different subject groups). When subjects were given descriptions such as Example 1, they found the description more representative of an engineer, regardless of whether they had been told that engineers or lawyers had a higher base rate. While this might be expected of Bayesian reasoning if subjects believed the description to be perfectly reliable, when asked about descriptions like Example 2, in which subjects rated the description as uninformative, subjects tended to report 50% without respect to the base rate. But see ref. [72]. (**b**) Another well-known cognitive illusion is the conjunction fallacy. Undergraduates were found to rate the probability of option 6, that Linda is a bank teller, lower than that of option 8, even though the conjunction of a being a bank teller and active in the feminist movement by basic rules of probability must be lower[150]. But see ref. [71].

(a) Base-rate neglect

Instructions: A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100.

Example 1: "Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles."

Example 2: "Dick is a 30-year-old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues."

(b) Conjunction fallacy

"Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

Linda 1) is a teacher in elementary school 2) works in a bookstore and takes Yoga classes 3) is active in the feminist movement 4) is a psychiatric social worker 5) is a member of the League of Women Voters 6) is a bank teller 7) is an insurance salesperson 8) is a bank teller and is active in the feminist movement

"decision makers can satisfice either by finding optimum solutions for a simplified world, or by finding satisfactory solutions for a more realistic world" (p 498)[63]

In short, optimality theory faced strong empirical challenges to the view of humans as rational economic actors when confronted with actual human behavior compared to theoretic ideals.

Optimality theory in nonperceptual decision-making has also been clearly threatened when it comes to making explicitly cognitive decisions involving probabilities. Daniel Kahneman and Amos Tversky famously demonstrated a number of cognitive illusions (Box 2) in which people demonstrate base-rate neglect. That is, they readily neglect prior probabilities in assigning probabilities to events, even when the priors are clearly given, the person knows the base rates, and the evidence given is demonstrably uninformative[64]. This base-rate neglect phenomenon is important because it affects how we evaluate all sorts of information, from genetic tests in which a prior probability of a disease must be considered to the validity of inferences that can be made with new evidence from scientific studies. Even when human subjects correctly use prior information, they tend not to update their decision criteria enough to account for new evidence, a phenomenon that was labeled[65,66] as being a "conservative Bayesian." People, even trained scientists, make a number of stereotyped errors in trying to evaluate the probability of random events[67], particularly when sample sizes are small[68]. Moreover, people have systematic biases when trying to generate random numbers[69]. These and other ways in which humans fail to use optimal statistical reasoning were extensively documented, and broad categories of these failings were labeled as different forms of heuristic, rather than optimal, reasoning[1,2]. It is worthwhile to note that there have been critiques of some of the conclusions of biases, based on what the meaning of probability is for a single event, how framing of questions can make some of the cognitive illusions disappear, and other considerations that might make biased behavior appear more rational[70–72]. Pulling back from these specific critiques, it should be obvious that there can be no single characterization of human behavior as being optimal or biased, as humans can perform the same behavioral task in different ways depending on context and task demands.

While optimal formulations of complex humanlike cognitive functions have been ascendant in cognitive science, many have corresponding heuristic solutions. Bayesian statistical algorithms have become increasingly powerful, able to solve reasoning problems such as one-shot classification of handwritten characters[73], linguistic communication[74] and learning[75], discovering structure in data[76], assigning responsibility for the consequences of another's actions[77], categorical perception[78], inductive reasoning[79], time perception[80], and sensorimotor integration[81]. While these algorithms elegantly find optimal solutions, underneath the hood are heavy computations that estimate posterior distributions using techniques like Gibbs sampling and the Metropolis–Hastings algorithm. While some have seen potential direct analogs to these sampling algorithms in the momentary fluctuations of neural responses[82,83], heuristic solutions can often accomplish the task with much more simplicity. For example, rather than recreating a full posterior distribution from many samples, often only a few[84] will suffice to get one close to the solution (Fig. 2b). When given partial evidence of the length of an event (for example, how long an Egyptian pharaoh has reigned[85]), human subjects (at least in aggregate[86]) can apparently use the correct form of the prior distribution to estimate the full length. While this optimal calculation is arrived by multiplying prior and likelihood together, heuristic solutions can quickly find the best estimate depending on the shape and parameterization of the prior distribution (Fig. 2f). For example, the correct answer for a prior with a power-law distribution is given by simply multiplying the evidence by a fixed multiplier, and for an Erlang distribution, only a constant needs be added to the evidence[85]. Notably, these heuristics simplify the computation by foregoing direct multiplication of prior and evidence distributions together, but still require subjects to have knowledge of the correct prior distribution.

These heuristics that appear to govern human cognitive and value-based decision behavior can be formally defined as efficient solutions to problems that ignore part of the relevant information or the full computation[87]. The mathematician George Pólya championed the idea of teaching students the process of discovering mathematical proofs through the application of various heuristics[88]. This use of heuristics spread early into the artificial intelligence literature as psychologists and computer scientists teamed up to solve problems that were thought to encompass many aspects of higher-level human cognition[89] by using Polya's ideas of heuristic solution to teach computers to generate logical proofs[90,91] or play chess[92]. It was quickly recognized that master chess players need not search the whole tree of possible moves, but search could use strategies like alpha-beta pruning of the decision tree and then assign values to board positions using heuristic evaluation[93].

