Journal of Mathematical Psychology 73 (2016) 0–22

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

1 change on this page

A tutorial on cue combination and Signal Detection Theory: Using changes in sensitivity to evaluate how observers integrate sensory information

Pete R. Jones

Institute of Ophthalmology, University College London (UCL), 11-43 Bath Street, London EC1V 9EL, UK

ARTICLE INFO

Article history:
Received 28 May 2015
Received in revised form
5 April 2016

Keywords:
Cue combination
Multisensory integration
Weighted linear summation
Signal Detection Theory
Internal noise

ABSTRACT

Many sensory inputs contain multiple sources of information ('cues'), such as two sounds of different frequencies, or a voice heard in unison with moving lips. Often, each cue provides a separate estimate of the same physical attribute, such as the size or location of an object. An ideal observer can exploit such redundant sensory information to improve the accuracy of their perceptual judgments. For example, if each cue is modeled as an independent, Gaussian, random variable, then combining N cues should provide up to a $\sqrt{N}$ improvement in detection/discrimination sensitivity. Alternatively, a less efficient observer may base their decision on only a subset of the available information, and so gain little or no benefit from having access to multiple sources of information. Here we use Signal Detection Theory to formulate and compare various models of cue-combination, many of which are commonly used to explain empirical data. We alert the reader to the key assumptions inherent in each model, and provide formulas for deriving quantitative predictions. Code is also provided for simulating each model, allowing expected levels of measurement error to be quantified. Based on these results, it is shown that predicted sensitivity often differs surprisingly little between qualitatively distinct models of combination. This means that sensitivity alone is not sufficient for understanding decision efficiency, and the implications of this are discussed.
© 2016 Elsevier Inc. All rights reserved.

Consider a simple sensory judgment, such as 'where was the source of a sound located'? When attempting to understand how such a decision is made, the sensory input can be thought of as containing multiple sources of information ('cues'). In general, each cue is a function of the sensory input, which conveys information about a particular physical attribute (Sahani & Whiteley, 2011). However, exactly how cues are conceptualized varies between scientific disciplines. In biochemistry, the output of each ionotropic receptor may be considered a distinct cue (DeVries, 2000). In electrophysiology, a cue is generally the firing-rate of a neuron (Stein & Stanford, 2008), or of a given population of neurons (Averbeck, Latham, & Pouget, 2006; Deneve, Latham, & Pouget, 2001; Ma, Beck, Latham, & Pouget, 2006; Pouget, Deneve, & Duhamel, 2002). In the behavioral sciences, which the present paper concerns, cues are typically defined with respect to the stimulus. Thus, interaural differences in intensity and phase may be thought of as separate cues in a sound-localization task (Macpherson & Middlebrooks, 2002). Similarly, texture and disparity may be thought of as separate cues when judging visual depth (Hillis, Watt, Landy, & Banks, 2004). Alternatively, cues may be defined with respect to time; for example, each interval in a two-alternative forced choice [2AFC] (Jones, Shub, Moore, & Amitay, 2013), or each sample in a sequential-observation (Alexander & Lutfi, 2008; Juni, Gureckis, & Maloney, 2012; Swets, 1959) task. Finally, in some cases, cues may be defined with respect to the observer themselves. Thus, each eye (Blake & Fox, 1973; Blake, Sloane, & Fox, 1981), ear (Langhans & Kohlrausch, 1992), area of skin (Geldard & Sherrick, 1965), or sensory modality (Fetsch, Turner, DeAngelis, & Angelaki, 2009) may be thought of as yielding a separate cue.

Irrespective of how exactly the various cues are defined, a number of interesting questions arise: Can observers exploit these multiple sources of information (Stein, Meredith, Huneycutt, & McDade, 1989; Zacharias & Young, 1981)? Do they do so in an optimal manner (Ernst & Banks, 2002; Landy, Maloney, Johnston, & Young, 1995)? Do they continue to do so when the statistics of the task vary (Alais & Burr, 2004; Fetsch, Pouget, DeAngelis, & Angelaki, 2012; Knill & Saunders, 2003; Nardini, Bedford, & Mareschal, 2010)? At what age does this ability to combine cues develop (Gori, Del Viva, Sandini, & Burr, 2008; Nardini, Jones, Bedford, & Braddick, 2008)? Is it preserved in old age (Bates &

E-mail address: p.r.jones@ucl.ac.uk.

http://dx.doi.org/10.1016/j.jmp.2016.04.006
0022-2496/© 2016 Elsevier Inc. All rights reserved.

0 changes on this page

Wolbers, 2014; Laurienti, Burdette, Maldjian, & Wallace, 2006)? Is it present in clinical populations where some information channels are degraded (Alexander & Lutfi, 2004; Moro, Harris, & Steeves, 2014), or have been previously deprived of input (Garcia, 0000; Landry, Guillemot, & Champoux, 2013; Putzar, Gondan, & Röder, 2012)?

In psychophysics, such questions are often addressed by comparing an empirical measure to the predictions of one or more theoretical model of decision making. Since psychophysical tasks often require observers to minimize error, the key empirical measure tends to be some index of sensitivity (e.g., d', or the slope of the psychometric function). Accordingly, one might measure d' when two cues (e.g., texture and disparity) are presented individually, and again when both cues are presented together. If d' in the multi-cue case exceeds that of the best single-cue, then this is strong evidence that observers are using information from both cues to make their decision; we can therefore rule out any model of decision making that relies solely on a single source of information.

If the underlying model of decision making is known, it can also be used as a yardstick to assess how effective observers are at exploiting the information available to them. Thus, by defining some putative 'ideal' level of performance, it becomes possible to compare observed performance to the ideal, and thereby to state whether the observer is behaving *optimally*. Furthermore, by measuring observed performance relative to the ideal, a measure of *efficiency* can be computed (defined formally in Eq. (1.1.5)). This allows cue-combination ability to be compared across observers, even when each individual's sensitivity is expected to vary (Berg, 1990). *Ideal observer* analyses are therefore of substantial practical and theoretical utility, and are used extensively throughout studies of sensory cue-combination (Fetsch, DeAngelis, & Angelaki, 2013; Trommershauser, Kording, & Landy, 2011) (for further discussion, see Ref. Landy, Banks, & Knill, 2011).

However, what has not always been made clear is the diversity of plausible ideal-observer models. Thus, depending on the specific model used, what constitutes 'ideal' performance may differ between papers, and human performance in one study can exceed the predictions of an ideal observer in another (e.g., contrast the factor of N improvement predicted by Saarela & Landy, 2012 with the factor of $\sqrt{N}$ improvement predicted by Knill & Saunders, 2003). A closely related issue is that readers are not always fully aware of the key assumptions that are often required in order to compute 'ideal' performance. As shall be discussed, these assumptions are rarely strictly correct, and depending on exactly what assumptions one makes, the inferences regarding underlying decision-process may differ markedly.

The present paper

The goal of the present paper is to detail exactly what conclusions regarding cue-combination can, and cannot, be inferred from behavioral estimates of sensitivity.

Note that because we are only considering sensitivity as our dependent variable, we will limit ourselves to tasks where the observer's goal is to minimize response error. Such tasks are in no way an exhaustive reflection of everyday sensory decision making (see Section 4), though they do constitute the substantial majority of tasks in the cue-combination literature.

Also note that, when quantifying sensitivity, we shall focus specifically upon d' and other related Signal Detection Theory (Green & Swets, 1974; Macmillan & Creelman, 2005; Wickens, 2002) [SDT] metrics. Other measures can also be used to study perceptual sensitivity, such as the slope parameter of the psychometric function (Ernst & Banks, 2002) or the variance of a continuously distributed response (Nardini et al., 2008). However, SDT metrics are of particular interest due to their prevalence in the literature (Ban, Preston, Meeson, & Welchman, 2012; Dekker et al., 2015; Gu, Angelaki, & DeAngelis, 2008; Machilsen & Wagemans, 2011; Nardini et al., 2010; Persike & Meinhardt, 2015; Saarela & Landy, 2012),[1] and because SDT provides a formal mathematical framework for exploring the key assumptions/ideas common across most studies of cue combination.

The paper is divided into four main sections. In Section 1, we introduce briefly the relevant background theory. In Section 2, we consider the different ways in which information from multiple cues can be used to make a decision, and derive quantitative predictions for each possible decision strategy. In doing so, we detail the assumptions implicit in the various models, and alert the reader to the difficulties that arise if these assumptions are not met. Working examples of each model are also provided in the Supplemental Materials (coded in MATLAB; The MathWorks, Natick, MA). In Section 3, we summarize the information presented and develop overall comparisons and corollaries. In Section 4 we highlight the limits of what can be inferred from sensitivity alone, and discuss other approaches to studying cue-combination.

1. Background theory

1.1. Using signal detection theory to measure perceptual sensitivity

Explicitly or implicitly, studies of cue-combination typically use the theoretical framework of Signal Detection Theory [SDT] to understand how observers make their perceptual judgments (Ernst, 2006). Here we detail its key tenets. For more comprehensive expositions, see Refs. Green and Swets (1974), Macmillan and Creelman (2005) and Wickens (2002).

In SDT, an incoming sensory signal is theorized to produce an *internal response*, typically represented as a single scalar variable, x (Fig. 1(A)). Exactly how this number is instantiated in the brain is irrelevant for present purposes; however, for the sake of example, it could be thought of as the firing rate of a neuron, or the maximum response of a neural population code. Now, consider a simple yes/no detection task. On signal-absent trials, the expected response will equal some baseline quantity that we shall call "0", while on signal-present trials the expected response will be proportional to the task-relevant stimulus feature, S (e.g., the intensity of a sound, in dB SPL, or the luminance of a light, in cd/m^2). Notably though, various neural (Javel & Viemeister, 2000), physiological (Soderquist & Lindsey, 1971), and cognitive processes mean that the internal response is *noisy*. Thus, on each observation (i.e., on each trial in a yes/no task, or each interval in a two-alternative forced-choice task) x may deviate slightly from the expected mean value of 0 or S (Fig. 1(B)). To classify any given value of x as either 'signal' or 'noise', the observed value of x must therefore be compared to some cut-off criterion, λ, thus:

$$\textit{Response} = \begin{cases} \text{'Signal Present'} & \text{if DV} > \lambda \\ \text{'Signal Absent'} & \text{otherwise} \end{cases} \quad \text{where DV} = x. \qquad (1.1.1a)$$

In Eq. (1.1.1a) the decision variable, DV, upon which the behavioral response is based ('Response'), is simply the internal response to a single cue, x. In more complex tasks, however, the DV will not be determined by a single internal response. For example, in a two alternative forced-choice task, the DV is

[1] The use of SDT metrics is particularly prevalent among paradigms where the intensity of the target stimulus is fixed or determined by an adaptive (threshold) algorithm, and/or in cases where responses are binary. For continuously distributed responses, experimenters may wish to dispense with SDT sensitivity metrics, and instead use the variability of the response distribution as a more 'direct' proxy for the precision of the observer's sensory estimate. However, not all tasks lend themselves to this type of experimental design, and more complex methods of response can also introduce unwanted (e.g., non-sensory) sources of noise or bias.

2 changes on this page

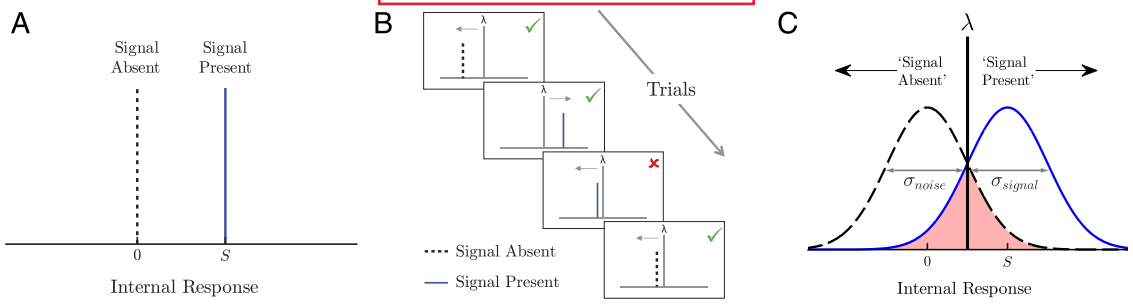


Fig. 1. Signal Detection Theory schema for a simple, yes/no detection task. On each trial, either a signal is presented, or nothing is presented. (A) The mean internal response is either 0 (Nothing), or S (Signal; where S is proportional to signal magnitude). (B) However, internal noise means that, on any given trial, the internal response may be slightly higher or slightly lower than ~~the mean~~. Any given internal responses must therefore be compared to a criterion, λ, in order to determine the appropriate behavioral action. (C) When averaging across many trials, a distribution of internal responses is observed for each condition (in this image the noise is additive and normally distributed). To the extent that the two distributions overlap (shaded region), the observer cannot distinguish the two conditions perfectly. This overlap is what is indexed by the sensitivity metric, d'.

generally considered to be the *difference* between the internal responses to the stimulus in each interval:

$$\textit{Response} = \begin{cases} \text{'Signal in Interval 2'} & \text{if DV} > \lambda \\ \text{'Signal in Interval 1'} & \text{otherwise} \end{cases} \quad \text{where}$$

$$\text{DV} = x_{\text{Int2}} - x_{\text{Int1}}. \tag{1.1.1b}$$

While in a compound-detection task (which the present paper focuses on primarily), the DV is the *sum* of N internal response values:

$$\textit{Response} = \begin{cases} \text{'Signal Present'} & \text{if DV} > \lambda \\ \text{'Signal Absent'} & \text{otherwise} \end{cases} \quad \text{where}$$

$$\text{DV} = x_1 + x_2. \tag{1.1.1c}$$

Irrespective of how exactly the DV is derived, to make a binary decision it must be compared to a criterion, λ. Ideally, λ will be placed so as to maximize some expected utility function (e.g., percent correct, N points won, etc.). If λ deviates from the ideal location then the observer is said to be *biased*. However, even with an unbiased criterion, when the sensory noise is continuously distributed some errors are inevitable. That is, no criterion will perfectly separate the signal-absent and signal-present distributions. An input of "0" may therefore sometimes trigger a 'Signal Present' response, or an actual signal, S, may trigger a 'Signal Absent' response (Fig. 1(C)). To the extent that the 'Noise' and 'Signal' internal-response distributions overlap (the shaded area in Fig. 1(C)), the observer is said to be less *sensitive* to differences between the two.

If we assume for now that the sensory noise is Gaussian distributed, then sensitivity (the degree of separation between the two internal response distributions) is determined formally by the difference in the means of the two internal response distributions, divided by their root mean variance:

$$\textit{sensitivity} = \frac{\mu_{\text{signal}} - \mu_{\text{noise}}}{\sqrt{\frac{1}{2}\left(\sigma^2_{\text{signal}} + \sigma^2_{\text{noise}}\right)}}. \tag{1.1.2}$$

Eq. (1.1.2) contains four parameters: the means of the two distributions (μ_{signal}, μ_{noise}) and their variances (σ^2_{signal}, σ^2_{noise}). Such a model is non-identifiable by most behavioral datasets, in that the four parameters cannot be uniquely constrained by observations of Hit (yes | signal), h, and False Alarm (yes | noise), f, rates. Fortunately, however, we are not typically interested in the individual parameters, and only want to index overall sensitivity. In that case, we can, without any loss of generality, recenter the means of the two distributions so that $\mu_{\text{noise}} = 0$ (while μ_{signal} remains an unknown parameter, to be estimated empirically). Furthermore, we shall for now assume that the internal noise is additive (i.e., is independent of internal response magnitude), in which case the variance of each distribution will be identical ($\sigma^2_{\text{signal}} = \sigma^2_{\text{noise}} = \sigma^2$). Finally, we can, again without any loss of generality, scale the means of the distributions by their standard deviations. In this way, the standard deviation becomes unity ($\sigma = \sigma/\sigma = 1$), and the mean becomes the mean scaled by the standard deviation ($\mu_{\text{signal}} = \mu_{\text{signal}}/\sigma$). By convention, this scaled mean is referred to as d', and can now be seen to be the sole determinant of sensitivity:

$$\begin{aligned} \textit{sensitivity} &= \frac{\mu_{\text{signal}} - 0}{\sqrt{\frac{1}{2}\left(\sigma^2_{\text{signal}} + \sigma^2_{\text{noise}}\right)}} = \frac{\mu_{\text{signal}}}{\sqrt{\frac{1}{2}\left(\sigma^2 + \sigma^2\right)}} \\ &= \frac{\mu_{\text{signal}}/\sigma}{\sqrt{\frac{1}{2}\left((\sigma/\sigma)^2 + (\sigma/\sigma)^2\right)}} = \frac{d'}{\sqrt{1}} = d'. \end{aligned} \tag{1.1.3}$$

In practice, the value of d' can be estimated empirically from the observed Hit rate, h, and False Alarm rate, f, thus:

$$d' = \Phi^{-1}(h) - \Phi^{-1}(f), \tag{1.1.4a}$$

where Φ^{-1} represents the inverse of the cumulative standard normal function (also commonly referred to as the z function ~~in statistics~~), which can be expressed in terms of the inverse error function:

$$\Phi^{-1}(p) = \sqrt{2}\,\text{erf}^{-1}(2p - 1). \tag{1.1.4b}$$

For the derivation of Eq. (1.1.4), see Refs. Berg (1990), Fetsch et al. (2013) and Trommershauser et al. (2011).

Finally, it is often useful to measure observed sensitivity, d'_{obs} relative to some putative ideal, d'_{ideal}. This allows easy comparison across conditions where d'_{ideal} is expected to vary—for example, when comparing compound-cue sensitivity across observers who are differentially sensitive to the constituent single-cues. Such a measure is known as *efficiency*, η, and is defined by Tanner and Birdsall (1958) as:

$$\eta = \left[d'_{\text{obs}}/d'_{\text{ideal}}\right]^2. \tag{1.1.5}$$

Thus, $\eta = 1$ indicates ideal sensitivity, while $\eta < 1$ indicates suboptimal performance.

1.2. Combining random variables

Section 1.1 expounded how, according to SDT, binary decisions are made by comparing a scalar Decision Variable, DV, to a

criterion, λ. If there is just a single cue, then t… determined by a single internal response variable, x (DV $= x$). However, as discussed, many models of decision-making imply multiple separate cues, each with its own associated source of noise. In that case, each cue can be considered a separate random variable, and in most models of decision making the DV is some combination of these variables (though cf. Sections 2.1, 2.2). It is therefore important to understand the expected properties of a random variable that is the combination of N random variables.

Firstly, let us assume for now that the process of combination is linear. The linear combination of two random variables is the additive sum of the individual values, each multiplied by some relative weighting constant. Thus, if x_1 and x_2 are two internal responses (two distinct cues), and ω_1 and ω_2 are their associated weight coefficients, then:

$$\mathrm{DV} = \omega_1 x_1 + \omega_2 x_2. \tag{1.2.1}$$

When $\omega_1 = 1$ and $\omega_2 = 1$ the two cues are summed completely (Total Summation). This may not always be desirable, however. For example, when detecting a change of heading, our sense of vision often provides more accurate information then our internal sense of balance (Nardini et al., 2008). If the goal is to minimize response error, then the less reliable balance cue should be given less relative weight ($\omega < 1$; Partial Summation).

