

TITLE

ORIGINAL: ~~A tutorial on cue combination theory: Using psychophysical data to evaluate how observers integrate sensory information.~~

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

RUNNING TITLE

Tutorial on cue combination

AUTHORS

Pete R. Jones¹

¹*Institute of Ophthalmology, University College London (UCL), UK.*

Correspondence: Pete R. Jones; Institute of Ophthalmology, University College London, 11-43 Bath Street, London EC1V 9EL; p.r.jones@ucl.ac.uk

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.

KEY WORDS

Cue Combination; Multisensory Integration; Weighted Linear Summation; Signal Detection Theory; Internal Noise

1 Consider a simple sensory judgment, such as ‘where was the source of a sound located’? When
2 attempting to understand how such a decision is made, the sensory input can be thought of as
3 containing multiple sources of information (‘cues’). In general, each cue is a function of the sensory
4 input, which conveys information about a particular physical attribute¹. However, exactly how
5 cues are conceptualized varies between scientific disciplines. In biochemistry, the output of each
6 ionotropic receptor may be considered a distinct cue². In electrophysiology, a cue is generally the
7 firing-rate of a neuron³, or of a given population of neurons⁴⁻⁷. In the behavioral sciences, which
8 the present paper concerns, cues are typically defined with respect to the stimulus. Thus,
9 interaural differences in intensity and phase may be thought of as separate cues in a sound-
10 localization task⁸. Similarly, texture and disparity may be thought of as separate cues when
11 judging visual depth⁹. Alternatively, cues may be defined with respect to time; for example, each
12 interval in a two-alternative forced choice [2AFC]¹⁰, or each sample in a sequential-observation¹¹⁻
13 ¹³ task. Finally, in some cases, cues may be defined with respect to the observer themselves. Thus,
14 each eye^{14,15}, ear¹⁶, area of skin¹⁷, or sensory modality¹⁸ may be thought of as yielding a separate
15 cue.

16 Irrespective of how exactly the various cues are defined, a number of interesting questions arise:
17 Can observers exploit these multiple sources of information^{19,20}? Do they do so in an optimal
18 manner^{21,22}? Do they continue to do so when the statistics of the task vary²³⁻²⁶? At what age does
19 this ability to combine cues develop^{27,28}? Is it preserved in old age^{29,30}? Is it present in clinical
20 populations where some information channels are degraded^{31,32}, or have been previously deprived
21 of input³³⁻³⁵?

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

31 If the underlying model of decision making is known, it can also be used as a yardstick to assess
32 how effective observers are at exploiting the information available to them. Thus, by defining some
33 putative ‘ideal’ level of performance, it becomes possible to compare observed performance to the
34 ideal, and thereby to state whether the observer is behaving *optimally*. Furthermore, by measuring
35 observed performance relative to the ideal, a measure of *efficiency* can be computed (defined
36 formally in Eq 1.1.5). This allows cue-combination ability to be compared across observers, even
37 when each individual’s sensitivity is expected to vary³⁶. *Ideal observer* analyses are therefore of
38 substantial practical and theoretical utility, and are used extensively throughout studies of sensory
39 cue-combination^{37,38} (for further discussion, see Ref~[³⁹]).

40 However, what has not always been made clear is the diversity of plausible ideal-observer models.
41 Thus, depending on the specific model used, what constitutes ‘ideal’ performance may differ
42 between papers, and human performance in one study can exceed the predictions of an ideal
43 observer in another (e.g., contrast the factor of N improvement predicted by Saarela and Landy⁴⁰
44 with the factor of \sqrt{N} improvement predicted by Knill and colleagues²³). A closely related issue is
45 that readers are not always fully aware of the key assumptions that are often required in order to
46 compute ‘ideal’ performance. As shall be discussed, these assumptions are rarely strictly correct,
47 and depending on exactly what assumptions one makes, the inferences regarding underlying
48 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 §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^{41–43} [SDT] metrics. Other measures can also be used to study perceptual sensitivity, such as the slope parameter of the psychometric function²² or the variance of a continuously distributed response²⁸. However, SDT metrics are of particular interest due to their prevalence in the literature^{26,40,44–48,i}, and the fact that 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 §1, we introduce briefly the relevant background theory. In §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 §3, we summarize the information presented and develop overall comparisons and corollaries. In §4 we highlight the limits of what can be inferred from sensitivity alone, and discuss other approaches to studying cue-combination.

ⁱ 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. Traditional psychophysical task therefore ask observers to make a discrete (e.g., yes/no) response. If Method of Constant Stimuli is used, then even binary responses can be used to recover a continuous psychometric function, from which a measure of sensory precision can be derived. However, such experimental designs are time consuming and not always practicable – for example, when performing multiple tests, or when working with children or clinical populations.

§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 judgments⁴⁹. Here we detail its key tenets. For more comprehensive expositions, see Refs~[41–43].

In SDT, an incoming sensory signal is theorized to produce an *internal response*, typically represented as a single scalar variable, x (**Fig 1A**). 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/m²). Notably though, various neural⁵⁰, physiological⁵¹, 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 1B**). 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:

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

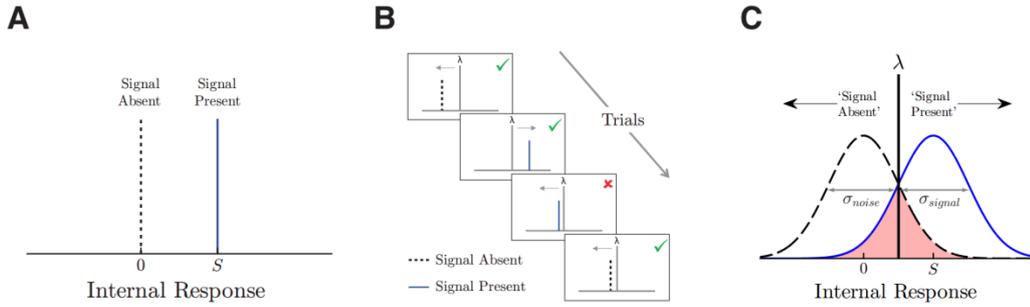
In **Eq 1.1.1a** the decision variable, DV, upon which the behavioral response is based (‘Response’), is simply the internal response to a single cue, x . In more complex tasks, however, the DV will not be determined by a single internal response. For example, in a two alternative forced-choice task, the DV is generally considered to be the *difference* between the internal responses to each interval:

$$Response = \begin{cases} \text{‘Signal in Interval 2’} & \text{if } DV > \lambda \\ \text{‘Signal in Interval 1’} & \text{otherwise} \end{cases} \quad \text{where } DV = x_{Int2} - x_{Int1}. \quad (1.1.1b)$$

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

$$Response = \begin{cases} \text{‘Signal Present’} & \text{if } DV > \lambda \\ \text{‘Signal Absent’} & \text{otherwise} \end{cases} \quad \text{where } DV = x_1 + x_2. \quad (1.1.1c)$$

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 1C**). To the extent that the ‘Noise’ and ‘Signal’ internal-response distributions overlap (the shaded area in **Fig 1C**), the observer is said to be less *sensitive* to differences between the two.