While heuristic solutions are often associated with bias and cognitive illusions, they need not be considered suboptimal in a larger context. For example, when diagnosing a disease, rather than weight each symptom or predictor optimally, for example as one would do in linear regression, a simpler heuristic solution can be achieved by tallying each cue as either positive or negative evidence (i.e., weighting each predictor equally). This simplifies the calculation as, of course, it would be hard to compute a linear regression in one's head and obviously a simple computer program can achieve better results if linear weighting is the right model. However, building a correct linear regression model, or for that matter, a modern deep-learning model, requires a great deal of data and stationarity of the system. A tallying heuristic in this real-world context can then be construed as approaching an optimal solution, similarly to using robust regression techniques to get a more stable solution[94]. Many fierce arguments have occurred in the literature[48,49,51] over asking whether humans are optimal or suboptimal, but heuristic solution to problems need not be viewed along these extreme dimensions. Instead, optimality can be viewed as deriving the goals of the system or what the nervous system should do and heuristics viewed as the way in which those goals are achieved by biological systems. Using Marr's levels of analysis[95], optimality is the computational goal and heuristics are the algorithms by which those goals are achieved.

Heuristics in perception

A slightly embarrassing fact for committed psychophysicists is that even well-trained observers in carefully designed experimental settings stray from the behavioral rules that psychophysical procedures assume, displaying a variety of different, sometimes idiosyncratic, biases not typically part of an ideal observer model. If given the choice of two intervals, subjects can be biased to one interval or the other, something that can be consistent among observers but idiosyncratic across different types of experiments[96]. Subjects rely on choices, rewards, and stimuli from previous trials even when these are irrelevant, for example, tending to repeat choices or to alternate more often than expected by chance[97–103], something that has been known for a century[104], sometimes rediscovered[105,106], and typically ignored even though it causes systematic underestimates of perceptual sensitivity[97,99]. After errors, subjects tend to slow down[107,108], perhaps to re-engage cognitive control[109]. Previous trial effects may

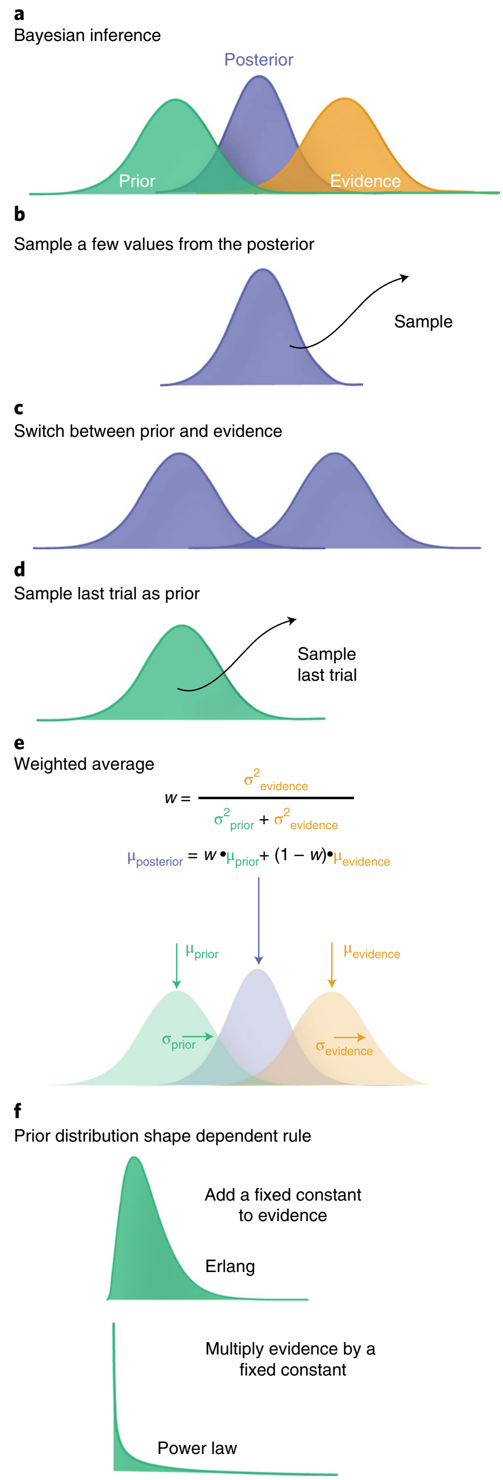


Fig. 2 | Heuristic solutions to sensory inference. a, When inferring properties of the sensory world (for example, the orientation of an edge), Bayesian inference provides the optimal solution of combining prior information (for example, the distribution of orientations in the environment) with noisy sensory evidence (for example, the stochastic spiking of orientation selective neurons) into a posterior distribution and choosing the best estimate using a read-out rule such as the maximum of this distribution, which optimizes the specified loss function of the decision-maker. While this computation in theory can be computed using probabilistic population codes[131], other heuristic solutions that simplify the computational complexity are possible. **b**, Computing only a few samples from the posterior may provide reasonably accurate estimates, particularly if there is a cost to how long it takes to compute samples[84]. **c**, Rather than multiplying prior and evidence together, observers can switch[118–120] between the two in proportion to their reliability. **d**, Instead of keeping a full prior distribution, one can use the last trial[121], or last encounter with the statistic, as a proxy for the full distribution. **e**, For Gaussian distributions, only two statistics need to be tracked from the prior and evidence: the mean and s.d. The posterior mean is then just a weighted sum of the prior and evidence means and the reliability of the posterior can be similarly computed[20]. This achieves the optimal solution while avoiding the full computation of multiplying the distributions together, thus making it a heuristic evaluation. **f**, For some types of prior distributions, a simpler rule[85], like adding or multiplying the evidence by a fixed constant, approximates the full computation.

be used as a prior[110,111], and sensory history representations in posterior parietal cortex bias choices in rats even when not relevant[103]. Though these behaviors may not be conscious, they parallel subjects' biases in producing random number sequences[69] or subjects' false beliefs that alternations or runs are less likely than they actually are, as in the gambler's fallacy. Being rewarded on previous trials typically results in win-stay, lose-switch bias even for seasoned psychophysical observers[99], and overtrained mice still tend to alternate choices[101]. While subjects have the ability to learn the statistical structure of trials, they are more adaptable when statistics confirm their original bias rather than push against it[99].