How to weight cues optimally shall be discussed in Section 2. However, irrespective of the specific weights employed, the mean of the weighted-sum of two random variables is the weighted-sum of the individual means:

$$\mu_{\mathrm{DV}} = \omega_1 \mu_1 + \omega_2 \mu_2. \tag{1.2.2}$$

Furthermore, if we assume for the moment that the noise associated with each cue is independent, then the variance of the weighted-sum is the sum of the individual variances, weighted by the squares of the coefficients:

$$\sigma_{\mathrm{DV}} = \sqrt{\omega_1^2 \sigma_1^2 + \omega_2^2 \sigma_2^2}. \tag{1.2.3}$$

By applying Eq. (1.2.3) to the SDT sensitivity formula given earlier (Eq. (1.1.2)), a general formula can be derived for expected sensitivity when two independent cues are linearly summed, d_{1+2}, thus:

$$d'_{1+2} = \frac{\mu_{\mathrm{DV}}}{\sigma_{\mathrm{DV}}} = \frac{\omega_1 \mu_1 + \omega_2 \mu_2}{\sqrt{\omega_1^2 \sigma_1^2 + \omega_2^2 \sigma_2^2}}. \tag{1.2.4}$$

Finally, recalling that $\mu_{\mathrm{signal}} = \mu_{\mathrm{noise}} = d'$, and $\sigma_{\mathrm{signal}} = \sigma_{\mathrm{noise}} = 1$:

$$d'_{1+2} = \frac{\omega_1 d'_1 + \omega_2 d'_2}{\sqrt{\omega_1^2 + \omega_2^2}}. \tag{1.2.5}$$

Note, however, that in deriving Eq. (1.2.5) we have assumed that the internal noise is Gaussian, additive, and independent, and that the process of combination is linear. As we shall see in Section 2, all of these assumptions are open to question, and expected sensitivity is liable to differ if any is violated.

2. Taxonomy

In this section we describe fourteen algorithms that a sensory system might plausible employ in order to make a judgment based on multiple sensory cues. Practical examples of each model are also given in Section A of the Supplemental Material, in the form of MATLAB code.

Some models differ in complexity, while others differ in terms of the assumptions they make about key properties of …ecision process (e.g., in terms of how error is introduced, or how it is distributed). In each case, we shall describe: (i) the key rules/principles underlying the model; (ii) how the decision variable is formed; and (iii) what the expected change in sensitivity would be for a given observer presented with one versus two cues. In particular, we specify what the upper-bound change in sensitivity would be for an ideal observer, by which an estimate of efficiency can be computed. Interested readers are also encouraged to read works by Macmillan and Creelman (2005, Ch. 6), Treisman (1998), and Wickens (2002, Ch. 10), in which some of the present information is also discussed. A broader overview of the modern cue-combination literature can also be found in Trommershauser et al. (2011).

For simplicity, we will constrain ourselves to situations where there are only two sources of information (Cue_1 and Cue_2), though many of the same principles generalize straightforwardly to more complex scenarios. We shall also further restrict ourselves to discussing simple detection tasks, rather than tasks requiring parameter discrimination, identification, or estimation. These latter tasks are liable to be more complicated to specify, and also tend to differ in terms of the language used to describe them (e.g., it makes more sense to think of sensory estimates being summed for detection, but averaged for estimation). Ultimately, however, many the same key conclusions often apply as with simple detection, including how much sensitivity is expected to improve for the ideal observer, and the form of the equation used to predict changes in sensitivity.

Overview of models

Table 1 and Fig. 2 provide overviews of the various decision models, each of which is defined in detail from Section 2.1 onwards.

Table 1 summarizes the form of each model, and provides an equation for computing multi-cue sensitivity given observed single-cue performance. It also specifies, in the final column, the expected change in sensitivity for an ideal observer. For example, given two equally informative cues, the '$\mathrm{Sum}_{\mathrm{early}}$' model (Row 7; Section 2.4) predicts up to a $\sqrt{2}$ improvement in sensitivity when a second cue is introduced. Notably, it can be seen that a given improvements in sensitivity is generally consistent with a range of possible decision models. For example, a 25% increase in d' is consistent with seven strategies, while an improvement of 100% or more is consistent with two models ('$\mathrm{Sum}_{\mathrm{add}}$', and 'Superadditivity').

Note also that the final column of Table 1 includes 95% confidence intervals, derived numerically for cases in which 100 empirical trials[2] are used to estimate d' (Grey shaded regions; for derivation see Section B in the Supplemental Material). From this, it can be seen that differences between the predictions of competing models are often liable to be rather small, relative to the amount of measurement error expected in a typical experiment. For example, to discriminate expected performance given a '1-look' or '2-look' strategy with a confidence level of 95% would require either at least 400 trials, or seven compliant observers performing 50 trials each (see Supplemental Material for derivations of confidence intervals).

An additional perspective on the various decision models is given in Fig. 2, which shows how the expected benefit of a second cue increases as the difference in sensitivity to each individual cues decreases (i.e., as the second cue becomes proportionately more useful).

2.1. One-look

In the 1-look strategy (Fig. 3), the observer bases their decision solely on a single cue (i.e., they 'look' only at one component of the

[2] 100 trials represents an arbitrary but realistic number.

3 changes on this page (all in Table 1 legend)

Table 1
Predicted sensitivity, given various models of cue-combination. See body text for details on each individual strategy. For each strategy, columns show: (C1) The name of the model; (C2) The trial-by-trial decision variable; (C3) The appropriate sensitivity metric (e.g., d' is inappropriate if the noise is not Gaussian or additive); (C4) The equation for computing predicted performance; (C5) A graphic representation of expected d' values, given single-cue sensitivities of $d'_1 = d'_2 = 1$, and assuming that the weight parameters are known and ideal. Markers in the final column show mean expected performance for an ideal observer. Where expected performance is contingent on a variable parameters (e.g., amount of correlation between noise sources), horizontal bars show the range of possible outputs. Grey shaded regions show ±1.96 *SD* of the d' sampling distribution, computed using 10,000 Monte Carlo simulations and 100 trials per simulation (i.e., 95% of d'_{1+2} estimates would be expected to fall within this region). In cases where d' is not the appropriate sensitivity metric, when computing d' values the criterion, λ, was placed such that the SDT bias metric c (see Macmillan & Creelman, 1996) was estimated to be zero. If $c \neq 0$, then estimates of d' could vary from those shown here (either higher or lower).

Model	DV (combination process)	Expected Sensitivity (d', d_a, or AUC)		Ideal d'_{1+2} if d'_1=d'_2=1 (1 – 1.5 – 2)
1-look$_{min}$	S_{min}+N_{min}	d'	$\min(d'_1, d'_2)$	
1-look$_{switch}$	S_i+N_i	d'	$\Phi^{-1}(H) - \Phi^{-1}(F)$ $H = P_1\Phi(d'_1 - \lambda_1) + (1-P_1)\Phi(d'_2 - \lambda_2)$ $F = P_1\Phi(-\lambda_1) + (1-P_1)\Phi(-\lambda_2)$	
1-look$_{max}$	S_{max}+N_{max}	d'	$\max(d'_1, d'_2)$	
Max-DV	$\max(S_1{+}N_1, S_2{+}N_2)$	d_a	$(\mu_{1+2} - 0.56)/\sqrt{(\sigma^2_{1+2} + 0.83^2)/2}$ $\mu_{1+2} = d'_1\Phi\left((d'_1 - d'_2)/\sqrt{2}\right) + d'_2\Phi\left((d'_2 - d'_1)/\sqrt{2}\right) + \sqrt{2}\phi\left((d'_1 - d'_2)/\sqrt{2}\right)$ $\sigma^2_{1+2} = \left[(1+d'^2_1)\Phi\left((d'_1 - d'_2)/\sqrt{2}\right) + (1+d'^2_2)\Phi\left((d'_2 - d'_1)/\sqrt{2}\right) + (d'_1 + d'_2)\sqrt{2}\phi\left((d'_1 - d'_2)/\sqrt{2}\right)\right] - \mu^2_s$	
2-look$_{\text{AND}}$	$(DV_1 > \lambda_1) \cap (DV_2 > \lambda_2)$	d'	$\Phi^{-1}(H) - \Phi^{-1}(F)$ $H = \Phi(d'_1 - \lambda_1)\,\Phi(d'_2 - \lambda_2)$ $F = \Phi(-\lambda_1)\,\Phi(-\lambda_2)$	
2-look$_{\text{OR}}$	$(DV_1 > \lambda_1) \cup (DV_2 > \lambda_2)$	d'	$\Phi^{-1}(H) - \Phi^{-1}(F)$ $H = 1 - \Phi(-d'_1 + \lambda_1)\,\Phi(-d'_2 + \lambda_2)$ $F = 1 - \Phi(\lambda_1)\,\Phi(\lambda_2)$	
Sum$_{early}$	$\omega_1(S_1{+}N_1) + \omega_2(S_2{+}N_2)$	d'	$(\omega_1 d'_1 + \omega_2 d'_2)/\sqrt{\omega_1^2 + \omega_2^2}$	
Sum$_{early}^{correlated}$	$\omega_1(S_1{+}N_1) + \omega_2(S_2{+}N_2)$	d'	$(\omega_1 d'_1 + \omega_2 d'_2)/\sqrt{\omega_1^2 + \omega_2^2 + 2\omega_1\omega_2\rho}$	
Sum$_{early}^{multi}$	$\omega_1(S_1{+}N_1) + \omega_2(S_2{+}N_2)$	d_a	$(\omega_1 d_{a1} + \omega_2 d_{a2})/\sqrt{\omega_1^2 + \omega_2^2}$	
Sum$_{early}^{non-norm}$	$\omega_1(S_1{+}N_1) + \omega_2(S_2{+}N_2)$	AUC	$< \Phi\left((\mu_{1+2}/\sigma_{1+2})/\sqrt{2}\right)$ $\mu_{1+2} = \omega_1\Phi^{-1}(AOC_1)\sqrt{2} + \omega_2\Phi^{-1}(AOC_2)\sqrt{2}$ $\sigma_{1+2} = \sqrt{\omega_1^2 + \omega_2^2}$	
Sum$_{late}$	$\omega_1 S_1 + \omega_2 S_2 + N_L$	d'	$\omega_1 d'_1 + \omega_2 d'_2$	
Sum$_{late}^{multi}$	$\omega_1 S_1 + \omega_2 S_2 + N_L\gamma^{(\omega_1 S_1 + \omega_2 S_2)}$	d_a	$(\omega_1\mu_1 + \omega_2\mu_2)/\sqrt{\left(1 + [\gamma^{(\omega_1\mu_1 + \omega_2\mu_2)}]^2\right)/2}$	
Sum$_{mixed}$	$\omega_1(S_1{+}N_1) + \omega_2(S_2{+}N_2) + N_L$	d'	$(\omega_1 d'_1 + \omega_2 d'_2)/\sqrt{\omega_1^2 P_{\text{E}} + \omega_2^2 P_{\text{E}} + (1{-}P_{\text{E}})}$	
Superadd.	$S_1 + S_2 + \sqrt{S_1 S_2}\delta + N_L$	d'	$d'_1 + d'_2 + \delta\sqrt{d'_1 + d'_2}$	

sensory input; see Ref. Wickens (2002) for further background on this nomenclature). The other cue is simply ignored. The observer therefore gains no benefit from multiple cues, and would never be expected to perform better than in the best single-cue condition.

To formalize the 1-look decision process, let the response to a signal-absent stimulus be 0, and let the response to a signal-present stimulus be S_1 and S_2 (for cues 1 and 2, respectively). On each trial, the stimulus response is jittered by noise samples, N_1 and N_2, each drawn independently from their corresponding noise distribution. For now, we shall assume that these distributions are zero-mean Gaussians, with standard deviations σ_1 and σ_2 (where σ represents the 'magnitude' of internal noise). We shall further assume that these values are combined additively with the initial stimulus response. Thus, the DV is $S_1 + N_1$ for Cue$_1$, and $S_2 + N_2$ for Cue$_2$. As described in Section 1.1, the relevant DV is then compared to a criterion, λ, in order to determine a response. Thus, the decision strategies for the two individual cue conditions are:

$$R_1 = \begin{cases} \text{'Signal Present'} & \text{if } (S_1 + N_1) > \lambda_1 \\ \text{'Signal Absent'} & \text{otherwise,} \end{cases} \tag{2.1.1a}$$

$$R_2 = \begin{cases} \text{'Signal Present'} & \text{if } (S_2 + N_2) > \lambda_2 \\ \text{'Signal Absent'} & \text{otherwise,} \end{cases} \tag{2.1.1b}$$

while the decision strategy in the combined condition is given by:

$$R_{\text{1-look}} = \begin{cases} \text{'Signal Present'} & \text{if } (S_i + N_i) > \lambda_{\text{1-look}} \\ \text{'Signal Absent'} & \text{otherwise,} \end{cases} \tag{2.1.1c}$$

where i in the two-cue case is equal to either 1 or 2. Note that there is no guarantee that the response criterion in the combined-cue condition (Eq. (2.1.1c)) is the same as that in either of the single-cue conditions (Eq. 2.1.1a-b). Thus, the simultaneous presence of the second cue could conceivably bias the observer towards or against a particular response. Such bias could affect some measures of performance, but should not affect d', which is specifically designed to be independent of bias (though cf. Ref. Soderquist & Lindsey, 1971). Similarly, any such bias should affect the point of subjective equality [PSE] of a psychometric function, but not its slope.

Sensitivity in the 1-look model will simply equal sensitivity in the corresponding single-cue condition. If both cues are not equally informative, then it becomes meaningful to ask which cue the observer attended to. An inefficient observer may base their decisions on the less informative of the two cues, in which case:

$$d'^{\,\min}_{\text{1-look}} = \min\left\{d'_1, d'_2\right\}. \tag{2.1.2a}$$

This could occur if, for example, the observer misjudged the relative utility of the two cues (e.g., due to a run of lucky guesses), because the poorer cue is somehow more convenient or easier to attend to, or because of some *a priori* bias in favor of that cue. Conversely, an ideal 1-look observer would base their decisions on the more informative of the two cues, in which case:

$$d'^{\,\max}_{\text{1-look}} = \max\left\{d'_1, d'_2\right\}. \tag{2.1.2b}$$

3 changes on this page

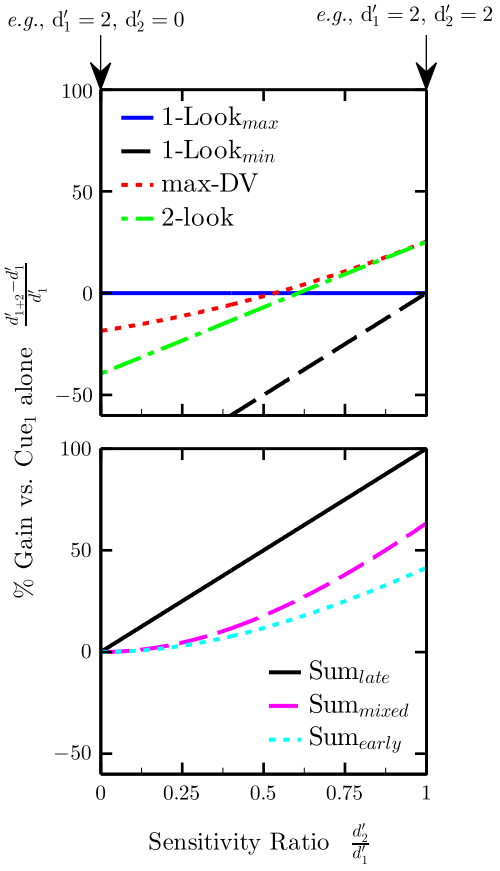


Fig. 2. Percent gain in combined-sensitivity (vs. Cue_1 alone), as relative sensitivity to Cue_2 increases. For example, in the "$\text{Sum}_{\text{early}}$" model the benefit of having a second cue increases exponentially as the difference in single-cue sensitivity decreases (Cyan dotted line, bottom panel). For display purposes only, models have been divided between two panels. Models for which d' is not an appropriate sensitivity metric are not shown (see Table 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, the observer may *alternate*: using one cue on some trials, and the other cue on other trials. Such alternation is often observed anecdotally in psychophysical experiments, where it may be variously attributed to: demand characteristics, exploration of the 'gain landscape' of the task (Gureckis & Love, 2009), an attempt to mitigate the effects of sensory adaptation, and/or simply an expression of boredom. If we assume for the moment that the trial-by-trial decision regarding which cue to attend to is made *a priori*, independent of the sensory evidence, then the result of switching is an intermediate level of sensitivity, somewhere between that of the best (d'^{max}) and worst (d'^{min}) single-cue.

More exactly, if the probability of using each cue, P_i, is known, then predicted sensitivity in the multi-cue case can be computed through linear interpolation. Note, however, that it is the raw Hit, h, and False Alarm, f, rates that must be interpolated, and not the single-cue sensitivity estimates. Thus:

$$\begin{aligned} h^{\text{switch}}_{\text{1-look}} &= P_1 h_1 + (1-P_1)h_2, \\ f^{\text{switch}}_{\text{1-look}} &= P_1 f_1 + (1-P_1)f_2, \end{aligned} \tag{2.1.3}$$

which can then be used to compute d' in the standard manner (Eq. (1.1.3)):

$$d'^{\,\text{switch}}_{\text{1-look}} = \Phi^{-1}(h^{\text{switch}}_{\text{1-look}}) - \Phi^{-1}(f^{\text{switch}}_{\text{1-look}}). \tag{2.1.4}$$

Although generally considered a poor strategy, several considerations actually favor the use of only a single cue. Firstly, finite cognitive (attention, memory) resources may prohibit the observer from attending to more than one cue. Secondly, the *cost* of attending to multiple cues may outweigh the perceived benefit. Thus, although cue-combination is often assumed to be without cost, processing a second cue may be effortful or confusing for some observers. In contrast, the gains of cue-combination are often small, either because the observer is already performing at ceiling, because the observer is content with their single-cue performance, or because the second cue genuinely provides relatively little additional information (see Fig. 2(A)). A rational observer may therefore trade-off cost against expected reward, and decide to predicate their decisions upon only a single cue. Whether observers do make such calculations, either implicitly or explicitly, is largely unknown, and to our knowledge, no attempts have been made to quantify perceived cost/benefits in sensory integration tasks.