104 **Fig 1.** Signal Detection Theory schema for a simple, yes/no detection task. On each trial, either a signal is
 105 presented, or nothing is presented. (A) The mean internal response is either 0 (Nothing), or S (Signal; where S is
 106 proportional to signal magnitude). (B) However, internal noise means that, on any given trial, the internal
 107 response may be slightly higher or slightly lower. Any given internal responses must therefore be compared to a
 108 criterion, λ , in order to determine the appropriate behavioral action. (C) When averaging across many trials, a
 109 distribution of internal responses is observed for each condition (in this image the noise is additive and normally
 110 distributed). To the extent that the two distributions overlap (red shaded region), the observer cannot distinguish
 111 the two conditions perfectly. This overlap is what is indexed by the sensitivity metric, d' .
 112

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

$$sensitivity = \frac{\mu_{signal} - \mu_{noise}}{\sqrt{\frac{1}{2} (\sigma_{signal}^2 + \sigma_{noise}^2)}}. \quad (1.1.2)$$

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

$$sensitivity = \frac{\mu_{signal} - 0}{\sqrt{\frac{1}{2} (\sigma_{signal}^2 + \sigma_{noise}^2)}} = \frac{\mu_{signal}}{\sqrt{\frac{1}{2} (\sigma^2 + \sigma^2)}} = \frac{\mu_{signal}/\sigma}{\sqrt{\frac{1}{2} ((\sigma/\sigma)^2 + (\sigma/\sigma)^2)}} = \frac{d'}{\sqrt{1}} = d'. \quad (1.1.3)$$

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

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

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

$$\Phi^{-1}(p) = \sqrt{2} \operatorname{erf}^{-1}(2p - 1). \quad (1.1.4b)$$

135 For the derivation of **Eq 1.1.4**, see Refs~[36–38].

136 Finally, it is often useful to measure observed sensitivity, d'_{obs} relative to some putative ideal, d'_{ideal} .
 137 This allows easy comparison across conditions where d'_{ideal} is expected to vary – for example,
 138 when comparing compound-cue sensitivity across observers who are differentially sensitive to the
 139 constituent single-cues. Such a measure is known as *efficiency*, η , and is defined by Tanner and
 140 Birdsall (1958)⁵² as:

$$\eta = [d'_{obs}/d'_{ideal}]^2. \quad (1.1.5)$$

141 Thus, $\eta=1$ indicates ideal sensitivity, and lower values indicate suboptimal performance.

142 §1.2 Combining random variables

143 §1.1 expounded how, according to SDT, binary decisions are made by comparing a scalar Decision
 144 Variable, DV, to a criterion, λ . If there is just a single cue, then the DV is fully determined by a single
 145 internal response variable, x ($DV = x$). However, as discussed, many models of decision-making
 146 imply multiple separate cues, each with its own associated source of noise. In that case, each cue
 147 can be considered a separate random variable, and in most models of decision making the DV is
 148 some combination of these variables (though cf. §2.1, §2.2). It is therefore important to
 149 understand the expected properties of a random variable that is the combination of N random
 150 variables.

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

$$DV = \omega_1 x_1 + \omega_2 x_2. \quad (1.2.1)$$

155 When $\omega_1 = 1$ and $\omega_2 = 1$ the two cues are summed completely (Total Summation). This may not
 156 always be desirable, however. For example, when detecting a change of heading, our sense of
 157 vision often provides more accurate information than our internal sense of balance²⁸. If the
 158 goal is to minimize response error, then the less reliable balance cue should be given less
 159 relative weight ($\omega < 1$; Partial Summation).

160 The exact way to weight cues optimally shall be discussed in §2. However, irrespective of the
 161 specific weights employed, the mean of the weighted-sum of two random variables is the
 162 weighted-sum of the individual means:

$$\mu_{DV} = \omega_1 \mu_1 + \omega_2 \mu_2. \quad (1.2.2)$$

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

$$\sigma_{DV} = \sqrt{\omega_1^2 \sigma_1^2 + \omega_2^2 \sigma_2^2}. \quad (1.2.3)$$

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

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

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

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

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

173 any is violated.

§2 Taxonomy

In this section we describe fourteen algorithms that a sensory system might plausibly employ in order to make a judgment based on multiple sensory cues. Practical examples of each model are also given in [Appendix A](#), 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 the underlying decision 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)⁴², Wickens (2002; Ch. 10)⁴³, and Triesman (1998)⁵³, 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, Kording and Landy (2011)³⁸.

For simplicity, we will constrain ourselves to situations where there are only two sources of information (Cue₁ and Cue₂), 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 I](#) and [Figure 2](#) provide overviews of the various decision models, each of which is defined in detail from [§2.1](#) onwards.

[Table I](#) 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 ‘Sum_{early}’ model (Row 7; [§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 (‘Sum_{add}’, and ‘Superadditivity’).

Note also that the final column of [Table I](#) includes 95% confidence intervals, derived numerically for cases in which 100 empirical trialsⁱⁱ are used to estimate d' (Grey shaded regions; for derivation see [Appendix B](#)). 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 [Appendix B](#) for derivations of confidence intervals).

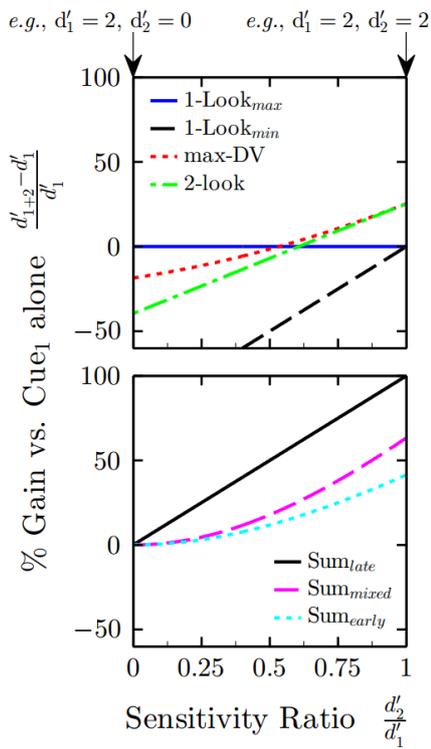
An additional perspective on the various decision models is given in [Figure 2](#), which shows how the expected benefit of a second cue increases as the difference in sensitivity to each individual cues decreases (i.e., as the second cue becomes proportionately more useful).