Close inspection of perceptual behavior that appears to satisfy the goals of optimal considerations can reveal hints of underlying heuristics. Criterion setting, the original obsession of the signal detection theorists, while shifted toward the optimal, never quite achieves an extreme enough setting demanded by the optimal solution[112–114], a behavior suggestive of the same 'conservative Bayesian' strategies that ultimately toppled the rational economic actor model. Heuristic models in which observers forego the full optimal calculation and instead simply set their criterion such that their response probability matches stimulus probability[115] can account for the nonoptimal criterion setting[114,116], and subjects can be biased by recent history[117]. While summary statistics of mean and s.d. of subject estimates in a motion-direction perceptual-estimation task conformed to the expectations of a Bayesian model, the actual distribution of trials suggest a switching heuristic (Fig. 2c) between prior and evidence on different trials in proportion to their relative reliability[118]. Children have been shown to use a similar switching heuristic in sensory cue-combination tasks before they learn to do full integration[119,120]. Other heuristic solutions have also been offered for Bayesian estimation tasks, such as using the past trial (or past few trials)[121] as an estimate of the prior (Fig. 2d). In studies of visual search, optimal strategies can be approximated by heuristic solutions[122] or rules that deviate from the statistically optimal combination of information from multiple sources, such as taking the max[11,12,123,124] or sum[124,125]. Models that weight imperfect cues to account for the perception of surface properties like gloss[126] suggest heuristic computations. Indeed, the tradition of Gestalt psychology was built on trying to understand perceptual organization through defining a set of heuristic rules for grouping stimuli, such as laws of proximity, similarity, closure, symmetry, and common fate.

Outstanding questions, approaches, and challenges

While arguing whether a perceptual behavior demonstrates an optimal or heuristic solution is futile in that it depends on how optimality is defined, determining the algorithm by which a behavior is achieved is absolutely critical for neuroscience. There have been proposals to more explicitly test whether behaviors are optimal, for example, by testing for whether the behavior shows rapid adaptation and transfer of priors to a new set of evidence[127] or by devising behavioral tests in which optimal and heuristic solutions can be thoroughly tested[128]. However, the goal for neuroscientists is not necessarily to determine whether a behavior is optimal, but to understand how it is implemented in the brain[129]. If a behavior appears optimal but is solved by a heuristic algorithm, chasing after ways for the nervous system to perform optimal computations will be fruitless. Take for example the central computation in statistical inference of multiplying together probability distributions. Theorists have proposed elegant ways in which the stochasticity of neural populations implicitly represents probability distributions. Under certain assumptions of the noise characteristics of individual neurons, representations need not be explicitly transformed into likelihoods[130], but instead computations like multiplying different distributions can be solved more simply through addition[131]. While there is an abundance of evidence that humans and animals[132] can integrate conflicting sensory information in nearly optimal fashion, this does not guarantee that the algorithm used requires multiplying probability distributions. Indeed, if observers use heuristic max or sum decision rules as has been suggested in the visual search literature, the neural implementation of these operators, rather than optimal integration, should be the target of inquiry.

A major challenge to discovering the algorithms that underlie perceptual judgements are the limits of standard behavioral techniques. Psychophysical procedures are rightly revered in perceptual science because of their ability to provide quantitative measurements of sensory capacities, but at the same time, they provide a dramatically impoverished view of what they set out to measure, i.e., perception. In an era when we can use functional MRI, electroencephalography, or magnetoencephalography to capture responses across the whole human brain in a fraction of a second or to capture detailed population responses of thousands of neurons at once using calcium imaging, it is remarkable that the most commonly used behavioral techniques typically capture a single binary choice per trial. These choices, combined with reaction times, have been the mainstay of perceptual measurement and have propelled drift-diffusion modeling, which can jointly account for both measurements[28–30]. But much has likely been hidden by psychophysics in which we ask a yes–no or two-alternative forced-choice question; understanding the algorithms and heuristics that humans use to attain, in some cases, nearly optimal estimation behavior will require behavioral frameworks that allow the hallmarks of these algorithms to peek through[133]. For example, estimation tasks in which subjects are required to provide a continuous, one-dimensional readout of the color, orientation, direction, or location of what they see provide full distributions of behavior from which more complex aspects of perceptual decision formation has become evident[80,111,118,134]. More complex stimuli have often been proposed as ways to enrich our understanding of perceptual neuroscience. In particular, white-noise stimuli provide means for estimating perceptual templates[135]. Moving forward, as more complex stimuli and more nuanced measurements of behavior develop, including eye-movement measures, we will need an arsenal of different behavioral models that incorporate potential algorithms and heuristics to be models to compare human behavior to.

One might worry that, without optimality theory to guide the search for neural implementations of perceptual behavior, we are left with simply a bag of tricks[136] from which no principle can be discerned and no single model may apply. Likely there is some truth to this idea, as vision is the result of evolutionary pressures and constraints that have adapted it to solve species- and niche-specific problems and not as a perfect inference machine that re-represents the visual world[137]. Nonetheless, viewed from this perspective, optimality and heuristic considerations of perceptual behavior play complementary and synergistic roles, as optimality theory can provide precise goals for what a perceptual behavior can possibly attain and the various constraints of computational complexity and evolutionary history are captured by the heuristic solutions that behavior may adopt to achieve those goals. Importantly, evolutionary and other pressures may only require satisficing of goals rather than optimizing. A solution that works under most circumstances may be good enough; visual perception can be fooled by illusions (Fig. 1), but the rarity of these special cases in which our visual system is in error is what makes them surprising and novel. While much effort has gone into rigorous and quantitative treatment of optimal goals of the system, less effort has been put into heuristic solutions, but this does not mean that heuristic behaviors cannot be computationally modeled. For example, prospect theory[138] is a quantitative framework that incorporated heuristic insights to correct the optimality driven theory that came before it. It built upon the expected utility framework, but showed how treating probabilities as subjective weights[139] rather than confining them to the mathematical rules of probability could better capture choice behavior (Box 3). A similar