2.2. Two-look (aka 'Probability Summation')

In the 2-look strategy (Fig. 4) – also commonly known as '*Probability Summation*' – each cue is used to make an independent

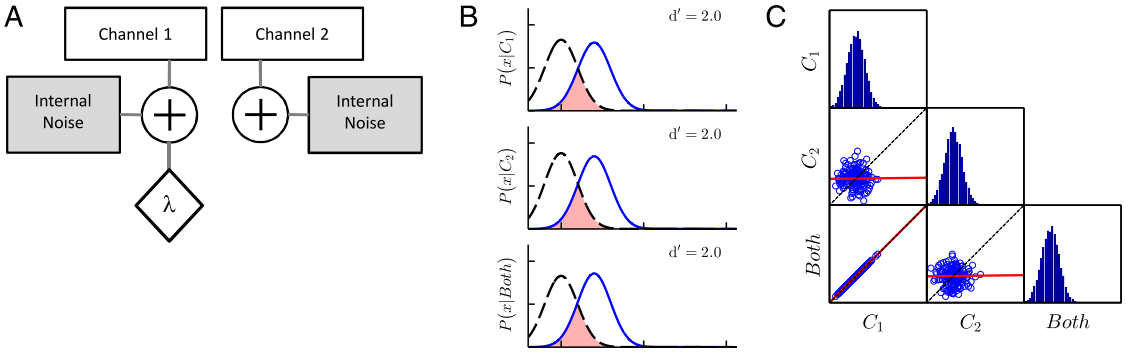


Fig. 3. 1-look cue-combination schema. (A) box-and-arrow logic diagram, showing how internal noise is added to the signals of each cue, and how signals are combined to make a decision. (B) Decision Variable (DV) distributions for signal-absent (black dashed) and signal-present (blue solid) conditions, for Cue_1 only (top), Cue_2 only (middle), and when both Cues are present together (bottom). These diagrams follow the same format as that shown previously in Fig. 1(C). Note that with the 1-look strategy, the DV in the combined condition is fully determined by the internal response to a single cue (see Eq. (2.1.1c)). The observer is therefore no more sensitive in the 'both' condition, than in either of the single cue conditions. (C) Matrix of scatter plots, showing correlations between 10,000 trial-by-trial DV values for each of the three conditions (Cue_1, Cue_2, Both; signal present condition only). Marginal histograms show the univariate distributions for each of the three conditions (i.e., these histograms correspond to the solid blue curves in Panel B). Red lines show the best fitting regression slope between each pair of conditions (flat line if no relationship between conditions). Only the lower-triangle of the matrix is shown, as the upper diagonal values are identical. In the case illustrated here, the observer only attends to Cue_1. The DV in the combined condition is therefore identical to that in the Cue_1 condition, and is totally uncorrelated with the DV in the Cue_2 condition. Alternatively, the observer could attend solely to Cue_2 (perfect correlation with Cue_2), or switch between cues (partial correlation with both cues). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

0 changes on this page

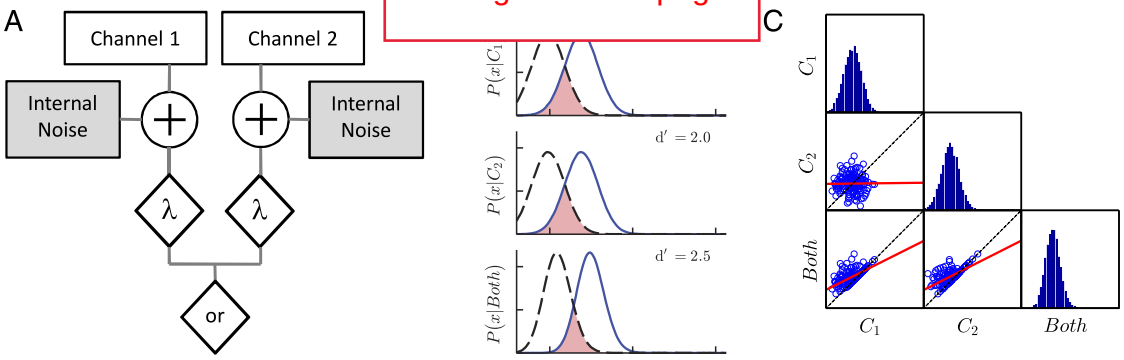


Fig. 4. 2-look$_{OR}$ cue-combination schema; same format as Fig. 3. The observer responds if *either* the Cue$_1$ or Cue$_2$ internal response exceeds criterion, λ. The DV distributions in the 'both' condition (bottom of Panel B) are the maximum of the two corresponding single-cue distributions, and upon close inspection can be seen to exhibit a slight rightward-skew.

decision (i.e., the observer 'looks' separately at each component of the sensory input). A third, 'meta' decision rule is then used to combine the two results. Thus, an observer might respond 'signal-present' if the internal response to *either* cue exceeded criterion:

$$R_{\text{2-look:OR}} = \begin{cases} \text{'Signal Present'} & \text{if } \exists i[(S_i + N_i) > \lambda_i] \\ \text{'Signal Absent'} & \text{otherwise,} \end{cases} \tag{2.2.1a}$$

or if *both* internal response exceeded criterion:

$$R_{\text{2-look:AND}} = \begin{cases} \text{'Signal Present'} & \text{if } \forall i[(S_i + N_i) > \lambda_i] \\ \text{'Signal Absent'} & \text{otherwise.} \end{cases} \tag{2.2.1b}$$

By the laws of probability, having multiple cues increases the likelihood that at least one will be successfully 'seen'. As such, expected sensitivity is greater than in any of the individual single-cue conditions. In both the AND or OR case, expected sensitivity is given approximately by:

$$d'_{\text{2-look}} \approx \frac{5}{8}d'_1 + \frac{5}{8}d'_2. \tag{2.2.2a}$$

More exactly, assuming an ideal criterion placement and additive internal noise, expected sensitivity is given by:

$$d'_{\text{2-look}} = \Phi^{-1}\left(\Phi\left(\frac{d'_1}{2}\right)\Phi\left(\frac{d'_2}{2}\right)\right) - \Phi^{-1}\left(\Phi\left(\frac{-d'_1}{2}\right)\Phi\left(\frac{-d'_2}{2}\right)\right). \tag{2.2.2b}$$

To understand how Eq. (2.2.2b) is derived, consider that the probability of both of two independent internal response exceeding criterion is equal to the product of the two individual probabilities (note that Probability "Summation" is therefore a misnomer, since probabilities are multiplied). Thus, the chance of a hit or a miss in the AND rule is given by:

$$\begin{aligned} h_{\text{2-look:AND}} &= h_1 h_2 = \Phi(d'_1 - \lambda_1)\Phi(d'_2 - \lambda_2), \\ f_{\text{2-look:AND}} &= f_1 f_2 = \Phi(-\lambda_1)\Phi(-\lambda_2). \end{aligned} \tag{2.2.3a}$$

While the chance of a hit or a miss in the OR rule (i.e., the complement of *neither* exceeding criterion) is given by:

$$\begin{aligned} h_{\text{2-look:OR}} &= 1 - (1 - h_1)(1 - h_2) \\ &= 1 - \Phi(-d'_1 + \lambda_1)\Phi(-d'_2 + \lambda_2), \\ f_{\text{2-look:OR}} &= 1 - (1 - f_1)(1 - f_2) = 1 - \Phi(\lambda_1)\Phi(\lambda_2). \end{aligned} \tag{2.2.3b}$$

In either case, using the outputs of Eqs. (2.2.3a) and (2.2.3b), d' for the combined condition can be computed in the standard manner (Green & Swets, 1974; Macmillan & Creelman, 2005; Wickens, 2002), based upon the difference in hits and false alarms (Eq. (1.1.2)). Notably, when the internal noise is additive (and signal and noise trials occur with equal frequency), then the ideal criterion location, λ, is equal to $d'/2$. Thus, if the observer is unbiased then the *difference* between hits and false alarms is equal in both the AND and the OR cases, and expected sensitivity is given by:

$$\begin{aligned} d'_{\text{2-look}} &= \Phi^{-1}\left(\Phi\left(\frac{d'_1}{2}\right)\Phi\left(\frac{d'_2}{2}\right)\right) \\ &\quad - \Phi^{-1}\left(\Phi\left(\frac{-d'_1}{2}\right)\Phi\left(\frac{-d'_2}{2}\right)\right) \\ &= \Phi^{-1}\left(1 - \Phi\left(\frac{-d'_1}{2}\right)\Phi\left(\frac{-d'_2}{2}\right)\right) \\ &\quad - \Phi^{-1}\left(1 - \Phi\left(\frac{d'_1}{2}\right)\Phi\left(\frac{d'_2}{2}\right)\right). \end{aligned} \tag{2.2.4}$$

However, when the criterion is *not* ideal (i.e., if the observer is biased in either of the single cue conditions), then the predicted sensitivities of the AND and OR strategies will differ slightly, and d' must be calculated for each strategy separately, as per Eq. (2.2.3a)/Eq. (2.2.3b).

Three further features of 2-look strategy are also worth noting. Firstly, although both the AND and OR strategies may predict the same improvement in sensitivity, they can be distinguished empirically by examining the SDT bias parameter, c (Macmillan & Creelman, 1990). Thus, the OR strategy would result in a bias towards responding 'Signal Present' (liberal observer), while the AND strategy would result in a bias towards responding 'Signal Absent' (conservative observer). Secondly, in both cases, percent correct would actually be no different to in the simple, 1-look strategy. Third and finally, note that although the 2-look strategy guarantees some improvement in sensitivity when the two cues are equally informative, when $d'_1 \neq d'_2$ sensitivity in the combined condition may actually be *lower* than in the best single-cue condition (Fig. 2(A)). This is in contrast to the linear-summation models that we consider below, where the ideal observer's sensitivity is guaranteed to increase as a function of the number of cues (Fig. 2(B)).

The formal equivalence of the* 2-look$_{OR}$ *and Max-DV models, and why* d' *is technically an invalid measure of sensitivity under either model

If two random variables are independent, then the probability that the maximum of the two will exceed criterion is equal to the probability that either of the two exceeds criterion:

$$P(\max\{X, Y\} \leq z) = P(X \leq z)P(Y \leq z). \tag{2.2.5}$$

With this in mind, it can be seen that 2-look$_{OR}$ (Eq. (2.2.1a)) is formally equivalent to a *prima facie* quite different strategy, in which the observer bases their response on whichever cue

3 changes on this page

produced the greatest internal response on that particular trial. We shall denote this strategy 'Max-DV',[3] and define it formally as:

$$R_{\text{max-DV}} = \begin{cases} \text{'Signal Present'} & \text{if } \max\{S_1 + N_1, S_2 + N_2\} > \lambda \\ \text{'Signal Absent'} & \text{otherwise.} \end{cases} \tag{2.2.6}$$

Note, however, that the maximum of two Gaussian random variables is not itself Gaussian distributed (see Ref. Nadarajah & Kotz, 2008). This is potentially problematic, because, as detailed in Section 1.1, the sensitivity metric d' implicitly assumes normality. A practical corollary of this is that empirical d' values will vary somewhat, depending on where observers place their criterion (see simulations in Section A of the Supplemental Material). This confound between bias and sensitivity is negligible when sensitivity to the individual cues is low (e.g., $d' = 1.0$), but increases when single-cue sensitivity is greater. For example, when $d'_1 = 4.0$ and $d'_2 = 4.0$, the value of d' in the combined condition may vary by approximately ±15%, depending on ~~the observer's exact criterion placement~~. The level of error also increases as the number of cues increases beyond two. For example, the maximum of 100 independent Gaussian variables exhibits a heavy rightward-skew.

For many experiments, this error is unlikely to be of substantive concern. However, for experimenters requiring exactitude, one way to deal with a deviation from internal-response normality is to measure performance as the criterion location is systematically varied (ROC analysis—see Section 2.6). In this way, the effect of criterion can be partialled out. As detailed in Section 2.6, such an analysis can provide criterion-invariant metric of sensitivity—for example, the term d^a, which we define later in Eq. (2.6.3). Notably, in the simple, two-cue case it is possible to predict d^a in the combined cue condition, using just the single cue measurements of d'. To see how this can be done, note that the DV in the Max-DV/2-look-OR model can be approximated by a Gaussian variable (Nadarajah & Kotz, 2008), with an expected mean and standard deviation of:

$$\begin{aligned} \mu_{\max} &= \mu_1 \Phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) + \mu_2 \Phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) + \theta\phi\left(\frac{\mu_1 - \mu_2}{\theta}\right), \\ \sigma_{\max} &= \left[(\sigma_1^2 + \mu_1^2)\Phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) + (\sigma_2^2 + \mu_2^2)\right. \\ &\quad \left. \times \Phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) + (\mu_1 + \mu_2)\phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) - \mu_{max}^2\right]^{0.5}, \end{aligned} \tag{2.2.7}$$

where ϕ and Φ are the probability distribution function (pdf) and cumulative distribution function (cdf) of the standard normal distribution (respectively), and where θ is determined by the internal noise magnitude for each cue, together with the degree of correlation, ρ, between them, thus:

$$\theta = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \tag{2.2.8}$$

By substituting the appropriate values for each individual cue, the DV for each of the noise ($\mu_i = 0$; $\sigma_i = 1$) and signal ($\mu_i = d'_i$; $\sigma_i = 1$) conditions can be derived. The resultant estimates of μ_{noise}, μ_{signal}, σ_{noise}, σ_{signal} can then be combined using Eq. (1.1.2) to compute expected d^a. Furthermore, since three of the DV values (μ_{noise}, σ_{noise}, σ_{signal}) are defined as constants (see Section 1.1), and if we assume that the cues are independent ($\rho = 0$; though cf. Section 2.4), then this model reduces to:

$$d^a_{\text{max-DV}} = \frac{\mu_{\text{signal}} - 0.56}{\sqrt{\frac{1}{2}\left(\sigma^2_{\text{signal}} + 0.68\right)}}, \tag{2.2.9a}$$

where

$$\begin{aligned} \mu_{\text{signal}} &= d'_1\Phi\left(\frac{d'_1 - d'_2}{\sqrt{2}}\right) + d'_2\Phi\left(\frac{d'_2 - d'_1}{\sqrt{2}}\right) \\ &\quad + \sqrt{2}\phi\left(\frac{d'_1 - d'_2}{\sqrt{2}}\right), \\ \sigma_{signal} &= \left[(1 + {d'_1}^2)\Phi\left(\frac{d'_1 - d'_2}{\sqrt{2}}\right) + (1 + {d'_2}^2)\right. \\ &\quad \left. \times\Phi\left(\frac{d'_2 - d'_1}{\sqrt{2}}\right) + (d'_1 + d'_2)\phi\left(\frac{d'_1 - d'_2}{\sqrt{2}}\right) - \mu^2_{max}\right]^{0.5} \end{aligned} \tag{2.2.9b}$$

To summarize, the 2-look$_{\text{OR}}$ decision strategy is equivalent to a strategy in which responses are based on the greatest internal response to all cues (Max-DV). In the Max-DV model, the DV is not Gaussian, and so d' is liable to misestimate sensitivity. Furthermore, Max-DV is formally identical to 2-look$_{\text{OR}}$. By extension, d' is therefore not an ideal metric for observers using the 2-look$_{\text{OR}}$/2-look$_{\text{AND}}$ strategies either. The error may be tolerable if sensitivity and/or the number of cues are low. However, in other cases alternative measures of sensitivity may be more appropriate, such as d^a (see Section 2.6). A formula for approximately predicting d^a based on single-cue d' estimates is given in Eq. (2.2.9).

Practical advantages of a 2-look decision strategy

As with 1-look strategies, 2-look (probability summation) models are generally considered relatively crude—providing only a modest improvement in sensitivity, relative to the linear-summation models that we shall consider next. However, as discussed previously with regard to the 1-look model (Section 2.1), practical considerations may make the 2-look model appealing. For example, consider a repeated-observation task. Observers are sequentially presented with N samples of the same stimulus (i.e., all 'signal present' or all 'signal absent'). They are then asked to make a single judgment as to whether or not the signal had been present throughout all the trials. To solve this task, more complex, linear-summation strategies (Section 2.3) would require the observer to remember the exact internal response value for each cue, and to estimate and store their relative reliabilities. In contrast the N-look strategy is computationally trivial. Nothing other than the chosen response needs to be stored in memory, and in terms of effort, the N-look observer can stop paying attention altogether once either a signal (if using OR) or ~~a~~ noise (if using AND) has been observed (see Section C of the Supplemental Material for example pseudocode).

Thus, an ideal observer with limited working memory might use a powerful linear-summation model (Section 2.3) when the number of cues is low, but revert to an N-look strategy (or similar), once some putative processing threshold is exceeded. In normal adults, and given only two cues, such a threshold is unlikely to be reached. However, if capacity is diminished (e.g., under conditions of high load Lavie, Hirst, de Fockert, & Viding, 2004; Repovš & Baddeley, 2006, or among children Jones, Moore, & Amitay, 2015; Luciana & Nelson, 1998 and older adults Hasher & Zacks, 1988), or if the demands of the task are increased (e.g., by increasing the number of cues), then processing costs may start to become a limiting factor. To our knowledge, this hypothesis has not been studied systematically within the cue-combination literature. However, limited supporting data do exist. For example, in the repeated-observation task described above, observers have

[3] Note, that the Max-DV rule (respond based on greatest trial-by-trial internal response) should not be confused with the 1-look$_{\text{max}}$ rule (respond based on cue with greater expected sensitivity). The Max-DV decision rule is not truly '1-look', in that all cues must be monitored before a decision is made. However, there is no integration of information across cues, and ultimately the decision is made based on information from one cue alone.

5 changes on this page

been shown to exhibit a $\sqrt{N}$ improvement in sensitivity when the number of cues is low ($N = 1$–5) (Swets, 1959). This exceeds the predictions of the N-look model, and is suggestive of an more complex linear-summation model (Section 2.3). However, when the number of cues is increased to seven (Juni et al., 2012) (i.e., close to Miller, 1956, or exceeding Cowan, 2010, the limit of human working memory), ~~relative performance~~ is diminished, and observers appear no longer able to integrate cues optimally. Similar effects have also been reported for observers asked to form 'summary statistics' (e.g., average the size or orientation) of an array of objects (Dakin, 2001; Gorea, Belkoura, & Solomon, 2014; Solomon, 2010). For example, given a mean-size-discrimination task, Solomon, Morgan, and Chubb (2011) reported that observers benefited from having two or four cues, but showed no significant improvement when the number of cues was increased to eight. The exact reason for this decrease in efficiency at high numbers of cues is unclear. For example, it may be that observers resort to a qualitatively different N-look strategy when the numbers of cues is high, or it may be that observers continue to integrate, but are quantitatively constrained in their efficiency by finite attention or memory[4] limits. Irrespective of the true explanation, that human observers behave in this way does at least highlight the importance of considering observers' limited processing resources when determining what constitutes 'ideal' performance. Thus, algorithmic limitations may prove instructive for understanding why observed sensitivity is often less than predicted in children (Gori et al., 2008; Nardini et al., 2010, 2008) and older adults (Bates & Wolbers, 2014), and in normal adults performing complex tasks (Juni et al., 2012).

2.3. *Linear summation with early, independent noise* ($\text{Sum}_{\text{early}}$)

In linear-summation models, the internal responses to each cue are summed to create the DV (see Section 1.2). Mathematically, this summation causes independent samples of noise to partially cancel out: improving the Signal to Noise Ratio [SNR], and thereby increasing sensitivity.[5] Physiologically, the process is plausible, *prima facie*, as it has long been known that individual neurons sum their dendritic inputs (Abbott, 1999), and there is growing evidence of sensory integration at the network level also (Angelaki, Gu, & DeAngelis, 2009; Fetsch et al., 2012; Gold & Shadlen, 2007).

The defining property of the *Early Noise* linear-summation model (Fig. 5) is that all the internal variability is introduced into the system prior to the two cues being combined (i.e., 'early'—see Section 2.5 for further definition of 'early' versus 'late'). Thus:

$$R_{\text{Sum:early}} = \begin{cases} \text{'Signal Present'} & \text{if } (S_1 + N_1 + S_2 + N_2) > \lambda \\ \text{'Signal Absent'} & \text{otherwise.} \end{cases} \quad (2.3.1)$$

The 'Early Noise' linear-summation model could be equivalently called the 'late integration' model, but that terminology becomes confusing when we go on to consider multiple sources of noise (i.e., situated both before and after the point of integration). The Early Noise model essentially implies that all uncertainty arises within the peripheral sensory system, and that the subsequent decision process is entirely noiseless. It also means that, when computing expected sensitivity, the final decision variable is the sum of N random variables. Typically, the noise associated with each cue is assumed to be independent, additive, and Gaussian distributed. As such, and as shown previously in Eq. (1.2.4), it follows that the final decision variable is also Gaussian distributed, with a mean equal to the linear weighted sum of the single-cue means, μ (where $\mu = d'$), and a variance equal to the linear weighted sum of the variances, σ^2 (where $\sigma = 1$). Sensitivity in the combined condition is therefore given by:

$$d'_{\text{Sum:early}} = \frac{\omega_1 d'_1 + \omega_2 d'_2}{\sqrt{\omega_1^2\sigma_1^2 + \omega_2^2\sigma_2^2}}$$

$$\text{where } \max(|\omega_i|) = 1,\ \sigma_i = 1. \quad (2.3.2)$$

When the relative weights, ω_i both equal one (equal attention paid to both cues), and when sensitivity is equal for both cues ($d'_1 = d'_2$), then this reduces to simply:

$$d'_{\text{Sum:early}} = \frac{1d' + 1d'}{\sqrt{1^2 1^2 + 1^2 1^2}} = \frac{d' + d'}{\sqrt{2}} = \sqrt{2}d'. \quad (2.3.3)$$

Notably though, Eq. (2.3.3) ceases to be the ideal decision strategy when one cue is more informative than the other ($d'_1 \neq d'_2$). In that case, the ideal weights, ~~ω_i~~, must be computed explicitly, and the appropriate values inserted into Eq. (2.3.2).