ⁱⁱ 100 trials represents a an arbitrary but realistic number

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 (see body text for details); (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), a horizontal bar show the range of possible outputs. Gray 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-look _{min}	$S_{min} + N_{min}$	d' $\min(d'_1, d'_2)$	1
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
1-look _{max}	$S_{max} + N_{max}$	d' $\max(d'_1, d'_2)$	1
Max-DV	$\max(S_1 + N_1, S_2 + N_2)$	d_a $(\mu_{1+2} - 0.50) / \sqrt{(\sigma_{1+2}^2 + 0.83)^2 / 2}$ $\mu_{1+2} = d'_1 \Phi\left(\frac{d'_1 - d'_2}{\sqrt{2}}\right) + d'_2 \Phi\left(\frac{d'_1 - d'_2}{\sqrt{2}}\right)$ $+ \sqrt{2} \Phi\left(\frac{d'_1 - d'_2}{\sqrt{2}}\right)$ $\sigma_{1+2}^2 = \left[(1 + d_1^2) \Phi\left(\frac{d_1 - d_2}{\sqrt{2}}\right) + (1 + d_2^2) \Phi\left(\frac{d_2 - d_1}{\sqrt{2}}\right) \right. \\ \left. + (d_1 + d_2) \sqrt{2} \phi\left(\frac{d_1 - d_2}{\sqrt{2}}\right) \right] - \mu_{1+2}^2$	1.5
2-look _{AND}	$(DV_1 > \lambda_1) \cap (DV_2 > \lambda_2)$	d' $\Phi^{-1}(H) - \Phi^{-1}(F)$ $H = \Phi(d'_1 - \lambda_1) \Phi(d'_2 - \lambda_2)$ $F = \Phi(-\lambda_1) \Phi(-\lambda_2)$	1
2-look _{OR}	$(DV_1 > \lambda_1) \cup (DV_2 > \lambda_2)$	d' $\Phi^{-1}(H) - \Phi^{-1}(F)$ $H = 1 - \Phi(-\lambda_1) \Phi(-\lambda_2)$ $F = 1 - \Phi(\lambda_1) \Phi(\lambda_2)$	1
Sum _{early}	$\omega_1(S_1 + N_1) + \omega_2(S_2 + N_2)$	d' $(\omega_1 d'_1 + \omega_2 d'_2) / \sqrt{\omega_1^2 + \omega_2^2}$	1
Sum _{correlated early}	$\omega_1(S_1 + N_1) + \omega_2(S_2 + N_2)$	d' $(\omega_1 d'_1 + \omega_2 d'_2) / \sqrt{\omega_1^2 + \omega_2^2 + 2\omega_1 \omega_2 \rho}$	1
Sum _{multis early}	$\omega_1(S_1 + N_1) + \omega_2(S_2 + N_2)$	d_a $(\omega_1 d_{a1} + \omega_2 d_{a2}) / \sqrt{\omega_1^2 + \omega_2^2}$	1
Sum _{non-norm early}	$\omega_1(S_1 + N_1) + \omega_2(S_2 + N_2)$	AUC $< \Phi\left(\frac{\mu_{1+2} + \sigma_{1+2}}{\sqrt{2}}\right)$ $\mu_{1+2} = \omega_1 \Phi^{-1}(AOC_1) \sqrt{2} + \omega_2 \Phi^{-1}(AOC_2) \sqrt{2}$ $\sigma_{1+2} = \sqrt{\omega_1^2 + \omega_2^2}$	1
Sum _{late}	$\omega_1 S_1 + \omega_2 S_2 + N_L$	d' $\omega_1 d'_1 + \omega_2 d'_2$	1
Sum _{multis late}	$\omega_1 S_1 + \omega_2 S_2 + N_L \gamma^{(\omega_1 S_1 + \omega_2 S_2)}$	d_a $(\omega_1 \mu_1 + \omega_2 \mu_2) / \sqrt{(1 + [\gamma^{(\omega_1 S_1 + \omega_2 S_2)}])^2} / 2$	1
Sum _{mixed}	$\omega_1(S_1 + N_1) + \omega_2(S_2 + N_2) + N_L$	d' $(\omega_1 d'_1 + \omega_2 d'_2) / \sqrt{\omega_1^2 P_E + \omega_2^2 P_E + (1 - P_E)}$	1
Superadd.	$S_1 + S_2 + \sqrt{S_1 S_2} \delta + N_L$	d' $d'_1 + d'_2 + \delta \sqrt{d'_1 + d'_2}$	1

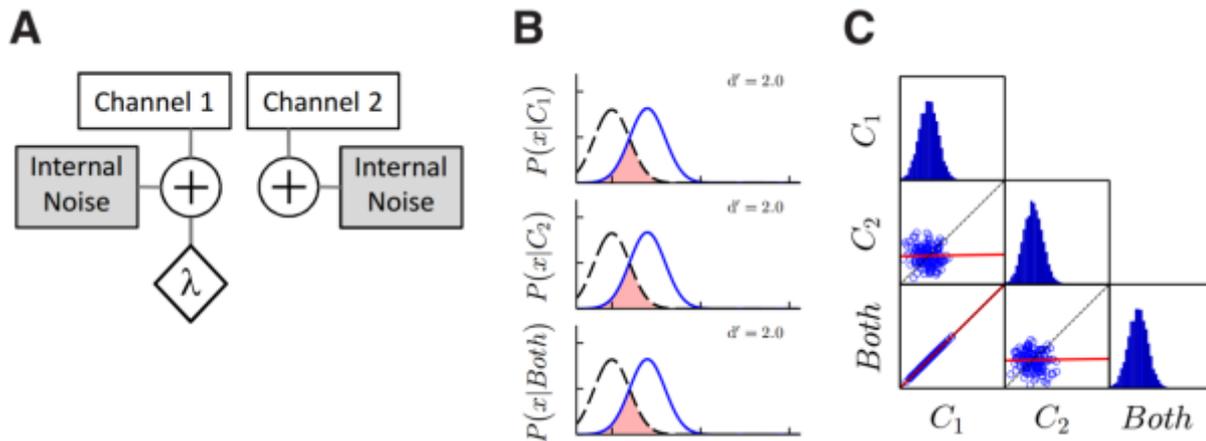
220



221
 222
 223
 224
 225

Fig 2. Percent gain in combined-sensitivity (vs. Cue₁ alone), as relative sensitivity to Cue₂ increases. For example, in the “Sum_{early}” model the benefit of having a second cue increases exponentially as the difference in single-cue sensitivity decreases (red line, right 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 I).