Box 3 | Quantitative models of nonoptimal behavior

Behavior that does not conform to predictions of optimal theory can still be quantitatively accounted for. For example, several aspects of human value-based decision-making that were not predicted by optimal calculations can nonetheless be specified in a rigorous, quantitative framework called prospect theory[138]. Prospect theory does require an initial nonquantitatively described step of editing gambles, after which value is described by the following equation

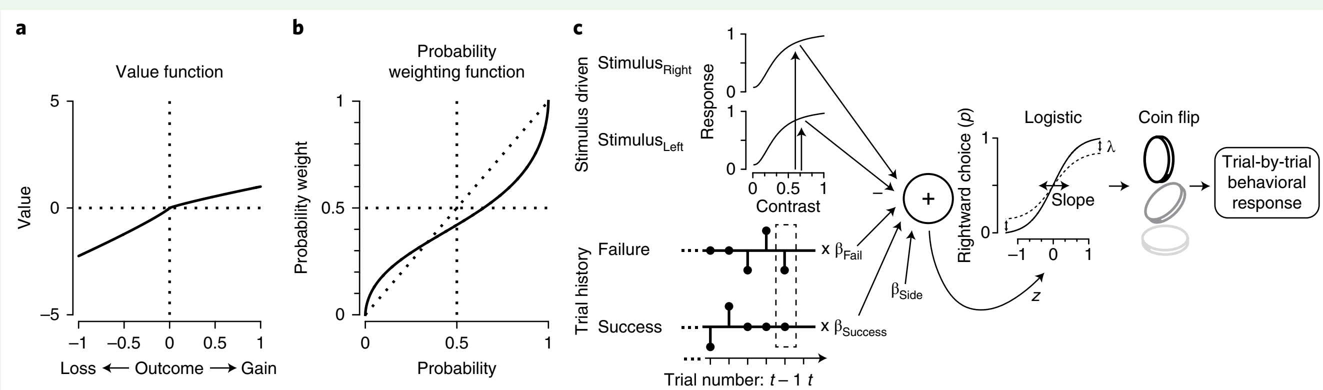
$$v(x) = \begin{cases} x^{a} \text{ if } x \geq 0 \\ -\lambda(-x)^{\beta} \text{ if } x < 0 \end{cases}$$

Where α makes the value function concave for gains, β makes the function convex for losses, and λ controls the scale for losses and accounts for loss aversion by making losses result in larger changes relative to the same-sized gains (**a**; $\alpha = \beta = 0.88$ and $\lambda = 2.25$). Probabilities are converted into subjective weights that no longer satisfy the rules of probability but account for a human subject's tendency to overweight small probabilities and underweight large ones, which can be captured by the following equation

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}}$$

where γ is a parameter that controls the shape of the subjective probability function (**b**; $\gamma = 0.61$).

In perceptual decision-making, typical psychometric analysis of visual choice behavior does not take into account biases from previous trials, but frameworks borrowed from reinforcement learning can estimate choice history biases and therefore better isolate sensitivity to the stimulus. For example, consider a contrast-discrimination task in which a subject must determine which of two stimuli presented to the left or right has higher contrast. Stimulus-driven effects (**c**, top) can be modeled as the difference between the evoked internal response to the two stimuli. Trial history effects (bottom row) such as previous failure or previous success can be multiplied by weights (β) and entered into a logistic function to predict trial-by-trial choices, assuming a small number of lapses (λ)[97,99,101]. By fitting the shape of the internal response functions (top), trial history weights, lapses, and the slope of the logistic function (which specifies the amount of internal noise), sensory-driven effects can be estimated separately from trial-history effects.



Prospect theory and probabilistic choice model. Value (**a**) and probability weighting (**b**) functions from prospect theory. (**c**) Probabilistic model of visual choice behavior incorporating choice history biases.

unifying theory of perceptual behavior is beginning to be built that incorporates mechanisms that are nonoptimal. For example, increasingly, psychophysical modeling is using fitting procedures familiar from reinforcement learning to add stimulus, choice, and reward history components[97,99,101] to better model perceptual decisions (Box 3). To be sure, the goal is not to turn the clock back on perceptual neuroscience to descriptive heuristic models, but instead to incorporate insights from this literature into quantitative models of perceptual behavior.

The search for the behavioral algorithms that humans use to achieve nearly optimal behavior has led some to computer science in which considerations of computing and memory resources have led to efficient algorithms that can approximate difficult-to-compute optimal solutions. For example, in statistical inference, where one might need to calculate the expectation of a complex posterior distribution, an algorithm known as importance sampling, which appropriately weights samples drawn from a simpler distribution, can provide an efficient approximation. This algorithm and its cousin, particle filtering, which applies the same principles to sequential inference problems by tracking a set of samples that are appropriately weighted and resampled through each iterative update, can be shown to produce solutions to cognitive problems that display characteristics similar to those of human behavior[140]. This approach, termed resource rationality, builds on the work of Simon[141] in his consideration of costs for decision-makers in searching for alternative choices to suggest a broader view of optimality which includes the cost of computation, resources, and accessibility of information[84,118,142] and which has provided new normative views of classic heuristic behaviors[142] such as the anchoring and adjustment heuristic[1]. Similarly, optimal solutions for valuing states or actions in reinforcement-learning models must satisfy the Bellman equations, and algorithms such as temporal difference or q-learning can be shown to converge to this solution[143]. But the speed of convergence, the complexity with which a model or model-free system represents the world, and the correct balance of exploration and exploitation make for trade-offs that might govern when a simpler, potentially more heuristic solution outweighs the flexibility and statistical efficiency of a more complex approach to using experience to optimally specify behavioral policies[144]. While these approaches are all promising, a major challenge for perceptual neuroscience is that, if one wants to understand neural computations underlying perceptual decisions, it is not enough to find heuristic solutions that share the same faults as human behavior; one needs to find the actual heuristic solution adopted in any given situation. Approximate algorithms developed for the limitations of modern computer architectures may not be the same ones that evolution has favored, as the costs of resources and computations are not directly comparable. Proponents of the use of these computational metaphors make no claim that sampling, for instance, is implemented by the brain[142], just that heuristic behaviors can be understood as coming from constraints imposed by limitations of resources.