How to compute ideal relative weight coefficients, ω_i

Intuitively, it is obvious that one should give less weight to less informative cues, and more weight to more informative cues. More formally, it has been shown by previous authors that the ideal strategy is to weight each cue proportional to its signal-to-variance ratio (Jacobs, 2002; Yuille & Buelthoff, 1996). Since d' has already been scaled to be a random Gaussian variable with a variance of one ($\sigma^2 = 1$), ideal relative weights are simply proportional to d'. Given our present convention of normalizing weights so that the greatest weight magnitude, ~~$|\omega_i|$~~, always equals one (see Section 1.2), the ideal weight for the ith cue is therefore:

$$\omega_i = \frac{d'_i}{\max_{j=1}^{N}\left[|d'_j|\right]}. \quad (2.3.4)$$

For example, if $d'_1 = 1$, and $d'_2 = 1.5$, then the ideal weights would equal $\omega_1 = 0.67$ and $\omega_2 = 1.00$, and ideal sensitivity would be $d'_{1+2} = 1.58$. More generally, by combining Eqs. (2.3.2) and (2.3.4), and assuming that all d' values are positive, ideal sensitivity reduces to simply:

$$d'_{\text{Sum:early}} = \sqrt{{d'_1}^2 + {d'_2}^2}. \quad (2.3.5)$$

A common alternative to Eq. (2.3.4) is to normalize weights so that their magnitudes *sum* to one:

$$\omega_i = \frac{d'_i}{\sum_{j=1}^{N}\left[|d'_j|\right]}. \quad (2.3.6)$$

When considering sensitivity, this method of normalization is formally equivalent to the 'max' approach of Eq. (2.3.4), and results in the same quantitative predictions (e.g., in terms of d'). Conceptually, however, Eq. (2.3.4) is more natural when considering detection (where cues are thought to be 'summed'), while Eq. (2.3.4) is more suited to estimation tasks (where cues are thought to be 'average d'). As discussed previously (Section 2), we

[4] For example, Gorea et al. (2014) discuss the possibility of a limited-memory ('Markovian') linear-summation model, in which the observer maintains a running-weighted-average of the observed data. In such a model, incoming information is effectively integrated with the *average* of what has come previously. Such models are outside the scope of the present paper, but have been discussed previously by a number of authors (Ossmy et al., 2013; Usher & McClelland, 2001).

[5] Due to the Law of Large Numbers (Grimmett & Stirzaker, 2003), an extreme noise sample for one cue is likely to be partially canceled out by an opposite value for another, such that if there were an infinite number of cues the internal noise samples internal noise would be effectively zero. As an interesting aside, this is also why sensitivity in a two interval, two alternative forced choice [2AFC] task is typically $\sqrt{2}$ better than in a one interval, yes/no task (see Macmillan & Creelman, 2005, pp. 166–170), and why, in statistics, Standard Error decreases with $\sqrt{N}$ data points.

1 change on this page

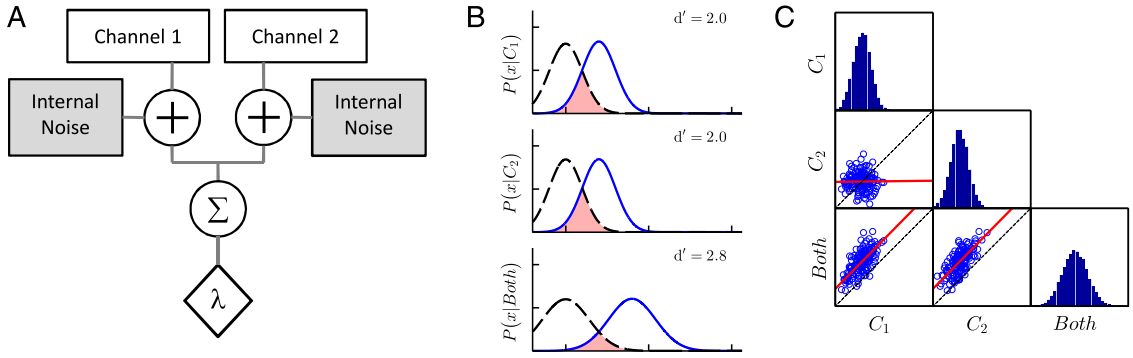


Fig. 5. Linear summation with early, independent noise; same format as Fig. 3. Independent noise is added to each cue prior to summation (Panel A). In the case illustrated here, the observer is equally sensitive to each cue ($d'_1 = d'_2$) and acts in the ideal manner by weighting each cue equally. Due to cancelation of noise between cues, the distributions in the combined cue condition therefore exhibit substantially ($\sqrt{2}$) less overlap than in either of the single cue conditions (Panel B).

shall, for simplicity, restrict ourselves to the case of detection, and so use Eq. (2.3.4) throughout. Eq. (2.3.4) also has the advantage that in many situations (e.g., when sensitivity is equal across all cues) all weights equal one; in which case, the ω_i terms can be omitted from many equations altogether.

How do observers know how to weight cues appropriately?

One possibility is that observers learn to weight cues appropriately through practice, for example via a process of supervised learning. In this case, weightings should improve over time, and, in the limiting case where the task is entirely novel, observers should be at chance on trial one. Another possibility is that reliability is estimated directly, within a single trial. This could be achieved if, for example, information was encoded by a distributed population code. Thus, if the internal response was a distribution, rather than a single scalar value, then the variance of the population-response distribution could be used as an estimate of the amount of signal uncertainty (Knill & Pouget, 2004; Ma et al., 2006; Pouget, Dayan, & Zemel, 2003). If this were the case, then observers could be capable of assigning appropriate weights to completely novel stimuli, and of responding on a trial-by-trial basis to random variations in relative SNR between cues. The evidence appears to somewhat favor the latter, 'dynamic reweighting', hypothesis. Thus, in experiments where the relative reliability of two cues is randomly varied trial-by-trial, both humans (Knill & Saunders, 2003) and macaque monkeys (Fetsch et al., 2012, 2009) have been shown to update weightings across trials. However, it is possible that such weights are also refined across trials, and there is some evidence in perceptual learning paradigms that cue weightings do improve with repeated practice (Gold, Sekuler, & Bennett, 2004; Jones, Moore, Shub, & Amitay, in press; Li, Levi, & Klein, 2004).

Expected sensitivity in the early weighted linear summation model

Under ideal conditions, with two cues of equal sensitivity, and equal weights given to each cue, sensitivity will increase by a factor of $\sqrt{2}$ (41%; as shown graphically in Table 1). More generally, ideal sensitivity will increase by $\sqrt{N}$, where N is the number of cues. If the two cues are not equally useful, but the weights are adjusted to remain ideal, then sensitivity is still guaranteed to be greater than in the best single cue (Eq. (2.3.5)). However, the magnitude of the benefit falls away exponentially as the observer's sensitivity to the two individual cues begins to differ (Fig. 2(B), black dashed line). For example, if an observer is only half as sensitive to the second cue ($d'_1 = d'_2/2$) then d' in the combined condition would only be expected to increase by a maximum of 12%. Given that estimates of d' often have a standard error on the order of ±50% (e.g., given 50 trials, see Section B in the Supplemental Material), any asymmetry will therefore greatly reduce the likelihood that observed sensitivity will improve significantly.

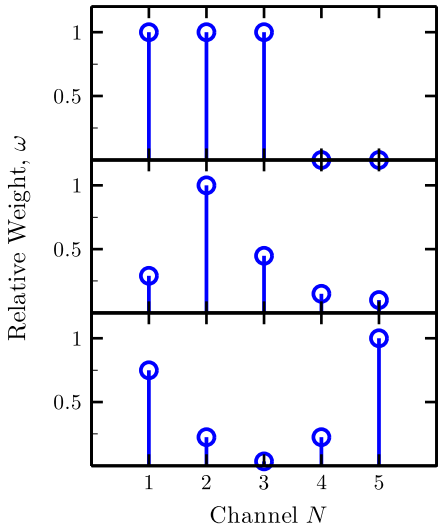


Fig. 6. Three ways of weighting multiple sensory cues ($N = 5$), for all of which the $\text{Sum}_{\text{early}}$ model predicts a $\sqrt{3}$ improvement in sensitivity (assuming that all cues are equally useful). Note that here we have scaled the weights so that the greatest weight is one, and the other weights are expressed as a fraction of this value (Eq. (2.3.4)). However, since we are only interested in the *relative* weight given to each cue, we could equivalently have normalized all the weights so that, for example, their magnitudes summed to one (Eq. (2.3.6)).

Finally, if weights are not optimal, then some information will be lost, and sensitivity in the combined condition can actually be *reduced* relative to the best single cue (Fig. 2(B), green dot-dashed line). Suboptimal weightings could occur if the observer misestimated the reliability of each cue, or if observers had some *a priori* bias towards a particular cue (for discussion, see Ref. Fetsch et al., 2009). An important, but often underappreciated point, is that although the early-noise linear-summation model predicts at most a $\sqrt{N}$ improvement in sensitivity, an increase of less than $\sqrt{N}$ does not mean that fewer than N cues are being used (as is sometimes, erroneously implied – e.g., Sweeny & Whitney, 2014). For example, Fig. 6 shows a number of relative-weight functions that would lead to a $\sqrt{3}$ improvement, only one of which involves giving no weight to two cues.

2.4. Linear summation with early, correlated noise ($\text{Sum}_{\text{early}}^{correlated}$)

A key assumption in the standard (early-noise) linear-summation model, described previously in Section 2.3, is that the internal noise associated with each cue is independent. However, this is unlikely to be strictly correct. Within a population of neurons, correlations exist between the firing patterns of individual cells (Dombeck, Khabbaz, Collman, Adelman, & Tank, 2007), and

1 change on this page

voltage signals across the cortex exhibit regular osci... over time (Salinas & Sejnowski, 2001).[6]

How would such correlations affect predicted performance? If all internal noise was correlated perfectly (Fig. 7), then the observer would gain no benefit whatsoever from having access to a second cue.[7] More realistically, internal noise may be partially correlated, with some internal variability shared between cues, and some internal variability independent across cues (Fig. 8). Mathematically, this can be expressed as follows:

$$d'_{\text{Sum:corr}} = \frac{\omega_1\mu_1 + \omega_2\mu_2}{\sqrt{\omega_1^2\sigma_1^2 + \omega_2^2\sigma_2^2 + 2\rho\omega_1\sigma_1\omega_2\sigma_2}} = \frac{\omega_1 d'_1 + \omega_2 d'_2}{\sqrt{\omega_1^2 + \omega_2^2 + 2\rho\omega_1\omega_2}}, \tag{2.4.1}$$

where ρ represents the correlation coefficient between the two internal noise sources. Notably, the ideal weights for such are model are different to those expressed previously in Eq. (2.3.4), and are given by:

$$\omega_i = \frac{d'_i - \sqrt{\rho}}{\max\left(|d'| - \sqrt{\rho}\right)}. \tag{2.4.2}$$

The formal proof of Eq. (2.4.2) is given elsewhere (Oruç, Maloney, & Landy, 2003). However, it can be understood intuitively as representing a 'relative utility' weighting, in which each cue is weighted proportional to its reliability, but only after correcting for redundancies between cues.

The effect of correlated noise is typically to reduce sensitivity in the multi-cue condition (e.g., relative to the ideal, independent, early noise linear-summation model). Specifically, the variance of the sum of two uncorrelated Gaussian variables is increased by a factor of $1 + \rho$. Thus, for example, when both cues are equally useful, then ideal (uncorrelated) sensitivity in the combined condition would be $\sqrt{2}d'$, while actual (correlated) sensitivity is:

$$d'_{\text{Sum:corr}} = \frac{\sqrt{2}d'}{\sqrt{1+\rho}}. \tag{2.4.3}$$

(N.B. when the cues are not equally useful then expected sensitivity values would have to be computed using Eq. (2.4.1).) Eq. (2.4.3) implies that a *negative* correlation between cues ($\rho < 0$) would actually cause sensitivity to *increase*. This is in fact the case, as can be seen in the simulation presented in Section A in the Supplemental Material, and follows from the fact that a noise sample for one cue would tend to be canceled out by an opposing/complimentary noise sample in the other.

***How to estimate the unknown correlation coefficient*, ρ**

One suggested technique (Wickens, 2002, pp. 184–186) for quantifying ρ is to infer the degree of correlation from the ratio of observed sensitivity, d'_{obs}, to predicted sensitivity given ...al noise, d'_{pred}:

$$\frac{d'_{\text{obs}}}{d'_{\text{pred}}} = \frac{d'_{\text{Sum:corr}}}{d'_{\text{Sum:early}}} = \frac{\frac{\omega_1 d'_1 + \omega_2 d'_2}{\sqrt{\omega_1^2+\omega_2^2+2\rho\omega_1\omega_2}}}{\frac{\omega_1 d'_1 + \omega_2 d'_2}{\sqrt{\omega_1^2+\omega_2^2}}}, \tag{2.4.4}$$

which, solving for ρ, yields:

$$\hat{\rho} = -\left(\omega_1^2 - \frac{\omega_1^2+\omega_2^2}{\left(d'_{\text{obs}}/d'_{\text{pred}}\right)^2} + \omega_2^2\right) \Big/ \left(2\omega_1\omega_2\right) = \left(\frac{d'_{\text{pred}}}{d'_{\text{obs}}}\right)^2 - 1. \tag{2.4.5}$$

The problem with this approach is that Eq. (2.4.4) assumes, implicitly, that correlation between internal noise sources is the *only* reason why observed sensitivity (the denominator) deviates from the optimal prediction (the numerator). In actuality though, observers may also deviate from the optimal prediction for other reasons: for example, because they are using suboptimal weights, or a less powerful decision strategy (e.g., 2-look). Without some independent means to rule out these other explanations, all that can be said based on sensitivity alone is that performance is *consistent* with a certain degree of internal noise correlation. However, by inspection of Table 1 (Column 3), it may also be consistent with a wide range of other explanations.

Ultimately, the only robust way to estimate ρ would be to compare the trial-by-trial internal responses to each cue (Fig. 8(C)). This would be difficult to achieve behaviorally, however, as it would require observers to make independent judgments of each cue simultaneously. This is conceptually possible, but we know of no study where this has been attempted, and the obvious concern is that the two stimuli/judgments would affect each other, either negatively (interference) or positively (compulsory integration Hillis, Ernst, Banks, & Landy, 2002; Nardini et al., 2010). Any such interaction would be evident by a change in sensitivity relative to the single-cue conditions, and would cause ρ to be over or under estimated, respectively. Alternatively, one might try to use neuroimaging to estimate an observer's internal responses directly (Tickle, Speekenbrink, Tsetsos, Michael, & Summerfield, 2016); however, such techniques are still in their infancy, and have yet to be proven robust.

2.5. *Linear summation with late noise* (Sum_{late})

Up until this point, we have assumed that all internal noise arises *early* in the decision-making process (i.e., before the point at which the internal responses to each sensory cue are integrated). However, internal noise may also be introduced at a later stage. For example, the circuits that perform the integration may themselves be subject to random error, or the observer may be unable to maintain a stable decision criterion (Jones, Moore, Shub, & Amitay, 2015; Kubovy & Healy, 1977), or the measured response of the observer may itself be stochastic (e.g., in a pointing or reaching task). Thus, Fig. 9 shows the opposite extreme, in which *all* variability is introduced after the sensory signals have been combined (Late Noise). The decision strategy for late-noise linear-summation is:

$$R_{\text{Sum:late}} = \begin{cases} \text{'Signal Present'} & \text{if } (S_1 + S_2 + N_L) > \lambda \\ \text{'Signal Absent'} & \text{otherwise.} \end{cases} \tag{2.5.1}$$

[6] Conversely, there have been reports that some neighboring cortical neurons appear to behave with statistical independence (Shadlen & Newsome, 1998). Furthermore, it may be that any correlations in part represent a shared global *gain* factor, which could in principle be factored out by an ideal decoder (Ecker et al., 2014; Goris, Movshon, & Simoncelli, 2014).

[7] NB: assuming that external noise levels are negligible. In the real world, and in some psychophysical studies also (Jesteadt, Nizami, & Schairer, 2003; Jones et al., 2013), cues may also be corrupted by external noise (i.e., noise arising from sources extrinsic to the observer, and which is therefore common across all observers). If this external noise is independently distributed between cues, then having multiple observations of it will confer a benefit in exactly the same way as with two independent internal noise sources.

0 changes on this page

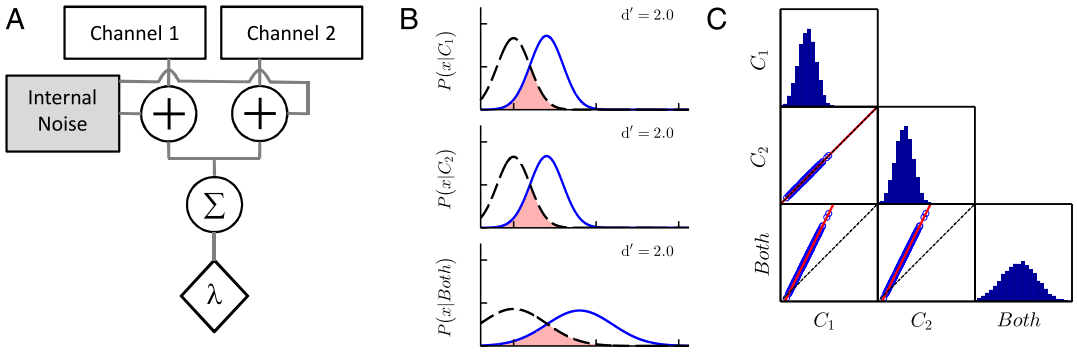


Fig. 7. Linear summation with early, correlated noise; same format as Fig. 3. As shown in Panel C, the internal responses to the two cues (Cue_1 vs. Cue_2) are perfectly correlated. Therefore, while the sum of these two variables is greater in magnitude than either alone, there is no cancelation between noise samples, and so no change in the amount of overlap between the 'Signal Absent' and 'Signal Present' distributions in the combined condition (Panel B, bottom).