226 §2.1 One-look



227

228 **Fig 3.** 1-look cue-combination schema. **(A)** box-and-arrow logic diagram, showing how internal noise is added to
 229 the signals of each cue, and how signals are combined to make a decision. **(B)** Decision Variable (DV) distributions
 230 for signal-absent (black dashed) and signal-present (blue solid) conditions, for Cue₁ only (top), Cue₂ only (middle),
 231 and when both Cues are present together (bottom). These diagrams follow the same format as that shown
 232 previously in Fig 1C. Note that with the 1-look strategy, the DV in the combined condition is fully determined by
 233 the internal response to a single cue (see Eq 2.1.1c). The observer is therefore no more sensitive in the ‘both’
 234 condition, than in either of the single cue conditions. **(C)** Matrix of scatter plots, showing correlations between
 235 10,000 trial-by-trial DV values for each of the three conditions (Cue₁, Cue₂, Both; signal present condition only).
 236 Marginal histograms show the univariate distributions for each of the three conditions (i.e., these histograms
 237 correspond to the solid blue curves in Panel B). Red lines show the best fitting regression slope between each pair
 238 of conditions (flat line if no relationship between conditions). Only the lower-triangle of the matrix is shown, as
 239 the upper diagonal values are identical. In the case illustrated here, the observer only attends to Cue₁. The DV in
 240 the combined condition is therefore identical to that in the Cue₁ condition, and is totally uncorrelated with the DV
 241 in the Cue₂ condition. Alternatively, the observer could attend solely to Cue₂ (perfect correlation with Cue₂), or
 242 switch between cues (partial correlation with both cues).

243 In the 1-look strategy (Fig 3), the observer bases their decision solely on a single cue (i.e., they
 244 ‘look’ only at one component of the sensory input; see Ref~[43] for further background on this
 245 nomenclature). The other cue is simply ignored. The observer therefore gains no benefit from
 246 multiple cues, and would never be expected to perform better than in the best single-cue condition.

247 To formalize the 1-look decision process, let the response to a signal-absent stimulus be 0, and let
 248 the response to a signal-present stimulus be S_1 and S_2 (for cues 1 and 2, respectively). On each trial,
 249 the stimulus response is jittered by noise samples, N_1 and N_2 , each drawn independently from their
 250 corresponding noise distribution. For now, we shall assume that these distributions are zero-mean
 251 Gaussians, with standard deviations σ_1 and σ_2 (where σ represents the ‘magnitude’ of internal
 252 noise). We shall further assume that these values are combined additively with the initial stimulus
 253 response. Thus, the DV, is $S_1 + N_1$ for Cue₁, and $S_2 + N_2$ for Cue₂. As described in §1.1, the relevant
 254 DV is then compared to a criterion, λ , in order to determine a response. Thus, the decision
 255 strategies for the two individual cue conditions are:

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

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

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

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

257 where i in the two-cue case is equal to either 1 or 2. Note that there is no guarantee that the
258 response criterion in the combined-cue condition (Eq 2.1.1c) is the same as that in the single-cue
259 condition (Eq 2.1.1a-b). Thus, the simultaneous presence of the second cue could conceivably bias
260 the observer towards or against a particular response. Such bias could affect some measures of
261 performance, but should not affect d' , which is specifically designed to be independent of bias
262 (though cf. Ref ~[51]). Similarly, any such bias should affect the point of subjective equality [PSE]
263 of a psychometric function, but not its slope.

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

$$d'_{1-look}{}^{min} = \min \{d'_1, d'_2\} \quad (2.1.2a)$$

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

$$d'_{1-look}{}^{max} = \max \{d'_1, d'_2\} \quad (2.1.2b)$$

272 Finally, the observer may *alternate*: using one cue on some trials, and the other cue on other trials.
273 Such alternation is often observed anecdotally in psychophysical experiments, where it may be
274 variously attributed to: demand characteristics, exploration of the 'gain landscape' of the task⁵⁴, an
275 attempt to mitigate the effects of sensory adaptation, and/or simply an expression of boredom. If
276 we assume for the moment that the trial-by-trial decision regarding which cue to attend to is made
277 *a priori*, independent of the sensory evidence, then the result of switching is an intermediate level
278 of sensitivity, somewhere between that of the best (d'^{max}) and worst (d'^{min}) single-cue.

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

$$\begin{aligned} h_{1-look}^{switch} &= P_1 h_1 + (1 - P_1) h_2, \\ f_{1-look}^{switch} &= P_1 f_1 + (1 - P_1) f_2, \end{aligned} \quad (2.1.3)$$

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

$$d'_{1-look}{}^{switch} = \Phi^{-1}(h_{1-look}^{switch}) - \Phi^{-1}(f_{1-look}^{switch}). \quad (2.1.4)$$

284 Although generally considered a poor strategy, several considerations actually favor the use of
285 only a single cue. Firstly, finite cognitive (attention, memory) resources may prohibit the observer
286 from attending to more than one cue. Secondly, the *cost* of attending to multiple cues may
287 outweigh the perceived benefit. Thus, although cue-combination is often assumed to be without
288 cost, processing a second cue may be effortful or confusing for some observers. In contrast, the
289 gains of cue-combination are often small, either because the observer is already performing at
290 ceiling, because the observer is content with their single-cue performance, or because the second
291 cue genuinely provides relatively little additional information (see Fig 2A). A rational observer

292 may therefore trade-off cost against expected reward, and decide to predicate their decisions upon
293 only a single cue. Whether observers do make such calculations, either implicitly or explicitly, is
294 largely unknown, and to our knowledge, no attempts have been made to quantify perceived
295 cost/benefits in sensory integration tasks.

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

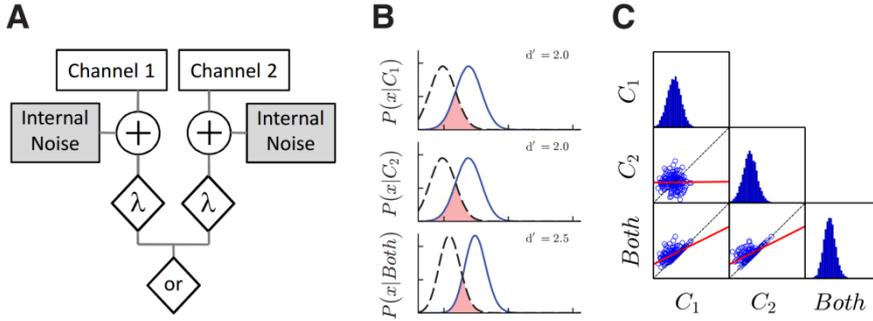


Fig 4. 2-look_{OR} cue-combination schema; same format as **Fig 3**. The observer responds if *either* the Cue₁ or Cue₂ internal response exceeds criterion, λ . The DV distributions in the 'both' condition (bottom of Panel B) are the maximum of the two corresponding single-cue distributions, and upon close inspection can be seen to exhibit a slight rightward-skew.