Optimal principles have served perceptual neuroscience well by strictly formulating testable hypotheses, but one should not forget that human behavior is endlessly full of surprises. Sometimes these surprises come in the form of clever, simple, heuristics. The understanding of these heuristics can expose how some simple and elegant shortcut can achieve an end that otherwise seems dauntingly complicated. When this is the case in perceptual neuroscience, we should rejoice, as it sets us on a much simpler course for how to understand the neural bases of perceptual behavior.

Received: 20 June 2018; Accepted: 16 January 2019;
Published online: 25 February 2019

References

1. Tversky, A. & Kahneman, D. Judgment under uncertainty: heuristics and biases. *Science* **185**, 1124–1131 (1974).
2. Tversky, A. & Kahneman, D. The framing of decisions and the psychology of choice. *Science* **211**, 453–458 (1981).
3. Swets, J., Tanner, W. P. Jr. & Birdsall, T. G. Decision processes in perception. *Psychol. Rev.* **68**, 301–340 (1961).
4. Tanner, W. P. Jr. & Swets, J. A. A decision-making theory of visual detection. *Psychol. Rev.* **61**, 401–409 (1954).
5. Green, D. M. Psychoacoustics and detection theory. *J. Acoust. Soc. Am.* **32**, 1189–1203 (1960).
6. Friedman, M. P., Carterette, E. C., Nakatani, L. & Ahumada, A. Comparisons of some learning models for response bias in signal detection. *Percept. Psychophys.* **3**, 5–11 (1968).
7. Nachmias, J. On the psychometric function for contrast detection. *Vision Res.* **21**, 215–223 (1981).
8. Green, D. M. & Moses, F. L. On the equivalence of two recognition measures of short-term memory. *Psychol. Bull.* **66**, 228–234 (1966).
9. Egan, J., Schulman, A. I. & Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. *J. Acoust. Soc. Am.* **31**, 768–773 (1959).
10. Burgess, A. Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation. *J. Opt. Soc. Am. A* **2**, 1498–1507 (1985).
11. Pelli, D. G. Uncertainty explains many aspects of visual contrast detection and discrimination. *J. Opt. Soc. Am. A* **2**, 1508–1532 (1985).
12. Palmer, J., Verghese, P. & Pavel, M. The psychophysics of visual search. *Vision Res.* **40**, 1227–1268 (2000).
13. Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E. & Pouget, A. Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron* **74**, 30–39 (2012).
14. Najemnik, J. & Geisler, W. S. Optimal eye movement strategies in visual search. *Nature* **434**, 387–391 (2005).
15. Eckstein, M. P., Peterson, M. F., Pham, B. T. & Droll, J. A. Statistical decision theory to relate neurons to behavior in the study of covert visual attention. *Vision Res.* **49**, 1097–1128 (2009).
16. Pestilli, F., Carrasco, M., Heeger, D. J. & Gardner, J. L. Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. *Neuron* **72**, 832–846 (2011).
17. Hara, Y. & Gardner, J. L. Encoding of graded changes in spatial specificity of prior cues in human visual cortex. *J. Neurophysiol.* **112**, 2834–2849 (2014).
18. Verghese, P., Kim, Y.-J. & Wade, A. R. Attention selects informative neural populations in human V1. *J. Neurosci.* **32**, 16379–16390 (2012).
19. Lee, D. K., Itti, L., Koch, C. & Braun, J. Attention activates winner-take-all competition among visual filters. *Nat. Neurosci.* **2**, 375–381 (1999).
20. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
21. Seung, H. S. & Sompolinsky, H. Simple models for reading neuronal population codes. *Proc. Natl. Acad. Sci. USA* **90**, 10749–10753 (1993).
22. Scolari, M., Byers, A. & Serences, J. T. Optimal deployment of attentional gain during fine discriminations. *J. Neurosci.* **32**, 7723–7733 (2012).
23. Purushothaman, G. & Bradley, D. C. Neural population code for fine perceptual decisions in area MT. *Nat. Neurosci.* **8**, 99–106 (2005).
24. Graham, N. & Robson, J. G. Summation of very close spatial frequencies: the importance of spatial probability summation. *Vision Res.* **27**, 1997–2007 (1987).
25. Jogan, M. & Stocker, A. A. Signal integration in human visual speed perception. *J. Neurosci.* **35**, 9381–9390 (2015).
26. Newsome, W. T., Britten, K. H. & Movshon, J. A. Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54 (1989).
27. Tolhurst, D. J., Movshon, J. A. & Dean, A. F. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23**, 775–785 (1983).
28. Link, S. W. & Heath, R. A. A sequential theory of psychological discrimination. *Psychometrika* **40**, 77–105 (1975).
29. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **85**, 59–108 (1978).
30. Roitman, J. D. & Shadlen, M. N. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* **22**, 9475–9489 (2002).
31. Barlow, H.B. in *Sensory Communication* (ed. Rosenblith, W.) 217–234 (MIT Press, Cambridge, MA, USA, 1961).
32. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
33. Yang, Z. & Purves, D. A statistical explanation of visual space. *Nat. Neurosci.* **6**, 632–640 (2003).
34. Girshick, A. R., Landy, M. S. & Simoncelli, E. P. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* **14**, 926–932 (2011).
35. Weiss, Y., Simoncelli, E. P. & Adelson, E. H. Motion illusions as optimal percepts. *Nat. Neurosci.* **5**, 598–604 (2002).
36. Vintch, B. & Gardner, J. L. Cortical correlates of human motion perception biases. *J. Neurosci.* **34**, 2592–2604 (2014).
37. Stocker, A. A. & Simoncelli, E. P. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* **9**, 578–585 (2006).
38. Sotiropoulos, G., Seitz, A. R. & Seriès, P. Changing expectations about speed alters perceived motion direction. *Curr. Biol.* **21**, R883–R884 (2011).
39. Ganguli, D. & Simoncelli, E. P. Implicit encoding of prior probabilities in optimal neural populations. *Adv. Neural Inf. Process. Syst.* **23**, 658–666 (2010).
40. Atick, J. J. Could information theory provide an ecological theory of sensory processing? *Network* **3**, 213–251 (1992).
41. Brainard, D. H. & Freeman, W. T. Bayesian color constancy. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **14**, 1393–1411 (1997).
42. Manning, J. R. & Brainard, D. H. Optimal design of photoreceptor mosaics: why we do not see color at night. *Vis. Neurosci.* **26**, 5–19 (2009).
43. Geisler, W. S., Perry, J. S., Super, B. J. & Gallogly, D. P. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* **41**, 711–724 (2001).
44. Lee, T. S. & Mumford, D. Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **20**, 1434–1448 (2003).
45. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
46. Gregory, R.L. *Concepts and Mechanisms of Perception* (Duckworth, London, 1974).
47. Rock, I. *The Logic of Perception* (MIT Press, Cambridge, MA, USA, 1983).
48. Bowers, J. S. & Davis, C. J. Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* **138**, 389–414 (2012).
49. Jones, M. & Love, B. C. Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behav. Brain Sci.* **34**, 188–231 (2011). 169–188.
50. Marcus, G. F. & Davis, E. How robust are probabilistic models of higher-level cognition? *Psychol. Sci.* **24**, 2351–2360 (2013).
51. Rahnev, D. & Denison, R. N. Suboptimality in perceptual decision making. *Behav. Brain Sci.* https://doi.org/10.1017/S0140525X18000936 (2018).
52. Geisler, W. S. Contributions of ideal observer theory to vision research. *Vision Res.* **51**, 771–781 (2011).
53. Griffiths, T.L., Chater, N., Norris, D. & Pouget, A. How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). *Psychol. Bull.* **138**, 415–422 (2012).
54. Goodman, N. D. et al. Relevant and robust: a response to Marcus and Davis (2013). *Psychol. Sci.* **26**, 539–541 (2015).
55. Frank, M. C. Throwing out the Bayesian baby with the optimal bathwater: response to Endress (2013). *Cognition* **128**, 417–423 (2013).
56. Neumann, Von, J. & Morgenstern, O. *Theory of Games and Economic Behaviour* (Princeton University Press, Princeton, NJ, USA, 1944).
57. Bernoulli, D. Exposition of a new theory on the measurement of risk. *Econometrica* **22**, 23–36 (1954).
58. Allais, M. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes del'ecole Americaine. *Econometrica* **21**, 503–546 (1953).