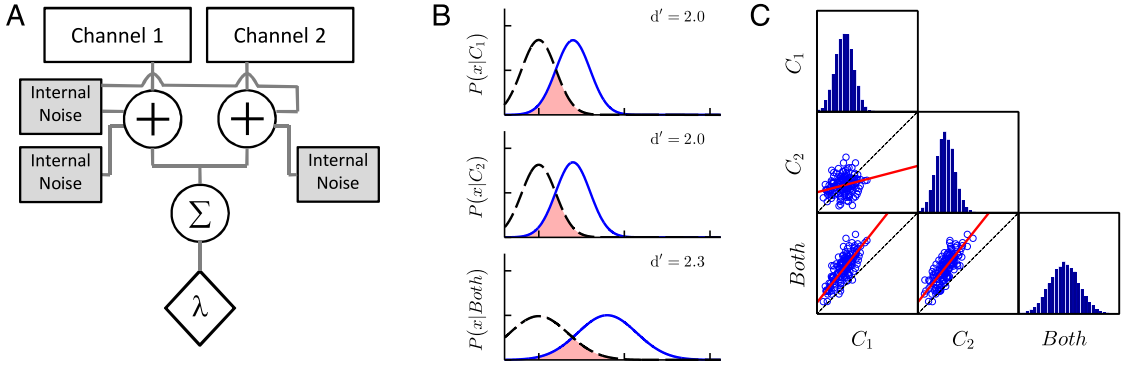


Fig. 8. Half-correlated cue-combination schema; same format as Fig. 3. Here the correlation coefficient was $\rho = 0.5$, though in practice ρ could be any value between 0 (total independence; Fig. 6) and 1 (total redundancy; Fig. 7). As a result, there was partial trial-by-trial correlation between internal responses to each (Panel C), and some increase in over-all sensitivity in the combined condition (Panel B).

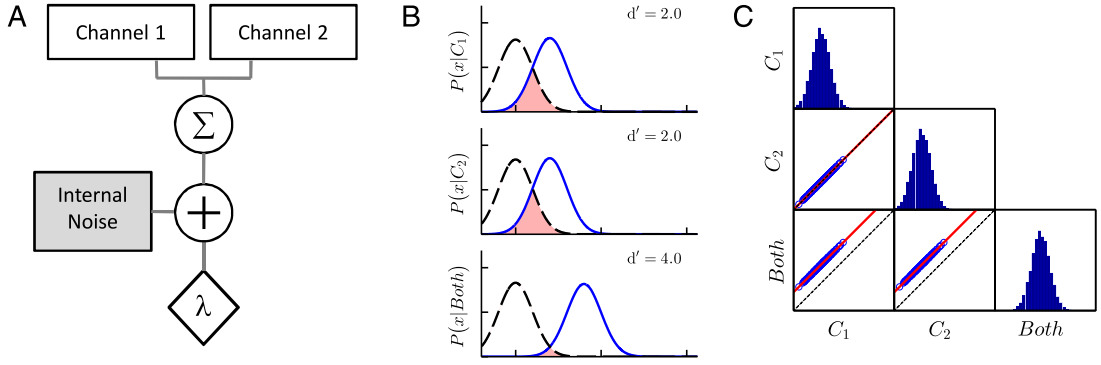


Fig. 9. Linear summation with late noise; same format as Fig. 3. In this case all internal noise is assumed to occur after the point at which the two cues are integrated (Panel A). As a result, the internal responses to each cue are perfectly correlated (Panel C). However, unlike in the correlated early-noise condition (Fig. 7), internal noise is constant, and does not increase as a function of N cues. Therefore, there sensitivity doubles when two equally-informative cues are optimally combined (Panel B).

It is important to note that the noise in this model, N_L, is 'late' relative to the point of integration, not in absolute terms within the neural processing hierarchy. Thus, stochastic variation in one brain region could potentially be a source of early noise for one pair of cues, but a source of late noise with respect to another pair of cues (Fig. 10). It is therefore not strictly correct to equate early/late noise with sensory/cognitive noise, although at times it may be convenient to do so (and in general the two dichotomies are likely to be related).

If, as is typically the case in psychophysical paradigms, the amount of external noise is minimal, the ideal late-noise observer should always weight each cue equally, as each cue provides only signal (in contrast to the early-noise model, where each cue also contributes additional noise). Ideally, the relative weights should therefore all equal 1, and the weight terms, ω, can simply be omitted. Accordingly, ideal performance is given by:

$$d'_{\text{Sum:late}} = \frac{1d'_1 + 1d'_2}{1} = d'_1 + d'_2. \tag{2.5.2a}$$

However, an observer may in practice under-weight one or more cues. In that case, the improvement in sensitivity would be less than the arithmetic sum($<d'_1 + 2d'_2$), and could be computed explicitly as:

$$d'_{\text{Sum:late}} = \frac{\omega_1 d'_1 + \omega_2 d'_2}{\sigma_L} \quad \text{where } \max(|\omega_i|) = 1, \sigma_L = 1. \tag{2.5.2b}$$

It is clear that the potential gains are far larger than with the models discussed previously. Thus, if all cues are equally

0 changes on this page

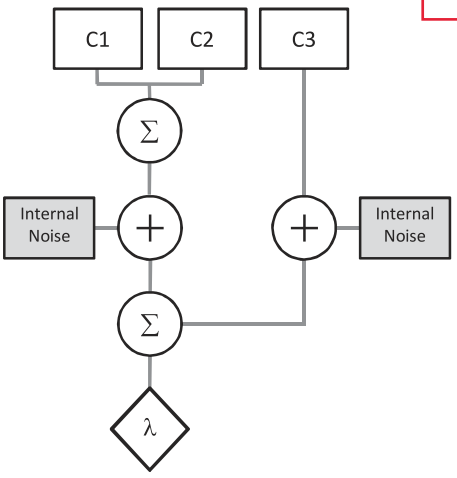


Fig. 10. Early versus late noise. The leftmost noise source would be considered early when comparing cues C1 and C2, but late when comparing cues C1 and C3.

informative then ideal sensitivity will increase by a factor of N, versus the $\sqrt{N}$ predicted by the early-noise linear-summation model (Eq. (2.3.3)). A minority of authors have reported cases of factor-of-N improvements (Saarela & Landy, 2012). However, for most tasks the late-noise model is actually logically implausible in its extreme form. For example, it would imply that there was some location in the brain where the two stimuli are represented perfectly (i.e., such that they could be summed together prior to any noise being introduced). Since even the peripheral mechanisms that encode incoming sensory information act in a stochastic manner (Field & Rieke, 2002; Jaramillo & Wiesenfeld, 1998), it is difficult to imagine where in the brain such a noiseless representation could exist. A more realistic scheme is therefore the hybrid model (Fig. 11), in which some noise is early, and some late:

$$R_{\text{Sum:late}} = \begin{cases} \text{'Signal Present'} \\ \quad \text{if } (S_1 + N_1 + S_2 + N_2 + N_L) > \lambda \\ \text{'Signal Absent'} \quad \text{otherwise.} \end{cases} \tag{2.5.3}$$

Expected performance in the hybrid model is given by:

$$d'_{\text{Sum:mix}} = \frac{\omega_1 d'_1 + \omega_2 d'_2}{\sqrt{\omega_1^2\sigma_1^2 + \omega_2^2\sigma_2^2 + \sigma_L^2}}$$

$$\text{where } \max(|\omega_i|) = 1, [\sigma_i + \sigma_L] = 1. \tag{2.5.4}$$

When the two cues are equally useful, the ideal weights will all remain equal to one (as in both the early- and late-noise models). When one cue is more useful than the other (Fig. 2(B)), the weights will depend on the ratio of early- to late-noise (see Section A in the Supplemental Material). If the late noise dominates, the ideal weights will all continue to equal (or approximately equal) one, as in the late-noise model. If the early noise dominates, the weights will be primarily determined by the relative reliability of the two cues, as in the early-noise model (Eq. (2.3.4)). In each case, sensitivity will lie somewhere between the predictions of the early- and late-noise models. For example, consider the case where signal and noise are equal in magnitude for all cues. If the majority of noise is late, then the benefit will be large:

$$d'_{\text{mix}} \approx 2^{\frac{1}{1}} d' = N d'. \tag{2.5.5a}$$

If the majority of the noise is early, the benefit will be smaller:

$$d'_{\text{mix}} \approx 2^{\frac{1}{2}} d' = \sqrt{N} d'. \tag{2.5.5b}$$

If half the noise is early and half is late, then the benefit will be an intermediate value:

$$d'_{\text{mix}} \approx 2^{\frac{1}{\sqrt{2}}} d' = \sqrt[\sqrt{2}]{N} d'. \tag{2.5.5c}$$

Empirical separating early and late noise

Is it possible to determine what proportion of noise is early/late on the basis of sensitivity alone? In principle, one could attempt to infer the ratio of early-to-late noise by comparing observed sensitivity to the ideal. For example, if d' equals Eq. (2.5.5a) then one might claim that all noise is late, or if d' equals Eq. (2.5.5b) then one might claim that all noise is early. However, as discussed previously with regards to estimating internal-noise correlation (Section 2.4), the difficulty with this inference-from-efficiency approach is that it requires us to assume that no other factors limit performance. Thus, the level of sensitivity predicted by early noise (Eq. (2.5.5a)) could also be explained by 100% late noise and suboptimal weights (see Fig. 2(B), dot-dashed green line), or by a mixture of late noise and correlated early noise.

Instead then, one possible way to disambiguate early and late noise would be to examine how sensitivity changes with cue asymmetry. Thus, it was noted earlier that multi-cue sensitivity decreases exponentially as the difference in single-cue sensitivity increases (Fig. 2(B), red dotted). In contrast, with late-noise, benefit varies linearly with cue asymmetry (Fig. 2(B), solid line), and

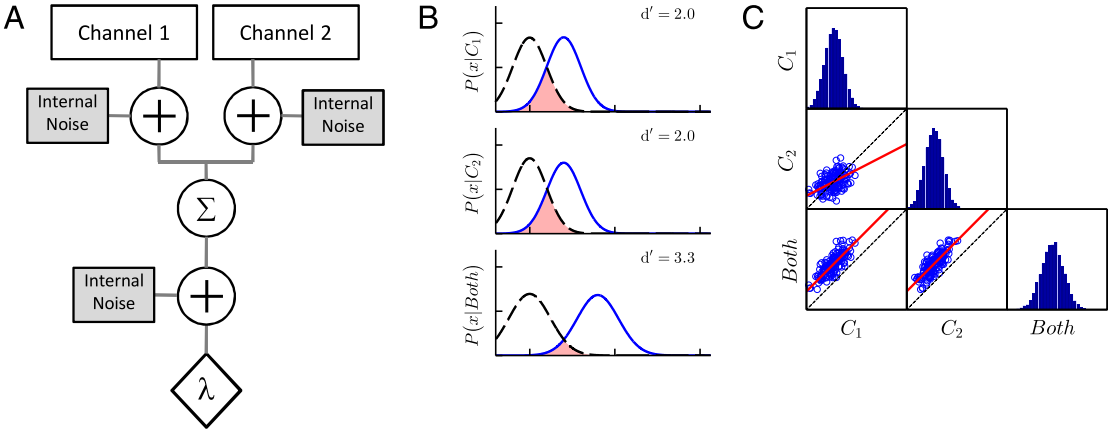


Fig. 11. Linear summation with half late noise, and half early noise; same format as Fig. 3. The early noise components are independent across cues, so the internal responses to each are only partially correlated (Panel C; Cue_1 vs. Cue_2). The overall increase in sensitivity (Panel B) is an intermediate value, between the lower bound of the early, independent noise model (Fig. 5), and the upper bound of the pure late noise model (Fig. 9).

various admixtures of early/late noise will fall along intermediate isobars (e.g., Fig. 2(B), black dashed). These differing predictions could ~~be used in principle~~ to differentiate between relative quantities of early vs. late noise. We are not aware of any attempts to perform such an analysis, and most existing datasets would be unsuitable, since experimenters tend to design cues to always be equally useful (since this is when expected gain is maximized—see Section 3.2). Moreover, such a technique implicitly assumes that the late noise is additive, and does not vary with combined internal response magnitude; an assumption which is itself open to question.

It is also important to note that the ratio of early-to-late-noise may not be stationary. For example, it has been well-established that sensory judgments improve with practice (*perceptual learning* Fine & Jacobs, 2002; Wright & Fitzgerald, 2001), suggesting that sensitivity changes over time. This can lead to some interesting predictions regarding how cue-combination strategies may differ between naïve and well-trained observers. For example, it has been shown that some of the learnt improvements in sensitivity can be attributed to reductions in nonstationary bias (Jones et al., 2015), which for present purposes can be considered a source of late internal noise.[8] Thus, while sensitivity may primarily reflect sensory limitations in well-trained observers (early noise), naïve observers may be more limited by late noise. Such considerations lead to the novel, and somewhat counterintuitive prediction, that naïve observers should gain *more* benefit from having multiple cues than well-practiced observers do (i.e., since before practice, decisions will be limited partially by late noise, and so sensitivity should increase by more than the Pythagorean sum of the two cues). In contrast, after practice, sensitivity should be determined by early noise alone, and so follow Eq. (2.3.3). If a greater proportion of internal noise were 'late' in naïve observers, then this would also mean that the ideal weights for a naïve and a trained observer would differ when the cues are of unequal utility. Thus, in well-practiced observers each cue should be weighted proportional to its reliability (as per the early-noise model), whereas in naïve observers the weights should tend more towards equality, irrespective of any asymmetry (as per the late-noise model). This may be an important consideration when determining whether, for example, observers can learn to optimize their decision weights with practice (Gold et al., 2004; Jones et al., in press; Li et al., 2004) (i.e., since an observer could conceivably change their weights while remaining equally efficient/inefficient).

Finally, it is worth noting that there is a potential contradiction between the correlated-early-noise model and the late-noise linear-summation model. The ~~late-noise~~ model predicts that when cues are similar, the information should be integrated early (i.e., more noise should be late), and so the benefits of integration should be *greater*. Conversely, it might be argued that when two cues activate similar sensory regions, more of the noise should be correlated across cues, and so the benefits would be *smaller*. The evidence tends to favor the former viewpoint, with the benefits of integration being greater when the two cues are more similar. Thus, when the cues are located in different modalities (Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002; van Beers, Wolpert, & Haggard, 2002) or spatial locations (Redmond, Zlatkova, Vassilev, Garway-Heath, & Anderson, 2013), improvements tend to follow the early-noise predictions (Eq. (2.3.2)). In contrast, reported improvements for two visual-depth-cues (Saarela & Landy, 2012), or two nearby retinal locations (Redmond et al., 2013), have tended to be greater, and to follow the late-noise prediction (Eq. (2.5.2b)).

2.6. Linear summation with multiplicative noise ($\text{Sum}^{multi}_{\text{early/late}}$)

Up until this point, we have assumed that all internal noise is additive. In reality though, internal noise may be *multiplicative* (Fig. 12), such that the degree of internal variability, σ, varies with the magnitude of the internal response, μ. We begin by considering the implications if the additivity assumption is breached, before considering the evidence for or against it.

The first point to note is that the effects of any multiplicative noise depend on whether on whether the internal noise is early or late. In the Early Noise model, the multiplicative gain is applied before the signals have been summed;

$$R^{\text{early}}_{\text{Sum:multi}} = \begin{cases} \text{'Signal Present'} \\ \quad \text{if } \left(S_1 + N_1\gamma^{S_1} + S_2 + N_2\gamma^{S_2}\right) > \lambda \\ \text{'Signal Absent'} \quad \text{otherwise.} \end{cases} \tag{2.6.1}$$

In both the single-cue and combined-cue condition, the strength of the sensory signal (and thus the amount of multiplication) is identical at the point when internal noise is introduced. The fact that the noise is multiplicative therefore has no effect on the predicted benefit of integration, which remains unchanged from the additive early noise model (Eq. (2.3.3)). The caveat to this, however, is that d' is no longer a valid measure of sensitivity when internal noise is multiplicative. Thus, recall that for a single parameter, d', to fully constrain sensitivity, it is necessary to assume that the signal and noise distributions have equal variance (Macmillan & Creelman, 1996; Wickens, 2002) (see Eq. (1.1.3)). Equal variance cannot be the case with multiplicative noise, since, tautologically, the signal magnitude will be greater in the 'signal' condition than in the 'noise' condition ($\gamma^S > \gamma^0$). Instead, recalling Eq. (1.1.2), sensitivity must therefore be estimated using the more general sensitivity measure, d^a:

$$d^a_i = \frac{\mu_{\text{signal}} - 0}{\sqrt{\frac{1}{2}\left(\sigma^2_{\text{signal}} + \sigma^2_{\text{noise}}\right)}} \neq \frac{\mu_{\text{signal}}}{\sqrt{\frac{1}{2}\left(\sigma^2 + \sigma^2\right)}}. \tag{2.6.2}$$

As discussed extensively elsewhere (Macmillan & Creelman, 1996, 2005; Wickens, 2002), d^a can be estimated empirically, using the equation:

$$d^a = \frac{\sqrt{2}a}{\sqrt{1+b^2}} \tag{2.6.3}$$

where a and b are, respectively, the intercept and slope of an observed Receiver Operator Characteristic (Macmillan & Creelman, 2005; Swets, 1973; Wickens, 2002) (ROC; see Fig. 13(A)). Failure to use the appropriate measure of sensitivity may lead to spurious variations in apparent sensitivity, depending on where precisely the observer places their criterion.

Fortunately, however, estimates of d^a can be used to predict multi-cue performance in exactly the same manner as with d' (Eq. (2.3.2)), thus:

$$d^{a\ \text{early}}_{\text{Sum:multi}} = \frac{\omega_1 d^a_1 + \omega_2 d^a_2}{\sqrt{\omega_1^2\sigma_1^2 + \omega_2^2\sigma_2^2}} = \sqrt{d_1^{a\,2} + d_2^{a\,2}}. \tag{2.6.4}$$

With d^a, ideal weights remain proportional to the sensitivity values of the two cues, as per Eq. (2.3.2). Thus, as with d', the ideal linear-weighted sum of two equally useful cues will cause d^a to improve by $\sqrt{2}$ (41%), and this benefit will diminish exponentially

[8] NB: response bias ought to be independent of sensitivity, but, as discussed in Ref. Jones et al. (2015), the two factors are liable to confounded when the bias varies between trials. Furthermore, although bias fundamentally reflects a *deterministic* process, it may still be considered a source of 'random noise' for present purposes, so long as it is uncorrelated with the task-relevant information.