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

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

or if *both* internal response exceeded criterion:

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

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:

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

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

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

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:

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

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

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

318 In either case, using the outputs of **Eq 2.2.3a** and **Eq 2.2.3b**, d' for the combined condition can be
319 computed in the standard manner⁴¹⁻⁴³, based upon the difference in hits and false alarms (**Eq**
320 **1.1.2**). Notably, when the internal noise is additive (and signal and noise trials occur with equal
321 frequency), then the ideal criterion location, λ , is equal to $d'/2$. Thus, if the observer is unbiased
322 then the *difference* between hits and false alarms is equal in both the AND and the OR cases, and
323 expected sensitivity is given by:

$$\begin{aligned} d'_{2\text{-look}} &= \Phi^{-1} \left(\Phi \left(\frac{d'_1}{2} \right) \Phi \left(\frac{d'_2}{2} \right) \right) - \Phi^{-1} \left(\Phi \left(\frac{-d'_1}{2} \right) \Phi \left(\frac{-d'_2}{2} \right) \right) \\ &= \Phi^{-1} \left(1 - \Phi \left(\frac{-d'_1}{2} \right) \Phi \left(\frac{-d'_2}{2} \right) \right) - \Phi^{-1} \left(1 - \Phi \left(\frac{d'_1}{2} \right) \Phi \left(\frac{d'_2}{2} \right) \right) \end{aligned} \quad (2.2.3)$$

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

327 Three further features of 2-look strategy are also worth noting. Firstly, although both the AND and
328 OR strategies may predict the same improvement in sensitivity, they can be distinguished
329 empirically by examining the SDT bias parameter, c ⁵⁵. Thus, the OR strategy would result in a bias
330 towards responding 'Signal Present' (liberal observer), while the AND strategy would result in a
331 bias towards responding 'Signal Absent' (conservative observer). Secondly, in both cases, percent
332 correct would actually be no different to in the simple, 1-look strategy. Third and finally, note that
333 although the 2-look strategy guarantees some improvement in sensitivity when the two cues are
334 equally informative, when $d'_1 \neq d'_2$ sensitivity in the combined condition may actually be *lower*
335 than in the best single-cue condition (**Fig 2A**). This is in contrast to the linear-summation models
336 that we consider below, where the ideal observer's sensitivity is guaranteed to increase as a
337 function of the number of cues (**Fig 2B**).

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

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

$$P(\max \{X, Y\} \leq z) = P(X \leq z)P(Y \leq z). \quad (2.2.4)$$

343 With this in mind, it can be seen that 2-look_{OR} (**Eq. 2.2.1a**) is formally equivalent to a *prima facie*
344 quite different strategy, in which the observer bases their response on whichever cue produced the
345 greatest internal response on that particular trial. We shall denote this strategy 'Max-DV'ⁱⁱⁱ, and
346 define it formally as:
347

$$R_{\text{max-DV}} = \begin{cases} \text{'Signal Present'} & \text{if } \max \{S_1 + N_1, S_2 + N_2\} > \lambda \\ \text{'Signal Absent'} & \text{otherwise} \end{cases}. \quad (2.2.5)$$

ⁱⁱⁱ Note, that the Max-DV rule (respond based on greatest trial-by-trial internal response) should not be confused with the 1-look_{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

348 Note, however, that the maximum of two Gaussian random variables is not itself Gaussian
349 distributed (see Ref~[56]). This is potentially problematic, because, as detailed in §1.1, the
350 sensitivity metric d' implicitly assumes normality. A practical corollary of this is that empirical d'
351 values will vary somewhat, depending on where observers place their criterion (see simulations in
352 **Appendix A**). This confound between bias and sensitivity is negligible when sensitivity to the
353 individual cues is low (e.g., $d' = 1.0$), but increases when single-cue sensitivity is greater. For
354 example, when $d_1' = 4.0$ and $d_2' = 4.0$, the value of d' in the combined condition may vary by
355 approximately $\pm 15\%$, depending on the observer's exact criterion placement. The level of error
356 also increases as the number of cues increases beyond two. For example, the maximum of 100
357 independent Gaussian variables exhibits a heavy rightward-skew.

358 For many experiments, this error is unlikely to be of substantive concern. However, for
359 experimenters requiring exactitude, one way to deal with a deviation from internal-response
360 normality is to measure performance as the criterion location is systematically varied (ROC
361 analysis -- see §2.6). In this way, the effect of criterion can be partial-out. As detailed in §2.6, such
362 an analysis can provide criterion-invariant metric of sensitivity -- for example, the term d^a , which
363 we define later in Eq. 2.6.3. Notably, in the simple, two-cue case it is possible to predict d^a in the
364 combined cue condition, using just the single cue measurements of d' . To see how this can be done,
365 note that the DV in the Max-DV/2-look-OR model can be approximated by a Gaussian variable⁵⁶,
366 with an expected mean and standard deviation of:

$$\begin{aligned} \mu_{max} &= \mu_1 \Phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) + \mu_2 \Phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) + \theta \phi\left(\frac{\mu_1 - \mu_2}{\theta}\right), \\ \sigma_{max} &= \sqrt{(\sigma_1^2 + \mu_1^2) \Phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) + (\sigma_2^2 + \mu_2^2) \Phi\left(\frac{\mu_2 - \mu_1}{\theta}\right) + (\mu_1 + \mu_2) \phi\left(\frac{\mu_1 - \mu_2}{\theta}\right) - \mu_{max}^2}, \end{aligned} \quad (2.2.6)$$

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

$$\theta = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \quad (2.2.7)$$

371 By substituting the appropriate values for each individual cue, the DV for each of the noise ($\mu_i = 0$;
372 $\sigma_i = 1$) and signal ($\mu_i = d'_i$; $\sigma_i = 1$) conditions can be derived. The resultant estimates of μ_{noise} , μ_{signal} ,
373 σ_{noise} , σ_{signal} can then be combined using Eq 1.1.2 to compute expected d^a . Furthermore, since three
374 of the DV values (μ_{noise} , σ_{noise} , σ_{signal}) are defined as constants (see §1.1), and if we assume that the
375 cues are independent ($\rho = 0$; though cf. §2.4), then this model reduces to:

$$d_{max-DV}^a = \frac{\mu_{signal} - 0.56}{\sqrt{\frac{1}{2}(\sigma_{signal}^2 + 0.68)}}, \quad (2.2.8a)$$