59. Kahneman, D. & Tversky, A. Choices, values, and frames. *Am. Psychol.* **39**, 341–350 (1984).
60. Hershey, J. C. & Schoemaker, P. J. H. Risk taking and problem context in the domain of losses: an expected utility analysis. *J. Risk Insur.* **47**, 111–132 (1980).
61. Samuelson, P. A. A note on the pure theory of consumer's behaviour. *Economica* **5**, 61–71 (1938).
62. Tversky, A. Intransitivity of preferences. *Psychol. Rev.* **76**, 31–48 (1969).
63. Simon, H. A. Rational decision making in business organizations. *Am. Econ. Rev.* **69**, 493–513 (1979).
64. Kahneman, D. & Tversky, A. On the psychology of prediction. *Psychol. Rev.* **80**, 237–251 (1973).
65. Edwards, W. in *Judgment under Uncertainty Heuristics and Biases* (eds. Kahneman, D., Slovic, P. & Tversky, A.) 359–369 (Cambridge University Press, Cambridge, UK, 1968).
66. Fischhoff, B. & Beyth-Marom, R. Hypothesis evaluation from a Bayesian perspective. *Psychol. Rev.* **90**, 239–260 (1983).
67. Kahneman, D. & TverskyA. Subjective probability: a judgment of representativeness. *Cogn. Psychol* **3**, 430–454 (1972).
68. Tversky, A. & Kahneman, D. Belief in the law of small numbers. *Psychol. Bull.* **76**, 105–110 (1971).
69. Bar-Hillel, M. & Wagenaar, W. A. The perception of randomness. *Adv. Appl. Math.* **12**, 428–454 (1991).
70. Gigerenzer, G. How to make cognitive illusions disappear: beyond 'heuristics and biases'. *Eur. Rev. Soc. Psychol.* **2**, 83–115 (2011).
71. Hertwig, R. & Gigerenzer, G. The 'conjunction fallacy' revisited: how intelligent inferences look like reasoning errors. *J. Behav. Decis. Making* **12**, 275–305 (1999).
72. Birnbaum, M. H. Base rates in Bayesian inference: signal detection analysis of the cab problem. *Am. J. Psychol.* **96**, 85–94 (1983).
73. Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. *Science* **350**, 1332–1338 (2015).
74. Frank, M. C. & Goodman, N. D. Predicting pragmatic reasoning in language games. *Science* **336**, 998 (2012).
75. Chater, N. & Manning, C. D. Probabilistic models of language processing and acquisition. *Trends Cogn. Sci.* **10**, 335–344 (2006).
76. Kemp, C. & Tenenbaum, J. B. The discovery of structural form. *Proc. Natl. Acad. Sci. USA* **105**, 10687–10692 (2008).
77. Gerstenberg, T. et al. Lucky or clever? From expectations to responsibility judgments. *Cognition* **177**, 122–141 (2018).
78. Feldman, N. H., Griffiths, T. L. & Morgan, J. L. The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* **116**, 752–782 (2009).
79. Kemp, C. & Tenenbaum, J. B. Structured statistical models of inductive reasoning. *Psychol. Rev.* **116**, 20–58 (2009).
80. Jazayeri, M. & Shadlen, M. N. Temporal context calibrates interval timing. *Nat. Neurosci.* **13**, 1020–1026 (2010).
81. Körding, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247 (2004).
82. Gershman, S. J., Vul, E. & Tenenbaum, J. B. Multistability and perceptual inference. *Neural Comput.* **24**, 1–24 (2012).
83. Moreno-Bote, R., Knill, D. C. & Pouget, A. Bayesian sampling in visual perception. *Proc. Natl. Acad. Sci. USA* **108**, 12491–12496 (2011).
84. Vul, E., Goodman, N., Griffiths, T. L. & Tenenbaum, J. B. One and done? Optimal decisions from very few samples. *Cogn. Sci.* **38**, 599–637 (2014).
85. Griffiths, T. L. & Tenenbaum, J. B. Optimal predictions in everyday cognition. *Psychol. Sci.* **17**, 767–773 (2006).
86. Mozer, M. C., Pashler, H. & Homaei, H. Optimal predictions in everyday cognition: the wisdom of individuals or crowds? *Cogn. Sci.* **32**, 1133–1147 (2008).
87. Gigerenzer, G. & Gaissmaier, W. Heuristic decision making. *Annu. Rev. Psychol.* **62**, 451–482 (2011).
88. Polya, G. *How to Solve It* (Princeton University Press, Princeton, NJ, USA, 2014).
89. Newell, A. You can't play 20 questions with nature and win: projective comments on the papers of this symposium. *Visual Information Processing*. https://doi.org/10.1016/B978-0-12-170150-5.50012-3 (1973).
90. Newell, A., Shaw, J.C. & Simon, H.A. Empirical explorations of the logic theory machine: a case study in heuristic. *Proceedings of the IRE-AIEE-ACM Western Joint Computer Conference: Techniques for Reliability* 218–230 (1957).
91. Lenat, D. B. *AM, an artificial intelligence approach to discovery in mathematics as heuristic search*. PhD thesis, Stanford Univ. (1976).
92. Newell, A. The chess machine: an example of dealing with a complex task by adaptation. *Proceedings of the AFIPS Western Joint Computer Conference* 101–108 (1955).
93. Newell, A. & Simon, H. A. Computer science as empirical inquiry: symbols and search. *Commun. ACM* **19**, 113–126 (1976).
94. Dawes, R. M. The robust beauty of improper linear models in decision making. *Am. Psychol.* **34**, 571–582 (1979).