2 changes on this page

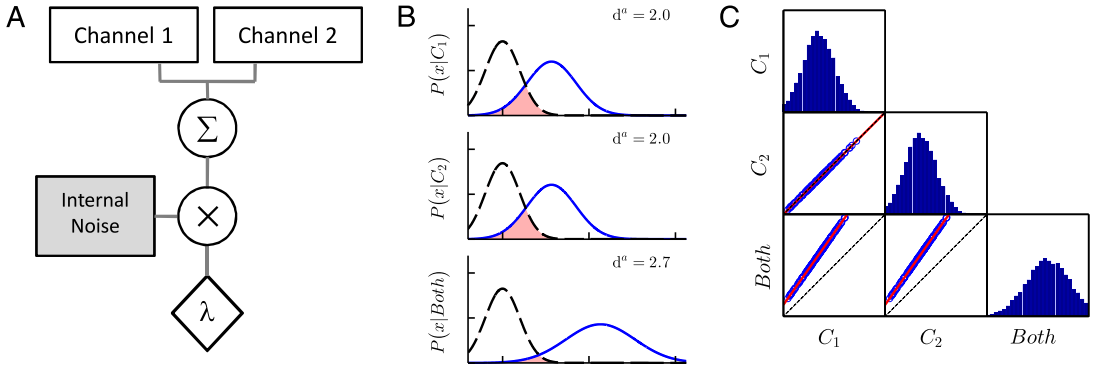


Fig. 12. Linear summation with multiplicative noise; same format as Fig. 3. In the example shown here, all internal noise is ~~assumed to be~~ late (Panel A). Since in this model internal noise magnitude increases as a function of signal strength, internal noise magnitude is greater in the Signal Present condition (Panel B; solid blue line) than in the Signal Absent condition (Panel B; dashed black line). This makes d^a a more appropriate metric of sensitivity than d'. Furthermore, internal noise magnitude will increase more as N cues increases (Panel B; bottom). As a result, sensitivity in the compound-cue condition will increase by less than that predicted by the additive late noise model (Fig. 9). In this case, the multiplicative factor is relatively low ($\gamma = 1.19$), so the overall increase in sensitivity is similar to the linear summation model with early, independent noise (Fig. 5). If γ were greater, the increase in sensitivity would be smaller, and could even become negative. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

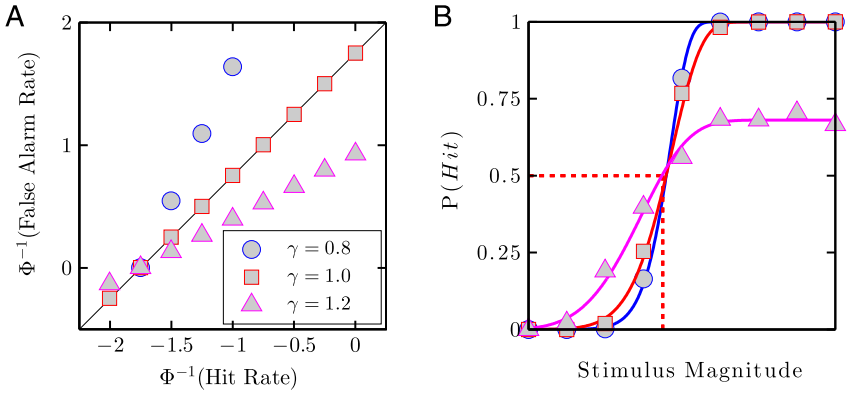


Fig. 13. Schematic tests of multiplicative noise, using (A) ROC analysis, and (B) Psychometric analysis. When internal noise is sublinear (purple triangles), internal noise is proportionally greater at higher stimulus levels, resulting in a smaller improvement in sensitivity than in the linear condition (red square). The supralinear (blue circle) condition results in the opposite asymmetry (relatively less noise at higher signal levels). As detailed elsewhere (Macmillan & Creelman, 2005; Wickens, 2002), an ROC curve is constructed by measuring hit rate and false alarm rates as the criterion, λ (see Fig. 1), is systematically shifted (e.g., by varying the benefit/cost of Hits/Misses). The psychometric function shown here plots proportion of hits, which is expected to increase monotonically as a function of stimulus magnitude. See body text for details. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as one cue is made less informative (Fig. 2(B)). In short then, if the multiplicative noise is *early*, then the practical method of computing sensitivity differs from the additive case, but predicted change in sensitivity remains invariant.

In contrast, if the internal noise is *late* (Section 2.5) then the presence multiplicative noise markedly affects predicted sensitivity in the multi-cue condition. For instance, let the level of multiplication be represented by the gain constant: γ^{S_i} (where S_i is the average internal response magnitude in the ith cue). In the late-noise model, this gain is applied *after* the individual signals have been summed, thus:

$$R^{\text{late}}_{\text{Sum:multi}} = \begin{cases} \text{'Signal Present'} \\ \quad \text{if } \left(S_1 + S_2 + N_L\gamma^{(S_1+S_2)}\right) > \lambda \\ \text{'Signal Absent'} \quad \text{otherwise.} \end{cases} \tag{2.6.5}$$

This introduces a non-linearity, such that the magnitude of noise in the multi-cue condition may not equal the sum of noise magnitudes in the single cue conditions:

$$d^a_{\text{Sum:multi}}{}^{\text{late}} = \frac{\omega_1 d_1^a + \omega_2 d_2^a}{\frac{\sqrt{[\sigma_L\gamma^{(d_1^a+d_2^a)}]^2+\sigma_L^2}}{2}} \neq \frac{\omega_1 d_1^a}{\sigma_L\gamma^{d_1^a}} + \frac{\omega_2 d_2^a}{\sigma_L\gamma^{d_2^a}}$$
$$\neq \omega_1 d_1^a + \omega_2 d_2^a. \tag{2.6.6}$$

When $\gamma = 1$, Eq. (2.6.6) reduces to the additive late noise model, and sensitivity in the multi-cue condition will equal the arithmetic sum of the individual sensitivities (Eq. (2.5.2a)). When $\gamma < 1$, the integration is *supralinear* (sensitivity will increase at a rate greater than predicted by the late noise model). When $\gamma > 1$, the integration is sublinear (sensitivity will increase at a rate less than that predicted by the late noise model). Notably though, even if the gain parameter, γ, were known, the unobservable DV parameters μ and σ would also have to be known in order to predict ideal performance in a combined-cue model (i.e., rather than the sensitivity ratio μ/σ, indexed by d').

In short ~~then~~, the possibility of multiplicative late noise complicates greatly any attempt to draw inferences from changes in sensitivity. *Any* observed sensitivity is consistent with *some* model in which the observer's decisions are limited by a source of internal noise that is multiplicative and late. This further complicates previous considerations, since now an improvement of less than Nd' could be caused by multiplicative late noise, as well as by independent early noise, correlated noise, a poorer decision strategy, or suboptimal decision weights.

Traditionally, however, it has been common to assume that levels of multiplicative noise are negligibly small. In part, this reflects mathematical convenience. However, there are also empirical arguments both for and against the presence of

multiplicative noise. In favor of multiplicative noise being present, there is converging evidence from psychophysics (McGill & Goldberg, 1968; Swets, 1959; Watson, Rilling, & Bourbon, 1964) and neurophysiology (Goris et al., 2014; Teich & Khanna, 1985; Tolhurst, Movshon, & Thompson, 1981; Winter & Palmer, 1991; Young & Barta, 1986). For example, Weber's law[9] is often taken to indicate the presence of a limiting noise source that increases with stimulus strength. Similarly, single cells are often shown to exhibit Poisson-like processes, with spike-rate variability increasing as a function of mean firing rate. *Prima facie*, these appear good evidence of multiplicative noise in decision-making. However, in the context of cue-combination, the arguments are misleading. Weber's law is typically observed over large (order of magnitude) stimulus ranges. In contrast, performance around threshold, which we are principally interested in, may be approximately, locally linear.[10] Furthermore, decisions are likely to be driven by populations of neurons, rather than individual cells. In that case, it follows from the Central Limit Theorem that the total product of the individual, Poisson-like process will rapidly converge to a Gaussian as the number of neurons increases (Ma et al., 2006).

In contrast, the evidence that multiplicative noise is of ~~relatively minor~~ concern is more robust. For example, additivity can be assessed empirically by constructing ROC curves. As shown in Fig. 13(A), when integration is linear (additive noise), the curve will have unit slope when plotted on Gaussian-transformed coordinates. In contrast, sublinear and supralinear conditions produce shallower or greater slopes, respectively. Such curves do not tend to be measured in studies of cue integration, but have been studied on a range of more basic sensory judgment tasks, where they tend to have approximately unit slope (albeit with some substantial variation, e.g., 0.5–2.0 Swets, 1986). Similarly, the effects of multiplicative noise should also be apparent in psychometric functions (Tyler & Chen, 2000) (Fig. 13(B)). For example, a sublinear process (greater noise in the combined condition) would result in an asymmetric function with a suppressed asymptote at high stimulus magnitudes ('saturation'; purple triangles). This is not typically observed in the types of tasks used in cue-integration studies, again allowing us to discount a substantial source of multiplicative noise.

To summarize, although the evidence is not conclusive, it appears that multiplicative noise is of relatively minor importance to decision making at a behavioral level. Multiplicative noise is, however, likely to be present to some extent in all tasks, and may vary in relative magnitude across tasks. If internal noise is early, then such noise will not affect predicted sensitivity. However, to the extent that internal noise is late, multiplicative noise may either slightly increase or depress ideal sensitivity.

2.7. Superadditivity

Superadditivity describes a class of decisions rules (Fig. 14) in which sensitivity to the combined stimulus is greater than the arithmetic sum of the individual sensitivities: $d'_{1+2} > (d'_1 + d'_2)$.

As with multiplicative noise, superadditivity introduces a nonlinearity into the decision process. Unlike multiplicative noise, this non-linearity: (i) is generally thought of as being applied to the signal rather than the noise (i.e., the numerator of Eq. (1.1.2)), (ii) depends on the presence of both activity in both cues, and, (iii) by definition, is always superlinear ('1 + 1 = 3'). An example superadditive decision rule is:

$$R_{\text{super}} = \begin{cases} \text{'Signal Present'} \\ \quad \text{if } \left(S_1 + S_2 + \sqrt{S1S2}\beta + N_L\right) > \lambda \\ \text{'Signal Absent'} \quad \text{otherwise,} \end{cases} \tag{2.7.1}$$

where β is some putative superadditivity mechanism. Sensitivity in such a model is specified by:

$$d'_{\text{super}} = \frac{d'_1 + d'_2 + \sqrt{d'_1 d'_2}\beta}{\sigma_L} \quad \text{where } \sigma_L = 1. \tag{2.7.2}$$

This predicts that sensitivity will always increase by a factor greater than the late-noise linear-summation model ($>N$).

How plausible is superadditivity? The notion of superadditivity is made credible, *prima facie*, by studies of physiology. For example, neurons in the Superior Colliculus (Perrault, Vaughan, Stein, & Wallace, 2005; Stanford, Quessy, & Stein, 2005; Stein & Stanford, 2008) and Superior Temporal Sulcus (Stevenson & James, 2009; Werner & Noppeney, 2010) have been shown to fire more than twice as often when presented with corresponding information from two modalities (e.g., sight and sound), versus either in isolation. This has been argued to reflect the linear summation of membrane potentials (Skaliora, Doubell, Holmes, Nodal, & King, 2004) followed by a static nonlinearity (threshold) in spike generation (Holmes & Spence, 2005).

However, as with multiplicative noise, it may be misleading to draw inferences regarding system-level decision-making from the dynamics of single-cells, and studies at the behavioral level have seldom reported superadditivity. Moreover, there is doubt over the extent to which, even in physiology, superadditivity generalizes beyond situations where both inputs are very weak/subthreshold (Morgan, DeAngelis, & Angelaki, 2008). Nonetheless, the possibility of superadditivity cannot be ruled out completely, and so caution is advised when attempting to infer decision efficiency from observations of sensitivity alone. For example, based on sensitivity alone, it is impossible to distinguish an early-noise linear-summation observer with ideal weights, from a superadditive observer with suboptimal weights.

2.8. Non-Gaussian noise models ($\text{Sum}^{\textit{non-norm}}_{\text{early/late}}$)

So far, we have relaxed the assumptions of independence, additivity, and linearity. The final assumption in most models of cue-combination is normality. Thus, while it is common to assume that internal noise is Gaussian distributed, in principle it may actually take many other forms, such as the noncentral t-distribution shown in Fig. 15(B).

What effect do deviations from normality have on expected cue-combination behavior? The answer is: surprisingly little (see Ref. Oruç et al., 2003). In theory, deviations from normality are a substantial complication. For example, as shown in Fig. 16, when the noise is highly asymmetric, performance in the combined-cue condition ceases to be predicted by the Pythagorean sum of the individual sensitivities (Eq. (2.3.5)), and the ideal weights for highly asymmetric noise deviate from those predicted by the Gaussian early-noise model (Eq. (2.3.4)). In practice though, the differences tend to be slight, and tend rapidly towards zero as the amount of skew reduces.

Furthermore, there is good reason to think that a strong departure from normality is unlikely. The arguments largely recapitulate those against multiplicative noise. In brief, the Central

[9] Weber's law states that the Just Noticeable Difference between two stimuli is proportional to their magnitude.

[10] One influential experiment by the early proponents of Signal Detection Theory actually appeared to indicate that substantial multiplicative noise *is* present at threshold (see Ref. Swets, Tanner, & Birdsall, 1961: "*Theoretical and Experimental Analysis of Second Choices*"). However, as discussed by Solomon (2007), the same data can be explained by other models of detection, such as an Intrinsic Uncertainty model in which perceptions are dictated by the maximum activity across multiple independent cues.

5 changes on this page (all in figure captions)

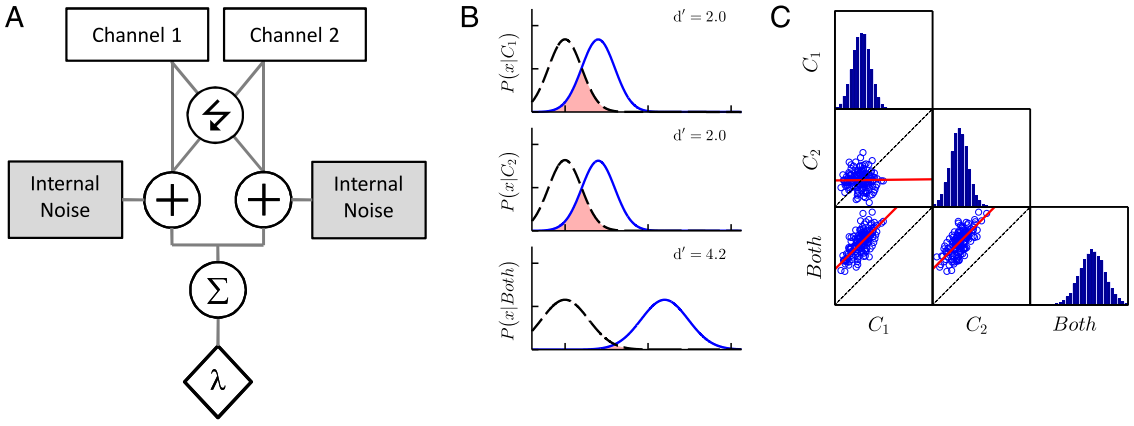


Fig. 14. Superadditive cue-combination schema; same format as Fig. 3. Some putative mechanism ~~is assumed to~~ increase the internal response to each cue~~,~~ when both are observed simultaneously (Panel A). As a result, sensitivity in the combined condition can increase by more than the sum of the individual cue conditions: $d'_{1+2} > (d'_1 + d'_2)$.

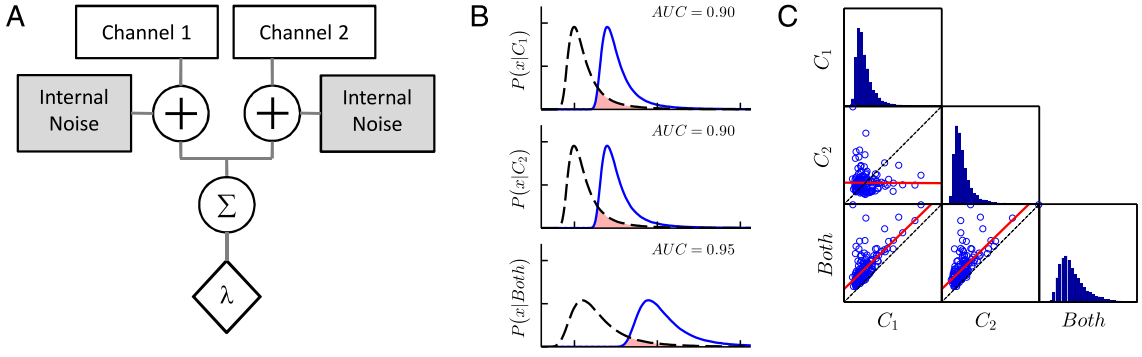


Fig. 15. Non-Gaussian cue-combination schema. Same format as Fig. 3. In this arbitrary example, internal noise ~~is assumed to~~ follow a noncentral t-distribution (an asymmetric distribution with a heavy right tail, which approximates normality as the shape parameter, v, increases). Due to this asymmetry, the non-parametric 'Area Under Curve' metric is a more appropriate measure of sensitivity than d' (Panel B).

Limit Theorem makes a strong asymmetry theoretically unlikely, and the fact that ROC curves tend to be linear on Gaussian-transformed coordinates is empirical evidence for normality.

3. Summary and comparisons between models

The question 'can people combine multiple sensory cues?' is trivial to answer empirically. If, for example, sensory judgments are better given sight and sound together than either alone, then it follows that observers are exploiting both sources of information. However, to quantify how well the observer is combining the information available, a measure of efficiency must be derived. This requires some hypothetical benchmark of what 'ideal' performance would be, which in turn requires a model of decision making. In Section 2, we described a range of such models, each of which predicts a quantitatively different level of ideal sensitivity.

In some models of decision-making, only a single internal response value is used, and the others discarded (*1-look, max-DV*). In others, multiple decisions are made independently, based upon each individual variable (*N-look*). Finally, in the most powerful strategies the decision is based on the linear-weighted sum of N random variables. Exactly how this integration takes place, however, depends on various properties of the model, such as whether the internal noise is independent for each cue, whether it combines additively with the signal, whether it occurs early or late, and whether it is normally distributed.

3.1. What can and cannot be inferred from observed sensitivity

As has been discussed, some models of cue-combination are more plausible than others. However, none of them can be ruled out *a priori*. Based on sensitivity alone, some models make quantitatively distinct predictions, and so can be delineated empirically (Table 1). For example, a $\sqrt{N}$ improvement is consistent with linear summation and superadditive models, but not with simpler 2-look or max-DV models. However, in many cases multiple models make overlapping predictions, and this is particularly the case once expected measurement error is taken into account. Accordingly, when it comes to understanding multi-cue decision making, surprisingly little can be inferred from changes in sensitivity alone. In general, any observed level of sensitivity is consistent with a range of possible decision models, and so may be more or less close to the ideal, to an unknown degree. More specifically:

1. An improvement in sensitivity relative to the best single cue is not proof of linear-weighted summation. To evidence such a process, performance must not just be better than the best single cue, but also better than any alternative cue-combination strategies would predict. The appropriate comparison is therefore not to the individual cues (e.g., Petrini, Remark, Smith, & Nardini, 2014; Weisswange, Rothkopf, Rodemann, & Triesch, 2011), but to the max-DV and N-look (probability summation) models (e.g., Nardini et al., 2010), which predict improvements of up to ~25%.
2. An improvement in sensitivity equal to the predictions of the Early Noise model is neither necessary or sufficient proof of an optimal decision strategy, *unless* one assumes (or can evidence) that internal noise is early, independent, and additive, and that the system is linear. Without these assumptions, it is not the case that the greatest possible improvement is $\sqrt{N}$. Nor is it the case, more generally, that the ideal sensitivity in the combined

1 change on this page

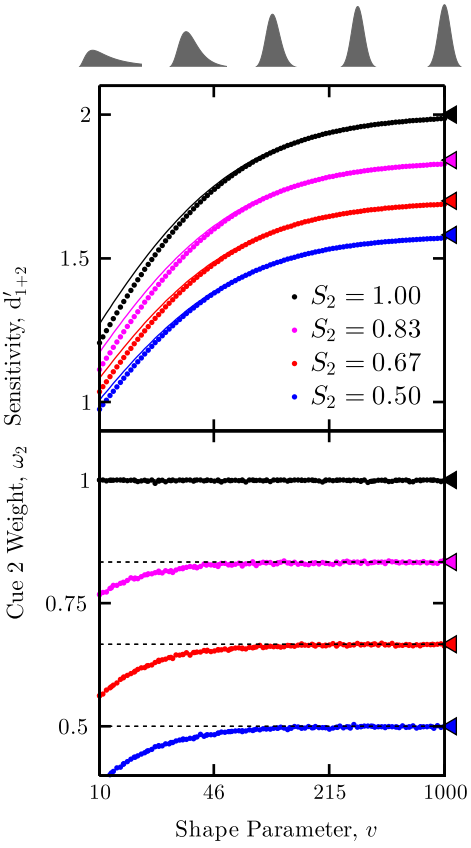


Fig. 16. Means of Monte Carlo simulations, showing how sensitivity (top) and ideal weights (bottom) vary when the skew of the internal noise distribution is varied, and also as the utility of the second cue (different colored curves) is varied. Each data point (100 points per curve, per level of skew) was estimated by mean-averaging over 100 independent simulations, using 2M trials per simulation (80B trials total). The signal magnitude, μ, of Cue$_1$ was fixed at 1.0 (i.e., $d'_1 \approx 1.0$, though, as discussed in Section 2.6, d' is not technically a valid measure of sensitivity when internal noise is not additive and Gaussian). For Cue$_2$, μ varied from 0.5 to 1.0 between curves. The appropriate non-parametric method of sensitivity is the Area Under the (ROC) Curve [AUC]. Nonetheless, estimated d' values in the combined condition were a near miss to the predictions the ideal, additive, Gaussian model (dashed line, top panel; derived using Eq. (2.3.5)). Similarly, the ideal weights (those that maximized d'_{1+2}), approximated the predictions of the ideal, additive, Gaussian model (dashed line, bottom panel), for all but the greatest levels of skew. Furthermore, even when differences in ideal weights appear large, the consequent differences in performance was very slight (i.e., as indicated in the top panel, comparing observed performance, to the predictions of the ideal Gaussian model). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

condition is equal to the Pythagorean sum of the individual sensitivities. The ideal observer would show a smaller increase in sensitivity if the noise is correlated or multiplicative (see below), or a greater increase in sensitivity if the noise is late or superadditive (see Table 1). Without ruling out these other possibilities, one cannot therefore infer whether an observer's decision strategy is more or less efficient, or make any claims as to why one observer's sensitivity differs from another's.