376 where

$$\begin{aligned} \mu_{signal} &= d_1' \Phi\left(\frac{d_1' - d_2'}{\sqrt{2}}\right) + d_2' \Phi\left(\frac{d_2' - d_1'}{\sqrt{2}}\right) + \sqrt{2} \phi\left(\frac{d_1' - d_2'}{\sqrt{2}}\right), \\ \sigma_{signal} &= \sqrt{(1 + d_1'^2) \Phi\left(\frac{d_1' - d_2'}{\sqrt{2}}\right) + (1 + d_2'^2) \Phi\left(\frac{d_2' - d_1'}{\sqrt{2}}\right) + (d_1' + d_2') \phi\left(\frac{d_1' - d_2'}{\sqrt{2}}\right) - \mu_{max}^2}, \end{aligned} \quad (2.2.8b)$$

377 To summarize, the 2-look_{OR} decision strategy is equivalent to a strategy in which responses are
378 based on the greatest internal response to all cues (Max-DV). In the Max-DV model, the DV is not
379 Gaussian, and so d' is liable to misestimate sensitivity. Furthermore, Max-DV is formally identical
380 to 2-look_{OR}. By extension, d' is therefore not an ideal metric for observers using the 2-look_{OR}/2-
381 look_{AND} strategies either. The error may be tolerable if sensitivity and/or the number of cues are
382 low. However, in other cases alternative measures of sensitivity may be more appropriate, such as
383 d^a (see §2.6). A formula for approximately predicting d^a based on single-cue d' estimates is given in
384 Eq 2.2.8.

385 ***Practical advantages of a 2-look decision strategy***

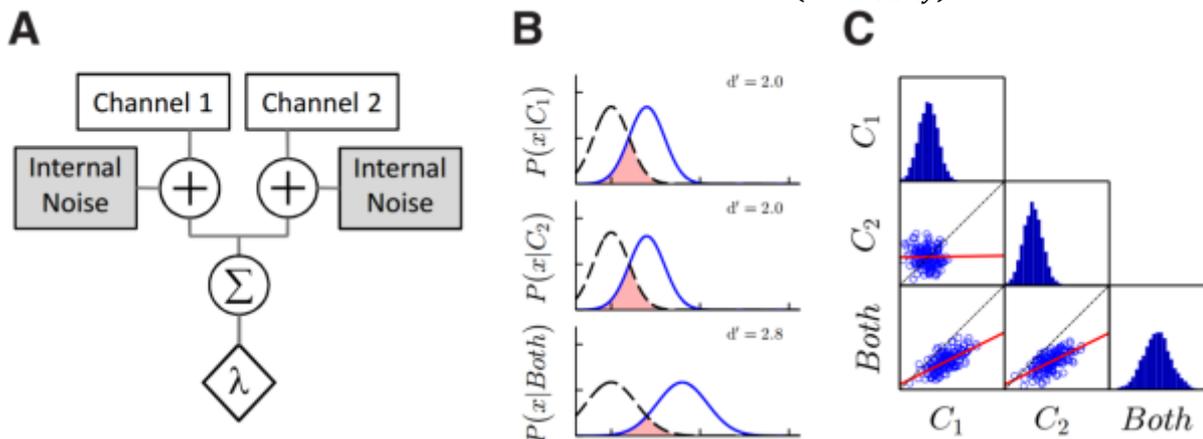
386 As with 1-look strategies, 2-look (probability summation) models are generally considered
387 relatively crude – providing only a modest improvement in sensitivity, relative to the linear-
388 summation models that we shall consider next. However, as discussed previously with regard to
389 the 1-look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

399 Thus, an ideal observer with limited working memory might use a powerful linear-summation
400 model (§2.3) when the number of cues is low, but revert to an N -look strategy (or similar), once
401 some putative processing threshold is exceeded. In normal adults, and given only two cues, such a
402 threshold is unlikely to be reached. However, if capacity is diminished (e.g., under conditions of
403 high load^{57,58}, or among children^{59,60} and older adults⁶¹), or if the demands of the task are
404 increased (e.g., by increasing the number of cues), then processing costs may start to become a
405 limiting factor. To our knowledge, this hypothesis has not been studied systematically within the
406 cue-combination literature. However, limited supporting data do exist. For example, in the
407 repeated-observation task described above, observers have been shown to exhibit a \sqrt{N}
408 improvement in sensitivity when the number of cues is low ($N = 1—5$)¹¹. This exceeds the
409 predictions of the N -look model, and is suggestive of a more complex linear-summation model
410 (§2.3). However, when the number of cues is increased to seven¹³ (i.e., close to⁶², or exceeding⁶³
411 the limit of human working memory), relative performance is diminished, and observers appear
412 no longer able to integrate cues optimally. Similar effects have also been reported for observers
413 asked to form ‘summary statistics’ (e.g., average the size or orientation) of an array of objects^{64–66}.
414 For example, given a mean-size-discrimination task, Solomon and colleagues⁶⁷ reported that
415 observers benefited from having two or four cues, but showed no significant improvement when
416 the number of cues was increased to eight. The exact reason for this decrease in efficiency at high
417 numbers of cues is unclear. For example, it may be that observers resort to a qualitatively different
418 N -look strategy when the numbers of cues is high, or it may be that observers continue to
419 integrate, but are quantitatively constrained in their efficiency by finite attention or memory^{iv}.

^{iv} For example, Gorea and colleagues⁶⁴ 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^{140,141}.

420 limits. Irrespective of the true explanation, that human observers behave in this way does at least
421 highlight the importance of considering observers' limited processing resources when determining
422 what constitutes 'ideal' performance. Thus, algorithmic limitations may prove instructive for
423 understanding why observed sensitivity is often less than predicted in children²⁶⁻²⁸ and older
424 adults³⁰, and in normal adults performing complex tasks¹³.

425 **§2.3 Linear summation with early, independent noise (Sum_{early})**



426
427 **Fig 5.** Linear summation with early, independent noise; same format as **Fig 3**. Independent noise is added to each
428 cue prior to summation (Panel A). In the case illustrated here, the observer is equally sensitive to each cue ($d'_1 =$
429 d'_2) and acts in the ideal manner by weighting each cue equally. Due to cancelation of noise between cues, the
430 distributions in the combined cue condition therefore exhibit substantially ($\sqrt{2}$) less overlap than in either of the
431 single cue conditions (Panel B).