95. Marr, D. *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information*. (MIT Press, Cambridge, MA, USA, 1982).
96. Yeshurun, Y., Carrasco, M. & Maloney, L. T. Bias and sensitivity in two-interval forced choice procedures: tests of the difference model. *Vision Res.* **48**, 1837–1851 (2008).
97. Fründ, I., Wichmann, F. A. & Macke, J. H. Quantifying the effect of intertrial dependence on perceptual decisions. *J. Vis.* **14**, 9 (2014).
98. Braun, A., Urai, A. E. & Donner, T. H. Adaptive history biases result from confidence-weighted accumulation of past choices. *J. Neurosci.* **38**, 2418–2429 (2018).
99. Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M. & Gardner, J. L. Adaptable history biases in human perceptual decisions. *Proc. Natl. Acad. Sci. USA* **113**, E3548–E3557 (2016).
100. Lueckmann, J.-M., Macke, J. H. & Nienborg, H. Can serial dependencies in choices and neural activity explain choice probabilities? *J. Neurosci.* **38**, 3495–3506 (2018).
101. Busse, L. et al. The detection of visual contrast in the behaving mouse. *J. Neurosci.* **31**, 11351–11361 (2011).
102. Marcos, E. et al. Neural variability in premotor cortex is modulated by trial history and predicts behavioral performance. *Neuron* **78**, 249–255 (2013).
103. Akrami, A., Kopec, C. D., Diamond, M. E. & Brody, C. D. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature* **554**, 368–372 (2018).
104. Fernberger, S. W. Interdependence of judgments within the series for the method of constant stimuli. *J. Exp. Psychol.* **3**, 126–150 (1920).
105. Senders, V. L. & Sowards, A. Analysis of response sequences in the setting of a psychophysical experiment. *Am. J. Psychol.* **65**, 358–374 (1952).
106. Howarth, C. I. & Bulmer, M. G. Non-random sequences in visual threshold experiments. *Q. J. Exp. Psychol.* **8**, 163–171 (1956).
107. Laming, D. Choice reaction performance following an error. *Acta Psychol. (Amst.)* **43**, 199–224 (1979).
108. Rabbitt, P. & Rodgers, B. What does a man do after he makes an error? an analysis of response programming. *Q. J. Exp. Psychol.* **29**, 727–743 (1977).
109. Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624–652 (2001).
110. Fischer, J. & Whitney, D. Serial dependence in visual perception. *Nat. Neurosci.* **17**, 738–743 (2014).
111. Fritsche, M., Mostert, P. & de Lange, F. P. Opposite effects of recent history on perception and decision. *Curr. Biol.* **27**, 590–595 (2017).
112. Ackermann, J. F. & Landy, M. S. Suboptimal decision criteria are predicted by subjectively weighted probabilities and rewards. *Atten. Percept. Psychophys.* **77**, 638–658 (2015).
113. Ulehla, Z. J. Optimality of perceptual decision criteria. *J. Exp. Psychol.* **71**, 564–569 (1966).
114. Creelman, C. D. & Donaldson, W. ROC curves for discrimination of linear extent. *J. Exp. Psychol.* **77**, 514–516 (1968).
115. Parks, T. E. Signal-detectability theory of recognition-memory performance. *Psychol. Rev.* **73**, 44–58 (1966).
116. Thomas, E. A. & Legge, D. Probability matching as a basis for detection and recognition decisions. *Psychol. Rev.* **77**, 65–72 (1970).
117. Norton, E. H., Fleming, S. M., Daw, N. D. & Landy, M. S. Suboptimal criterion learning in static and dynamic environments. *PLoS Comput. Biol.* **13**, e1005304 (2017).
118. Laquitaine, S. & Gardner, J. L. A Switching observer for human perceptual estimation. *Neuron* **97**, 462–474.e6 (2018).
119. Nardini, M., Jones, P., Bedford, R. & Braddick, O. Development of cue integration in human navigation. *Curr. Biol.* **18**, 689–693 (2008).
120. Adams, W. J. The development of audio-visual integration for temporal judgements. *PLoS Comput. Biol.* **12**, e1004865 (2016).
121. Raviv, O., Ahissar, M. & Loewenstein, Y. How recent history affects perception: the normative approach and its heuristic approximation. *PLoS Comput. Biol.* **8**, e1002731 (2012).
122. Najemnik, J. & Geisler, W. S. Simple summation rule for optimal fixation selection in visual search. *Vision Res.* **49**, 1286–1294 (2009).
123. Baldassi, S. & Verghese, P. Comparing integration rules in visual search. *J. Vis.* **2**, 559–570 (2002).
124. Kramer, P., Graham, N. & Yager, D. Simultaneous measurement of spatial-frequency summation and uncertainty effects. *J. Opt. Soc. Am. A* **2**, 1533–1542 (1985).
125. Baldassi, S. & Burr, D. C. Feature-based integration of orientation signals in visual search. *Vision Res.* **40**, 1293–1300 (2000).
126. Marlow, P. J., Kim, J. & Anderson, B. L. The perception and misperception of specular surface reflectance. *Curr. Biol.* **22**, 1909–1913 (2012).
127. Maloney, L. T. & Mamassian, P. Bayesian decision theory as a model of human visual perception: testing Bayesian transfer. *Vis. Neurosci.* **26**, 147–155 (2009).