3. A $\sqrt{N}$ improvement in sensitivity does not mean that the observer is using N cues to make their decision. This is the case firstly, because noise may be multiplicative-and-late, or early and correlated across cues. And secondly, because even assuming independent, additive, early noise, there are many possible combinations of weights that predict a given level of performance (i.e., only a minority of which involve giving zero-weight to any single cue). Thus, changes in sensitivity cannot reveal the number of cues an observer has used to make their decision (although, as some authors have noted (Dakin, Mareschal, & Bex, 2005; Manning, Dakin, Tibber, & Pellicano, 2014), sensitivity can be quantified in terms of the 'effective' number of cues that would be required to produce the observed performance, assuming binary – ~~0.0 or 1.0~~ – weights).

3.2. Why multi-cue sensitivity may differ between observers

To summarize, when presented with two cues, why might observer A exhibit lower sensitivity than observer B? If the observers have been equated for their individual-cue sensitivities (e.g., by using a metric of *efficiency*), then two categories of explanation are possible. Firstly, observer A may be using a qualitatively poorer strategy to combine the available information. Secondly, A may be using the same or better strategy to B, but be implementing it less optimally (e.g., suboptimal weights). Reasons for both of these eventualities have been discussed throughout the present manuscript. Thus, a less powerful strategy may be easier to implement, placing fewer demands on memory and attention. While, in terms of implementation, more complex strategies contain numerous parameters, each of which may have a level of estimation error associated with it. Accurate parameter estimates may therefore require a requisite level of skill or practice, and some observers may have *a priori* biases that affect their computation or use.

A third alternative is that the two observers do *not* differ in sensitivity, and that the apparent difference is a statistical artifact. This may happen for either of two reasons. Firstly, when assumptions of either normality or additivity are breached, then d' ceases to be an appropriate measure of sensitivity. This can cause apparent sensitivity to differ spuriously between tests/observers, depending on where they place their criterion. Such differences would be eradicated by using an appropriate, non-parametric measure, such as d^a or AUC. Secondly, as shown in Table 1, the amount of measurement error associated with estimates of sensitive are non-trivial. This can lead to Type I (false difference reported) or Type II (true difference missed) errors, either of which can lead to misleading conclusions as to how observers compare in terms of sensitivity.

4. Beyond measures of sensitivity

In the present work, we have assumed that observers are attempting to minimize response error by combining redundant cues, and have outlined how changes in sensitivity can be used to evaluate how observers integrate sensory information. In doing so, we have only scratched the surface of compound-cue decision making. Experimentally, sensitivity is only one of many possible dependent variables that we wish to measure (see below), and when performing a psychophysical task, observers may be attempting to optimize parameters other than response error (e.g., response bias Landy et al., 1995; Scarfe & Hibbard, 2011 or response time Grice, Canham, & Boroughs, 1984; Hershenson, 1962). Furthermore, in the real world observers also need to consider a range of extraneous factors, such as causation ('do both cues pertain to the same common source?'), prior information, and the relative costs/pay-offs associated with each possible response outcome. A full exposition of all these facets is beyond the scope of the present work, and interested readers are instead encouraged to consult Ref. Trommershauser et al. (2011). However, in this final section we highlight two key ways in which our understanding of compound-cue decision making can be improved by looking beyond measurements of sensitivity alone.

2 changes on this page

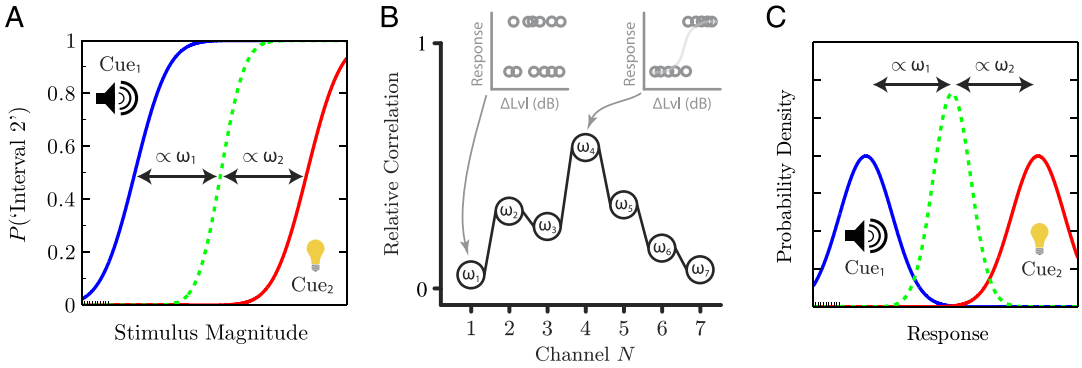


Fig. 17. Two methods of computing relative decision weights. (A) Cue Conflict Paradigm. To the extent that the Point of Subjective Equality (PSE) is shifted laterally towards the Point of Objective Equality (POE) in Cue_1, the observer can be inferred to be relying upon (giving weight to) Cue_1. To the extent that the PSE is shifted towards the POE in Cue_2, the observer is giving weight to Cue_2 (see Section D in the Supplemental Material). (B) Reverse Correlation Paradigm. The relative magnitude of correlation between the trial-by-trial variations in Cue-level and observer-response, indicates the degree to which the observer attends to (weights) that Cue. (C) Continuous Cue Conflict Paradigm. A modified version of (A), in which the dependent variable is continuously distributed, allowing weights to be inferred from a single trial. If the decision strategy varies across trials, a Gaussian mixture model will be observed. If the decision strategy is constant, then a Gaussian distribution will be observed, with a standard deviation determined by the observer's sensitivity (as per the slope of the green dashed curve in A).

4.1. Relative weights

In Section 3.1 we saw that sensitivity alone is a relatively poor indicator of how efficiently an observer is performing, or of what their underlying decision strategy is. As a result, researchers interested in human decision-making are increasingly looking to quantify relative decision weights, ω. These indicate not only how *well* people are performing a given task, but also *how* they are performing it, and can provide additional information with which to disambiguate between models of decision making (see below).

How to measure relative decision weights

In practice, relative decision weights can be measured by introducing a discrepancy between the response predicted by each cue, and recording how observers respond (irrespective of whether the response is correct or incorrect). The discrepancy between cues may be a constant (Cue Conflict paradigm De Gelder & Bertelson, 2003), in which case the effect is to laterally shift the psychometric function leftwards/rightwards in favor of the more weighted cue (Fig. 17(A)). Alternatively, the discrepancy between cues may be introduced randomly, by adding uncorrelated (external) noise to each cue on a trial-by-trial basis (Reverse Correlation paradigm Dai & Micheyl, 2010; Gorea et al., 2014; Lutfi, 19[illegible] Richards & Zhu, 1994). In this case, the relative *correlation* between the trial-by-trial value of each cue and the observer's response is used to index weights (Fig. 17(B)). For example, if a certain cue strongly dictates responses, then the relative correlation (weight) will be high. Conversely, if a cue is largely ignored, then the relative correlation between cue-value and observer-response (weight) will be low. The Reverse Correlation approach can be more easily generalized to more than two cues, and since the expected *mean* disparity between each cues is zero, it may discourage observers from modifying their decision strategy due to one cue being perceived as 'better' (less biased) (Roach, Heron, & McGraw, 2006).

In both the Cue Conflict and Reverse Correlation paradigms described above, relative weights are computed using data from multiple trials. This provides a measure of the *average* reliance placed on each individual cue. However, this does not indicate what weights were used on any specific trial. In cases where relative weights are not constant across trials, these methods may therefore provide misleading results. For example, an observer who alternates, trial-by-trial, between giving full weight to each of two cues ($P_1 = 0.5$; $P_2 = 0.5$), may appear indistinguishable from an observer who always integrates both cues on every trial, but gives equal weight to each ($\omega_1 = 1.0$; $\omega_2 = 1.0$). One way to validate whether weight measures are valid is to reapply them to the original data, and use them to predict an observer's trial-by-trial responses (i.e., since predicted and measured sensitivity would be different—see below). Alternatively, the presence of a nonstationary decision strategy can be evidenced by making the response measure continuous, thereby allowing weights to be estimated within a single trial. For example, if asked to point at the location of a conflicting flash-beep compound, then the relative weight given to the visual and auditory cues can be observed directly from a single response. An alternating observer will produce a bimodal distribution of responses, while the observer who maintains constant weightings will produce a Gaussian distribution of responses (Fig. 17(C)).

Why measure relative weights

Irrespective of precisely how they are measured, relative weight coefficients may be of interest for two main reasons. The first is that, when used in combination with sensitivity measurements, they can help to constrain the number of possible decision models. Thus, while many models may predict observed sensitivity given *some* combination of cues (see Table 1), generally only a small subset of models are consistent with both a particular level of sensitivity, d', and a particular set of relative weights, ω (Fig. 18). Therefore, by independently measuring both d' and ω, the underlying decision model (and thus efficiency) can be inferred empirically, without the need to make a series of debatable assumptions.

Unfortunately, while this approach is elegant in principle, there is a practical difficulty. Namely, the amount of measurement error typically associated with estimates of both d' and ω means that their union may fail to adequately constrain the range of possible models (Fig. 18). Thus, without uncommonly precise measures of d' and ω, it remains impossible to say with certainty precisely which model underlies performance, and therefore whether a given level of sensitivity was more or less close to ideal. Researchers looking to ensure that they have sufficient data to distinguish between rival hypotheses are encouraged to inspect the sampling distributions of d' given in Section B in the Supplemental Material.

Furthermore, the potential for an experiment to be underpowered increases as expected benefit decreases—e.g., either because cues are sub-optimally weighted (Fig. 18(B)), or because the observer is relative insensitive to the second cue (Fig. 18(C); see also Fig. 2). In either case, the difference in predicted sensitivity is further diminished between models. In this light, it is perhaps unsurprising that so many studies have found humans to act consistent with a 'Bayesian (early-noise) ideal observer' (Alais & Burr, 2004;

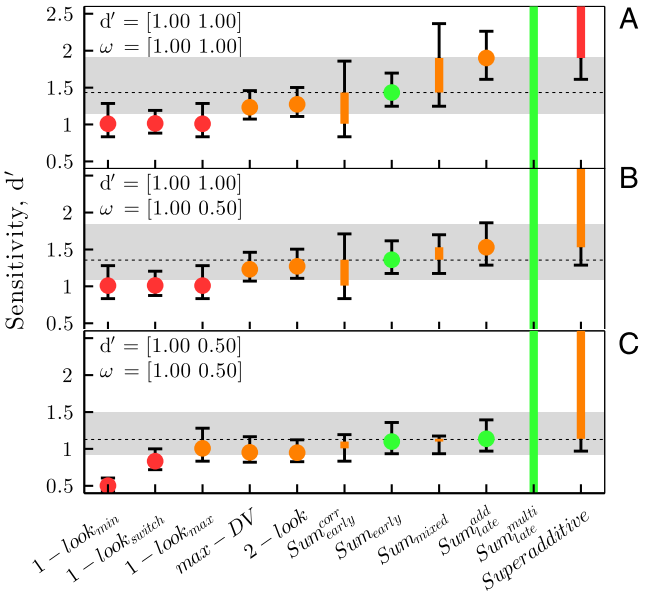


Fig. 18. Observed (horizontal dashed line) and predicted (markers) sensitivity given various cue-combination models (*X* Axis). Panels show three example combinations of single-cue sensitivities and weights. Amber and green markers indicate model predictions that would not be expected to differ significantly from observed sensitivity (i.e., and which therefore cannot be discounted). For the sake of example, observed sensitivity (horizontal dashed line) was computed using the $\text{Sum}_{\text{early}}$ model (Section 2.3). The value shown represents the mean expected d' value, averaged across 20,000 independent simulations. In each simulation, a raw d' value was computed by fitting a psychometric function to 600 simulated trials (see Section D in the Supplemental Material). The shaded grey region shows the 95% Confidence Interval of the mean ($\text{CI}_{95\%}$ = 1.96 × 1 SD of the N = 20,000 sample distribution). To compute the various model predictions (colored markers) analogous simulations were also run for each single cue condition (to derive estimates of d'_1 and d'_2), and for a two-cue conflict-condition (to derive estimates of ω_1 and ω_2). The $\text{CI}_{95\%}$ error bars for the markers therefore incorporate measurement error from each of these three conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Ernst & Banks, 2002; Fetsch et al., 2012; Gori et al., 2008; Knill & Saunders, 2003; Nardini et al., 2010, 2008), as with only two cues it would require an acutely poor strategy, or an uncommonly small degree of measurement error, to prove otherwise.

These practical difficulties can be attenuated by increasing the number of trials used to make each estimate, or by averaging estimates to reduce sampling error. However, caution is required in each case. Long trial sequences can introduce learning effects (Jones, Moore, Shub, & Amitay, 2014), while if averaging across observers there is a danger that qualitative differences in decision strategy (e.g., $\omega_A = \{0\ 1\}$ vs. $\omega_A = \{1\ 0\}$) may alias to produce ~~an unrepresentative mean~~. Alternatively, effect sizes can be increased by increasing the number of cues, in which case the expected differences in sensitivity (between models) will increase independent of measurement error. However, as discussed previously in Section 2.3 there is no guarantee that an observer will necessarily rely on the same qualitatively decision strategy when cues are few versus many.

The second reason why weight measurements may be of interest is because they often constitute a more direct measurement of what we have been so far trying to infer indirectly through estimates of sensitivity. Thus, it is often assumed (perhaps not always correctly) that the limiting factor in decision-making, and the only component of the decision process that observers have any control over, is the relative weight given to each cue. That being the case, the question of "how efficient is an observer's decision-process?" reduces to "how appropriately does the observer weight each cue?" However, when quantifying weight efficiency, one soon encounters the same difficulty as with sensitivity (Section 3). Namely, that knowledge of the underlying model is necessary in order to establish what the appropriate (ideal) weights are in the first place. Thus, for example, it is not the case that "optimally, weights are chosen to be proportional to the reliability of a given signal" (Ernst, 2006), unless one assumes that the internal noise is early, independent, and additive, and that summation is linear. To the extent that the noise is late, both cues should always be given equal weight. To the extent that the noise is correlated or multiplicative,[11] the ideal weights will vary in more complex ways. As discussed, some of these assumptions are more likely to be valid than others. Notably, deviations from normality will not tend to substantially affect the optimal weightings, though may affect the taking of measurements on a practical level (Richards & Zhu, 1994).

In principle then, neither measurements of d' or ω alone are sufficient to estimate an observer's efficiency. Both are required to constrain the underlying model of decision making, and neither are intelligible without knowledge of the underlying model. There are, however, two exceptions. The first is when all cues are equally useful. In that case almost all decision strategies predict equal weights, so there is no need to commit to any specific model. The second occurs when external noise is introduced to the stimuli by the experimenter. Such noise is generally assumed to be exclusively early in effect,[12] can be specified as being independent and Gaussian, and when it is great enough in magnitude will swamp the decision-process, making any internal noise negligible. Given these assumptions, the ideal weights are guaranteed to be those predicted by the early-noise model (i.e., directly proportional to reliability, Eq. (2.3.4)), and so the efficiency of the weights can be meaningfully considered independent of observed sensitivity.

4.2. Response times

Throughout the present paper, we have assumed that the observer's goal is always to minimize response error. However, observers may also wish to optimize other properties of the decision making process, such as *response time*. A reduction in response time may thus constitute an entirely separate reason to attend to a second cue, and a second cue may thereby confer benefit even when accuracy/sensitivity in the single-cue conditions is near ceiling.

Many authors have used response time data to make inferences regarding compound-signal decision making (Grice et al., 1984; Hershenson, 1962; Hugenschmidt, Hayasaka, Peiffer, & Laurienti, 2010; Otto & Mamassian, 2012; Schröger & Widmann, 1998), and a full exposition of this literature is beyond the scope of the present work. However, it is worth noting that many commonalities exist between the approaches used to study response times, and the models of sensitivity in Section 2.

In brief, classic response-time theory posits that noisy sensory data is accumulated over time until a criterion is reached (a 'drift-diffusion' process Ratcliff & McKoon, 2008), at which point a decision is made. Given this framework, authors typically advance three alternative hypotheses, illustrated in Fig. 19, for how observers respond to two cues:

[11] Or, more generally, to the extent that the system is non-linear through superadditivity, or in some other way.

[12] This assumption is not necessarily correct, however. For example, it is possible to imagine a situation in which adding random jitter to the sensory inputs causes observers to also vary their ability to integrate information or maintain a stable criterion, for example due to simple confusion or perceptual load. Evidence against strategy changes can be found in Pelli (1990). Evidence for strategy changes (for some, but not all types of noise) can be found in works by Allard and Cavanagh (2011, 2012) and Allard and Faubert (2013).

2 changes on this page

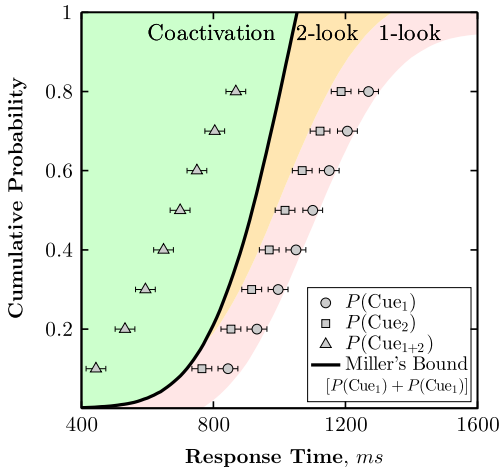


Fig. 19. Example response-time analysis. Markers show simulated single-cue performance (squares, circles), and two-cue performance (triangles). Shaded regions show predicted two-cue performance for three models of decision making. The solid black line represents Miller's Bound (or: 'the Race Model Bound'), which is defined as the sum of the two single-cue cumulative probability density functions (for details, see Refs. Miller, 1982 and Ulrich et al., 2007). Miller's Bound constitutes a nonparametric model of the greatest improvement over single-cue performance that could be achieved if each cue were processed independently (i.e., were combined through probability summation alone). In the data shown here, the simulated observer's response times in the compound condition (triangles) exceeded Miller's Bound, so are consistent with coactivation.