432 In linear-summation models, the internal responses to each cue are summed to create the DV (see
433 **§1.2**). Mathematically, this summation causes independent samples of noise to partially cancel out:
434 improving the Signal to Noise Ratio [SNR], and thereby increasing sensitivity^v. Physiologically, the
435 process is plausible, *prima facie*, as it has long been known that individual neurons sum their
436 dendritic inputs⁶⁸, and there is growing evidence of sensory integration at the network level
437 also^{24,69,70}.

438 The defining property of the *Early Noise* linear-summation model (**Fig 5**) is that all the internal
439 variability is introduced into the system prior to the two cues being combined (i.e., ‘early’ – see
440 **§2.5** for further definition of ‘early’ versus ‘late’). Thus:

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

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

^v Due to the Law of Large Numbers¹⁴², an extreme noise sample for one cue is likely to be partially cancelled 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 and Creelman, 2005, pp. 166 – 170⁴²), and why, in statistics, Standard Error decreases with \sqrt{N} data points.

451 therefore given by:

$$d'_{Sum:early} = \frac{\omega_1 d'_1 + \omega_2 d'_2}{\sqrt{\omega_1^2 \sigma_1^2 + \omega_2^2 \sigma_2^2}} \quad \text{where} \quad \max(|\omega_i|) = 1, \sigma_i = 1. \quad (2.3.2)$$

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

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

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

457 **How to compute ideal relative weight coefficients, ω_i**

458 Intuitively, it is obvious that one should give less weight to less informative cues, and more
459 weight to more informative cues. More formally, it has been shown by previous authors that
460 the ideal strategy is to weight each cue proportional to its signal-to-variance ratio^{71,72}. Since d'
461 has already been scaled to be a random Gaussian variable with a variance of one ($\sigma^2 = 1$), ideal
462 relative weights are simply proportional to d' . Given our present convention of normalizing
463 weights so that the greatest weight magnitude, $|\omega|$, always equals one (see **§1.2**), the ideal
464 weight for the i th cue is therefore:

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

465 For example, if $d'_1 = 1$, and $d'_2 = 1.5$, then the ideal weights would equal $\omega_1 = 0.67$ and $\omega_2 =$
466 1.00 , and ideal sensitivity would be $d'_{1+2} = 1.58$. More generally, by combining **Eq. (2.3.2)** and
467 **Eq. (2.3.4)**, and assuming that all d' values are positive, ideal sensitivity reduces to simply:

$$d'_{Sum:early} = \sqrt{d_1'^2 + d_2'^2} \quad (2.3.5)$$

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

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

469 When considering sensitivity, this scheme is formally equivalent to the ‘max’ approach of **Eq**
470 **2.3.4**, and results in the same quantitative predictions (e.g., in terms of d'). Conceptually,
471 however, **Eq 2.3.4** is more natural when considering detection (where cues are thought to be
472 ‘summed’), while **Eq 2.3.4** is more suited to estimation tasks (where cues are thought to be
473 ‘averaged’). As discussed previously (**§2**), we shall, for simplicity, restrict ourselves to the case
474 of detection, and so use **Eq 2.3.4** throughout. **Eq 2.3.4** also has the advantage that in many
475 situations (e.g., when sensitivity is equal across all cues) all weights equal one; in which case,
476 the ω_i terms can be omitted from many equations altogether.

477 **How do observers know how to weight cues appropriately?**

478 One possibility is that observers learn to weight cues appropriately through practice, for
 479 example via a process of supervised learning. In this case, weightings should improve over
 480 time, and, in the limiting case where the task is entirely novel, observers should be at chance on
 481 trial one. Another possibility is that reliability is estimated directly, within a single trial. This
 482 could be achieved if, for example, information was encoded by a distributed population code.
 483 Thus, if the internal response was a distribution, rather than a single scalar value, then the
 484 variance of the population-response distribution could be used as an estimate of the amount of
 485 signal uncertainty^{4,73,74}. If this were the case, then observers could be capable of assigning
 486 appropriate weights to completely novel stimuli, and of responding on a trial-by-trial basis to
 487 random variations in relative SNR between cues. The evidence appears to somewhat favor the
 488 latter, ‘dynamic reweighting’, hypothesis. Thus, in experiments where the relative reliability of
 489 two cues is randomly varied trial-by-trial, both humans²³ and macaque monkeys^{18,24} have been
 490 shown to update weightings across trials. However, it is possible that such weights are also
 491 refined across trials, and there is some evidence in perceptual learning paradigms that cue
 492 weightings do improve with repeated practice^{75–77}.

493 **Expected sensitivity in the early weighted linear summation model**

494 Under ideal conditions, with two cues of equal sensitivity, and equal weights given to each cue,
 495 sensitivity will increase by a factor of $\sqrt{2}$ (41%; as shown graphically in **Table I**). More
 496 generally, ideal sensitivity will increase by \sqrt{N} , where N is the number of cues. If the two cues
 497 are not equally useful, but the weights are adjusted to remain ideal, then sensitivity is still
 498 guaranteed to be greater than in the best single cue (**Eq 2.3.5**). However, the magnitude of the
 499 benefit falls away exponentially as the observer’s sensitivity to the two individual cues begins
 500 to differ (**Fig 2B**, black dashed line). For example, if an observer is only half as sensitive to the
 501 second cue ($d'_1 = d'_2/2$) then d' in the combined condition would only be expected to increase
 502 by a maximum of 12%. Given that estimates of d' often have a standard error on the order of
 503 $\pm 50\%$ (e.g., given 50 trials, see **Appendix B**), any asymmetry will therefore greatly reduce the
 504 likelihood that observed sensitivity will improve significantly.

505 Finally, if weights are not optimal, then some information will be lost, and sensitivity in the
 506 combined condition can actually be *reduced* relative to the best single cue (**Fig 2B**, green dot-
 507 dashed line). Suboptimal weightings could occur if the observer misestimated the reliability of
 508 each cue, or if observers had some *a priori* bias towards a particular cue (for discussion, see
 509 Ref~[18]). An important, but often underappreciated point, is that although the early-noise linear-
 510 summation model predicts at most a \sqrt{N} improvement in sensitivity, an increase of less than \sqrt{N}
 511 does not mean that fewer than N cues are being used (as is sometimes, erroneously implied⁷⁸). For
 512 example, **Figure 6** shows a number of relative-weight functions that would lead to a $\sqrt{3}$
 513 improvement, only one of which involves giving no weight to two cues.