128. Shen, S. & Ma, W. J. A detailed comparison of optimality and simplicity in perceptual decision making. *Psychol. Rev.* **123**, 452–480 (2016).
129. Darlington, T. R., Beck, J. M. & Lisberger, S. G. Neural implementation of Bayesian inference in a sensorimotor behavior. *Nat. Neurosci.* **21**, 1442–1451 (2018).
130. Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nat. Neurosci.* **9**, 690–696 (2006).
131. Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**, 1432–1438 (2006).
132. Fetsch, C. R., DeAngelis, G. C. & Angelaki, D. E. Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nat. Rev. Neurosci.* **14**, 429–442 (2013).
133. Huk, A., Bonnen, K. & He, B. J. Beyond trial-based paradigms: continuous behavior, ongoing neural activity, and natural stimuli. *J. Neurosci.* **38**, 7551–7558 (2018).
134. Zhang, W. & Luck, S. J. Discrete fixed-resolution representations in visual working memory. *Nature* **453**, 233–235 (2008).
135. Beard, B.L. & Ahumada, A.J. in *Human Vision and Electronic Imaging III* **3299**, 79–86 (1998).
136. Ramachandran, V.S. in *Vision: Coding and Efficiency* (ed. Blakemore, C.) 346–360 (Cambridge Univ. Press, Cambridge, UK, 1991).
137. Purves, D., Wojtach, W. T. & Lotto, R. B. Understanding vision in wholly empirical terms. *Proc. Natl. Acad. Sci. USA* **108**(Suppl 3), 15588–15595 (2011).
138. Kahneman, D. & Tversky, A. Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–291 (1979).
139. Savage, L. J. *The Foundations of Statistics* (Wiley, New York, 1954).
140. Shi, L., Griffiths, T. L., Feldman, N. H. & Sanborn, A. N. Exemplar models as a mechanism for performing Bayesian inference. *Psychon. Bull. Rev.* **17**, 443–464 (2010).
141. Simon, H. A. A behavioral model of rational choice. *Q. J. Econ.* **69**, 99–118 (1955).
142. Lieder, F., Griffiths, T. L., M Huys, Q. J. & Goodman, N. D. The anchoring bias reflects rational use of cognitive resources. *Psychon. Bull. Rev.* **25**, 322–349 (2018).
143. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: an Introduction* (MIT Press, Cambridge, MA, USA, 1998).
144. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
145. Helmholtz, H. V. *Treatise on Physiological Optics. Volume 3: The Perceptions of Vision* (Optical Society of America, New York, 1867).
146. Ramachandran, V. S. Perception of shape from shading. *Nature* **331**, 163–166 (1988).
147. Freeman, W. T. The generic viewpoint assumption in a framework for visual perception. *Nature* **368**, 542–545 (1994).
148. Nakayama, K. & Shimojo, S. Experiencing and perceiving visual surfaces. *Science* **257**, 1357–1363 (1992).
149. Ma, W. J. Organizing probabilistic models of perception. *Trends Cogn. Sci.* **16**, 511–518 (2012).
150. Tversky, A. & Kahneman, D. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* **90**, 293–315 (1983).

Acknowledgements
I thank E. Thomas, H. Gweon, S. Wu, D. Birman, G. Riesen, A. Jagadeesh, and M. Lee for discussions and comments on earlier drafts of this manuscript. I am grateful for the generous support of Research to Prevent Blindness and Lions Clubs International Foundation and the Hellman Fellows Fund.

Competing interests
The authors declare no competing interests.

Additional information
Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to J.L.G.