1. The observer responds based upon a single cue only. In this case, response times will be no quicker than response times in the faster single cue condition. This is equivalent to the 1-look model for response-accuracy presented in Section 2.1.
2. The observer responds based upon whichever random walk reaches threshold first. In this case, response times will, by chance, tend to be faster than in either single-cue condition, but should never exceed the sum of both (Miller's Bound; Fig. 19, black line). This represents probability summation of response times, and is equivalent to the 2-look accuracy model presented in Section 2.2.
3. Evidence from both cues 'coactivate' some central decision-making process (i.e., evidence from both is accumulated together within a single random walk). In this case, response times in the two-cue condition may reduce by more than the sum of the single-cue conditions. (Typically, no parametric predictions are made, though are possible if assumptions are made regarding the underlying error distribution and sampling rate.) In its non-parametric form, this hypothesis is a superset of all the models presented in Section 2.3 onwards.

In short, decision-models of response times have typically attempted to address the same basic questions as decision-models of sensitivity, and many studies of response times neatly parallel those of sensitivity (see, for example, Ref. Nardini et al. (2010) for a study of sensitivity that explicitly attempts to disambiguate between the three hypotheses shown graphically in Fig. 19). In doing so, studies of response times also suffer from the same theoretical caveats (e.g., failure to exceed Miller's Bound does not rule out more powerful 'coactivation' models, unless one assumes independent noise and unlimited processing capacity), as well as many of the same practical limitations (e.g., relatively small differences in effect size between models). As a dependent measure, response times are also complicated by their relative volatility. Thus, researchers often find it necessary to exclude as outliers empirical estimates below and/or above an arbitrary threshold. Such trimming can in turn introduce non-trivial artifacts into the data, which, if not corrected for, can lead to fallacious conclusions (Rach, Diederich, Steenken, & Colonius, 2010). For practical reasons, many authors therefore to prefer measures of sensitivity over response times.

Recently, however, there has been interest in response time data, for two, related reasons. Firstly, because many sensory decisions intuitively represent a *trade-off* between speed and accuracy. This is most obviously the case in situations where the stimulus duration is under the participant's control, but trade-offs may even occur with brief stimulus presentations, given that evidence continues to be accumulated even after the physical input is removed (e.g., as evidenced by physiological recordings Schall, 2001, and by backward masking Breitmeyer & Ogmen, 2000). Thus, an observer may choose to spend less time gathering evidence at the cost of decreased accuracy, or may deliberate longer to improve accuracy. A number of models have been proposed to account for speed/accuracy trade-off (see Refs. Drugowitsch, DeAngelis, Klier, Angelaki, & Pouget, 2014; Palmer, Huk, & Shadlen, 2005; Ratcliff & Smith, 2004), and these can, for example, reveal highly efficient decision-process even in situations where response accuracy in the multi-cue condition did not improve significantly (Drugowitsch et al., 2014).

Secondly, response times may be of interest because they can provide additional information to complement sensitivity metrics. Thus, a recurring conclusion of the present work is that it can be remarkably difficult to distinguish between competing models of decision making, based on accuracy alone (e.g., see final column of Table 1). Often, however, differences in accuracy are comorbid with differences in response time. For example, age-related sensory decline is characterized by both slower and less accurate responses. Combining both response times and accuracy into a single unified measure – such as in the non-parametric 'integration coefficient' proposed by Altieri, Townsend, and Wenger (2014) and Townsend and Altieri (2012) – might therefore provide a more sensitive test of cue-combination ability. Currently, however, the data are lacking to conclusively validate such a compound measure.

Acknowledgments

The author thanks Joshua Solomon for comments on an earlier draft. This work was supported by the Special Trustees of Moorfields Eye Hospital (#ST 12 03 C), and by the NIHR Biomedical Research Centre at (both) Moorfields Eye Hospital and the UCL Institute of Ophthalmology (#Moorfields BRC2_010).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmp.2016.04.006.

References

Abbott, L. F. (1999). Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Research Bulletin*, *50*, 303–304.
Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.
Alexander, J. M., & Lutfi, R. A. (2004). Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *Journal of the Acoustical Society of America*, *116*, 2234–2247.
Alexander, J. M., & Lutfi, R. A. (2008). Sample discrimination of frequency by hearing-impaired and normal-hearing listeners. *Journal of the Acoustical Society of America*, *123*, 241–253.
Allard, R., & Cavanagh, P. (2011). Crowding in a detection task: External noise triggers change in processing strategy. *Vision Research*, *51*, 408–416.
Allard, R., & Cavanagh, P. (2012). Different processing strategies underlie voluntary averaging in low and high noise. *Journal of Vision*, *12*, 1–12.
Allard, R., & Faubert, J. (2013). Zero-dimensional noise is not suitable for characterizing processing properties of detection mechanisms. *Journal of Vision*, *13*, 25.

Altieri, N., Townsend, J. T., & Wenger, M. J. (2014). A measure for assessing the effects of audiovisual speech integration. *Behavior Research Methods*, *46*, 406–415.

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology*, *19*, 452–458.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*, 358–366.

Ban, H., Preston, T. J., Meeson, A., & Welchman, A. E. (2012). The integration of motion and disparity cues to depth in dorsal visual cortex. *Nature Neuroscience*, *15*, 636–643.

Bates, S. L., & Wolbers, T. (2014). How cognitive ageing affects multisensory integration of navigational cues. *Neurobiology of Aging*, *35*, 2761–2769.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*, 1391–1397.

Berg, B. G. (1990). Observer efficiency and weights in a multiple observation task. *Journal of the Acoustical Society of America*, *88*, 149–158.

Blake, R., & Fox, R. (1973). The psychophysical inquiry into binocular summation. *Perception & Psychophysics*, *14*, 161–185.

Blake, R., Sloane, M., & Fox, R. (1981). Further developments in binocular summation. *Perception & Psychophysics*, *30*, 266–276.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, *62*, 1572–1595.

Cowan, N. (2010). The magical mystery four how is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*, 51–57.

Dai, H., & Micheyl, C. (2010). Psychophysical reverse correlation with multiple response alternatives. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 976–993.

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A*, *18*, 1016–1026.

Dakin, S. C., Mareschal, I., & Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Research*, *45*, 3027–3049.

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*, 460–467.

Dekker, T. M., et al. (2015). Late development of cue integration is linked to sensory fusion in cortex. *Current Biology*, *25*, 2856–2861.

Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, *4*, 826–831.

DeVries, S. H. (2000). Bipolar cells use kainate and AMPA receptors to filter visual information into separate channels. *Neuron*, *28*, 847–856.

Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L., & Tank, D. W. (2007). Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron*, *56*, 43–57.

Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *Elife*, *3*, e03005.

Ecker, A. S., et al. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron*, *82*, 235–248.

Ernst, M. O. (2006). In G. Knoblich (Ed.), *Hum. body percept. from insid. out*. Oxford University Press.

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

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*, 429–442.

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*, 146–154.

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *Journal of Neuroscience*, *29*, 15601–15612.

Field, G. D., & Rieke, F. (2002). Mechanisms regulating variability of the single photon responses of mammalian rod photoreceptors. *Neuron*, *35*, 733–747.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of Vision*, *2*, 190–203.

Garcia, S., Petrini, K., Rubin, G. S., da Cruz, L., & Nardini, M. (2015). Visual and non-visual navigation in blind patients with a retinal prosthesis. *PLoS One*, *10*, e0134369.

Geldard, F. A., & Sherrick, C. E. (1965). Multiple cutaneous stimulation: The discrimination of vibratory patterns. *Journal of the Acoustical Society of America*, *37*, 797–801.

Gold, J. M., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, *28*, 167–207.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision*, *14*, 22.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, *18*, 694–698.

Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, *17*, 858–865.

Green, D.M., & Swets, J.A. (1974) (pp. 1–455). Krieger.

Grice, G. R., Canham, L., & Boroughs, J. M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics*, *35*, 451–463.

Grimmett, G. R., & Stirzaker, D. R. (2003). *Probability and random processes*. Oxford Univ. Press.

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, *11*, 1201–1210.

Gureckis, T., & Love, B. (2009). Learning in noise: Dynamic decision-making in a variable environment. *Journal of Mathematical Psychology*, *53*, 180–193.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and ageing: A review and a new view. *Psychology of Learning and Motivation*, *22*, 193–225.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, *63*, 289–293.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, *298*, 1627–1630. (80-.).

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*, 967–992.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: space, time and superadditivity. *Current Biology*, *15*, R762–R764.

Hugenschmidt, C. E., Hayasaka, S., Peiffer, A. M., & Laurienti, P. J. (2010). Applying capacity analyses to psychophysical evaluation of multisensory interactions. *Information Fusion*, *11*, 12–20.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*, 345–350.

Jaramillo, F., & Wiesenfeld, K. (1998). Mechanoelectrical transduction assisted by Brownian motion: a role for noise in the auditory system. *Nature Neuroscience*, *1*, 384–388.

Javel, E., & Viemeister, N. F. (2000). Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application to intensity discrimination. *Journal of the Acoustical Society of America*, *107*, 908–921.

Jesteadt, W., Nizami, L., & Schairer, K. S. (2003). A measure of internal noise based on sample discrimination. *Journal of the Acoustical Society of America*, *114*, 2147–2157.

Jones, P. R., Moore, D. R., & Amitay, S. (2015). Development of auditory selective attention: Why children struggle to hear in noisy environments. *Developmental Psychology*, *51*, 353–369.

~~Jones, P.R., Moore, D.R., Shub, D., & Amitay, S. Learning to detect a tone in unpredictable noise.~~

Jones, P. R., Moore, D. R., Shub, D. E., & Amitay, S. (2014). Learning to detect a tone in unpredictable noise. *Journal of the Acoustical Society of America*, *135*, EL128–EL133.

Jones, P. R., Shub, D. E., Moore, D. R., & Amitay, S. (2013). Reduction of internal noise in auditory perceptual learning. *Journal of the Acoustical Society of America*, *133*, 970–981.

Jones, P. R., Moore, D. R., Shub, D. E., & Amitay, S. (2015). The role of response bias in perceptual learning. *Journal of experimental psychology. Learning, memory, and Cognition*, *41*, 1456–1470.

Juni, M. Z., Gureckis, T. M., & Maloney, L. T. (2012). Effective integration of serially presented stochastic cues. *Journal of Vision*, *12*, 1–16.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*, 712–719.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*, 2539–2558.

Kubovy, M., & Healy, A. F. (1977). The decision rule in probabilistic categorization: What it is and how it is learned. *Journal of Experimental Psychology General*, *106*, 427–446.

Landry, S. P., Guillemot, J.-P., & Champoux, F. (2013). Temporary deafness can impair multisensory integration a study of cochlear-implant users. *Psychological Science*, 0956797612471142.

Landy, M. S., Banks, M. S., & Knill, D. C. (2011). In J. Trommershauser, K. Kording, & M. S. Landy (Eds.), *Sensory cue integration*. Oxford University Press, http://dx.doi.org/10.1093/acprof:oso/9780195387247.003.0001.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389–412.

Langhans, A., & Kohlrausch, A. (1992). Differences in auditory performance between monaural and diotic conditions. I: Masked thresholds in frozen noise. *Journal of the Acoustical Society of America*, *91*, 3456–3470.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging*, *27*, 1155–1163.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology General*, *133*, 339–354.

Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision 'template' for position discrimination. *Nature Neuroscience*, *7*, 178–183.

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four-to eight-year-old children. *Neuropsychologia*, *36*, 273–293.

Lutfi, R. A. (1995). Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks. *Journal of the Acoustical Society of America*, *97*, 1333–1334.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438.

Machilsen, B., & Wagemans, J. (2011). Integration of contour and surface information in shape detection. *Vision Research*, *51*, 179–186.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and nonparametric indexes. *Psychological Bulletin*, *107*, 401–413.

1 change on this page

Macmillan, N. A., & Creelman, C. D. (1996). Triangles in ROC space: History and theory of "nonparametric" measures of sensitivity and response bias. *Psychonomic Bulletin & Review*, *3*, 164–170.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide*. (pp. 3–184). Lawrence Erlbaum Associates.

Macpherson, E. A., & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *Journal of the Acoustical Society of America*, *111*, 2219–2236.

Manning, C., Dakin, S. C., Tibber, M. S., & Pellicano, E. (2014). Averaging, not internal noise, limits the development of coherent motion processing. *Developmental Cognitive Neuroscience*, *10*, 44–56.

McGill, W. J., & Goldberg, J. P. (1968). A study of the near-miss involving Weber's law and pure-tone intensity discrimination. *Attention, Perception, & Psychophysics*, *4*, 105–109.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279.

Morgan, M. L., DeAngelis, G. C., & Angelaki, D. E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, *59*, 662–673.

Moro, S. S., Harris, L. R., & Steeves, J. K. E. (2014). Optimal audiovisual integration in people with one eye. *Multisensory Research*, *27*, 173–188.

Nadarajah, S., & Kotz, S. (2008). Exact distribution of the max/min of two gaussian random variables. *IEEE Transactions on Very Large Scale Integration Systems*, *16*, 210–212.

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 17041–17046.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, *18*, 689–693.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*, 2451–2468.

Ossmy, O., et al. (2013). The timescale of perceptual evidence integration can be adapted to the environment. *Current Biology*, *23*, 981–986.

Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology*, *22*, 1391–1396.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*, 1.

Pelli, D. G. (1990). In C. Blakemore (Ed.), *Vision: coding and efficiency* (pp. 3–24). Cambridge University Press.

Perrault, T. J., Jr., Vaughan, J. W., Stein, B. E., & Wallace, M. T. (2005). Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *Journal of Neurophysiology*, *93*, 2575–2586.

Persike, M., & Meinhardt, G. (2015). Cue combination anisotropies in contour integration: The role of lower spatial frequencies. *Journal of Vision*, *15*, 17.

Petrini, K., Remark, A., Smith, L., & Nardini, M. (2014). When vision is not an option: children's integration of auditory and haptic information is suboptimal. *Developmental Science*, *17*, 376–387.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, *26*, 381–410.

Pouget, A., Deneve, S., & Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, *3*, 741–747.

Putzar, L., Gondan, M., & Röder, B. (2012). Basic multisensory functions can be acquired after congenital visual pattern deprivation in humans. *Developmental Neuropsychology*, *37*, 697–711.

Rach, S., Diederich, A., Steenken, R., & Colonius, H. (2010). The race model inequality for censored reaction time distributions. *Attention, Perception, & Psychophysics*, *72*, 839–847.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.

Redmond, T., Zlatkova, M. B., Vassilev, A., Garway-Heath, D. F., & Anderson, R. S. (2013). Changes in Ricco's Area with background luminance in the S-Cone pathway. *Optometry and Vision Science*, *90*, 66–74.

Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, *139*, 5–21.

Richards, V. M., & Zhu, S. (1994). Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. *Journal of the Acoustical Society of America*, *95*, 423–434.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 2159–2168.

Saarela, T. P., & Landy, M. S. (2012). Combination of texture and color cues in visual segmentation. *Vision Research*, *58*, 59–67.

Sahani, M., & Whiteley, L. (2011). In J. Trommershauser, K. Kording, & M. S. Landy (Eds.), ~~*Sens. cue integr.*~~, (pp. 82–100). Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195387247.003.0001.

Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience*, *2*, 539–550.

Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision*, *11*, 12.

Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, *2*, 33–42.

Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*, 755–759.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, *18*, 3870–3896.

Skaliora, I., Doubell, T. P., Holmes, N. P., Nodal, F. R., & King, A. J. (2004). Functional topography of converging visual and auditory inputs to neurons in the rat superior colliculus. *Journal of Neurophysiology*, *92*, 2933–2946.

Soderquist, D. R., & Lindsey, J. W. (1971). Physiological noise as a low-frequency masker: The cardiac cycle. *Journal of the Acoustical Society of America*, *50*, 143.

Solomon, J. A. (2007). Intrinsic uncertainty explains second responses. *Spatial Vision*, *20*, 45–60.

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, *10*, 19.

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, *11*, 13.

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *Journal of Neuroscience*, *25*, 6499–6508.

Stein, B., Meredith, M., Huneycutt, W., & McDade, L. (1989). Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience*, *1*, 12–24.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–266.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage*, *44*, 1210–1223.

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention ensemble perception of a crowd's gaze. *Psychological Science*, *25*, 1903–1913.

Swets, J. A. (1959). Multiple observations of signals in noise. *Journal of the Acoustical Society of America*, *31*, 514–521.

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science*, *182*, 990–1000.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychological Bulletin*, *99*, 100–117.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*, 301–340.

Tanner, J. W. P., & Birdsall, T. G. (1958). Definitions of d' and n as psychophysical measures. *Journal of the Acoustical Society of America*, *30*, 922–928.

Teich, M. C., & Khanna, S. M. (1985). Pulse-number distribution for the neural spike train in the cat's auditory nerve. *Journal of the Acoustical Society of America*, *77*, 1110–1128.

Tickle, H., Speekenbrink, M., Tsetsos, K., Michael, E., & Summerfield, C. (2016). Near-optimal integration of magnitude in the human parietal cortex. *Journal of Cognitive Neuroscience*, *EPub*, 1–15.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, *41*, 414–419.

Townsend, J. T., & Altieri, N. (2012). An accuracy–response time capacity assessment function that measures performance against standard parallel predictions. *Psychological Review*, *119*, 500–516.

Treisman, M. (1998). Combining information: Probability summation and probability averaging in detection and discrimination. *Psychological Methods*, *3*, 252–265.

Trommershauser, J., Kording, K., & Landy, M. S. (2011). *Sensory cue integration*. Oxford University Press.

Tyler, C. W., & Chen, C.-C. (2000). Signal detection theory in the 2AFC paradigm: attention, channel uncertainty and probability summation. *Vision Research*, *40*, 3121–3144.

Ulrich, R., Miller, J., & Schröter, H. (2007). Testing the race model inequality: an algorithm and computer programs. *Behavior Research Methods*, *39*, 291–302.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Current Biology*, *12*, 834–837.

Watson, C. S., Rilling, M. E., & Bourbon, W. T. (1964). Receiver-operating characteristics determined by a mechanical analog to the rating scale. *Journal of the Acoustical Society of America*, *36*, 283–288.

Weisswange, T. H., Rothkopf, C. A., Rodemann, T., & Triesch, J. (2011). Bayesian cue integration as a developmental outcome of reward mediated learning. *PLoS One*, *6*, e21575.

Werner, S., & Noppeney, U. (2010). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex*, *20*, 1829–1842.

Wickens, T. D. (2002). *Elementary signal detection theory*. (pp. 3–110). Oxford University Press.

Winter, I. M., & Palmer, A. R. (1991). Intensity coding in low-frequency auditory-nerve fibers of the guinea pig. *Journal of the Acoustical Society of America*, *90*, 1958–1967.

Wright, B. A., & Fitzgerald, M. B. (2001). Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 12307–12312.

Young, E. D., & Barta, P. E. (1986). Rate responses of auditory nerve fibers to tones in noise near masked threshold. *Journal of the Acoustical Society of America*, *79*, 426–442.

Yuille, A. L., & Buelthoff, H. H. (1996). In D. Knill, & W. Richards (Eds.), *Bayesian approaches to percept*. Cambridge University Press.

Zacharias, G. L., & Young, L. R. (1981). Influence of combined visual and vestibular cues on human perception and control of horizontal rotation. *Experimental Brain Research*, *41*, 159–171.