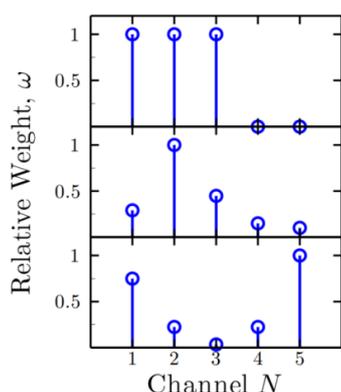
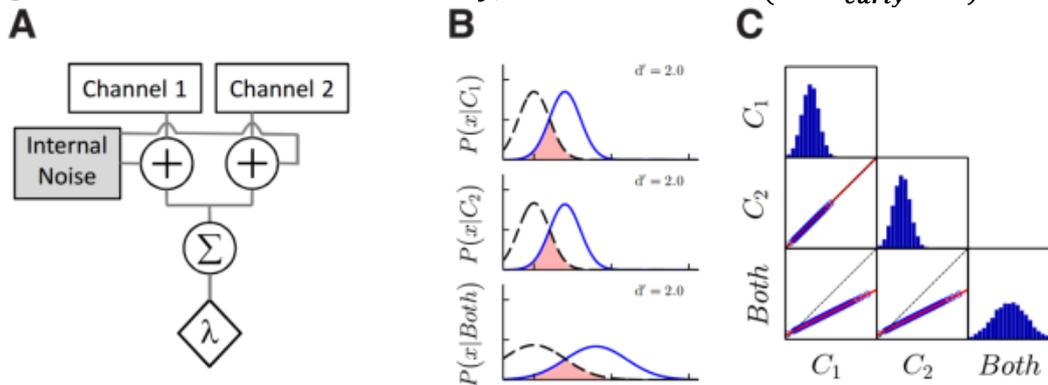
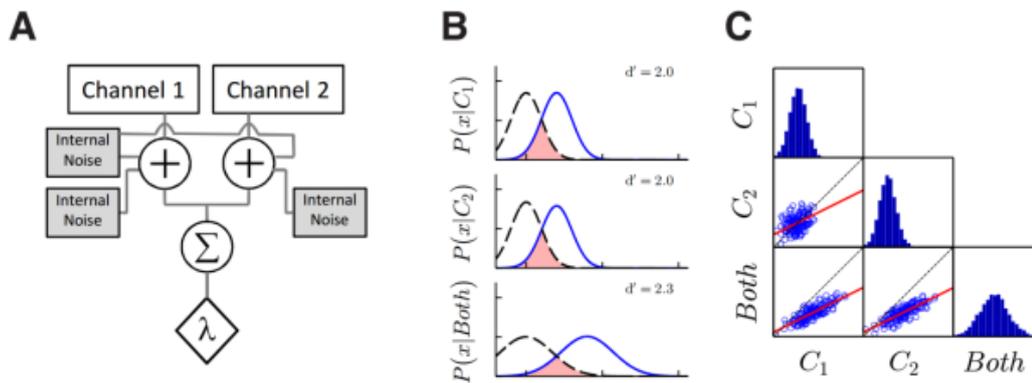


Fig 6. Three ways of weighting multiple sensory cues ($N = 5$), each of which predicts a $\sqrt{3}$ improvement in sensitivity (assuming that all cues are equally useful). Note that here we have scaled the weights so that the greatest weight is one, and the other weights are expressed as a fraction of 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**).

523 §2.4 Linear summation with early, correlated noise ($Sum_{early}^{correlated}$)



524
525 **Fig 7.** Linear summation with early, correlated noise; same format as Fig 3. As shown in Panel C, the internal
526 responses to the two cues (Cue₁ vs. Cue₂) are perfectly correlated. Therefore, while the sum of these two variables
527 is greater in magnitude than either alone, there is no cancelation between noise samples, and so no change in the
528 amount of overlap between the ‘Signal Absent’ and ‘Signal Present’ distributions in the combined condition (Panel
529 B, bottom).



530
531 **Fig 8.** Half-correlated cue-combination schema; same format as Fig 3. Here the correlation coefficient was $\rho = 0.5$,
532 though in practice ρ could be any value between 0 (total independence; Fig 6) and 1 (total redundancy; Fig 7). As
533 a result, there was partial trial-by-trial correlation between internal responses to each (Panel C), and some
534 increase in over-all sensitivity in the combined condition (Panel B).

535 A key assumption in the standard (early-noise) linear-summation model, described previously in
536 §2.3, is that the internal noise associated with each cue is independent. However, this is unlikely to
537 be strictly correct. Within a population of neurons, correlations exist between the firing patterns of
538 individual cells⁷⁹, and voltage signals across the cortex exhibit regular oscillations over time^{80vi}.

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

^{vi} Conversely, there have been reports that some neighboring cortical neurons appear to behave with statistical independence¹⁴³. 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^{100,144}

^{vii} NB: assuming that external noise levels are negligible. In the real world, and in some psychophysical studies also^{10,145}, 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.

$$d'_{Sum:corr} = \frac{\omega_1\mu_1 + \omega_2\mu_2}{\sqrt{\omega_1^2\sigma_1^2 + \omega_2^2\sigma_2^2 + 2\rho\omega_1\sigma_1\omega_2\sigma_2}} = \frac{\omega_1d'_1 + \omega_2d'_2}{\sqrt{\omega_1^2 + \omega_2^2 + 2\rho\omega_1\omega_2}}, \quad (2.4.1)$$

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

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

547 The formal proof of [Eq \(2.4.2\)](#) is given elsewhere⁸¹. However, it can be understood intuitively as
548 representing a ‘relative utility’ weighting, in which each cue is weighted proportional to its
549 reliability, but only after correcting for redundancies between cues.

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

$$d'_{Sum:corr} = \frac{\sqrt{2}d'}{\sqrt{1 + \rho}}. \quad (2.4.3)$$

555 (N.B. when the cues are not equally useful then expected sensitivity values would have to be
556 computed using [Eq 2.4.1](#).) [Equation 2.4.3](#) implies that a *negative* correlation between cues (ρ
557 < 0) would actually cause sensitivity to *increase*. This is in fact the case, as can be seen in the
558 simulation presented in [Appendix A](#), and follows from the fact that a noise sample for one cue
559 would tend to be cancelled out by an opposing/complimentary noise sample in the other.

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

561 One suggested technique (Wickens, 2002, pp 184 - 186⁴³) for quantifying ρ is to infer the
562 degree of correlation from the ratio of observed sensitivity, d'_{obs} , to predicted sensitivity given
563 uncorrelated internal noise, d'_{pred} :

$$\frac{d'_{obs}}{d'_{pred}} = \frac{d'_{Sum:corr}}{d'_{Sum:early}} = \frac{\frac{\omega_1d'_1 + \omega_2d'_2}{\sqrt{\omega_1^2 + \omega_2^2 + 2\rho\omega_1\omega_2}}}{\frac{\omega_1d'_1 + \omega_2d'_2}{\sqrt{\omega_1^2 + \omega_2^2}}}, \quad (2.4.4)$$

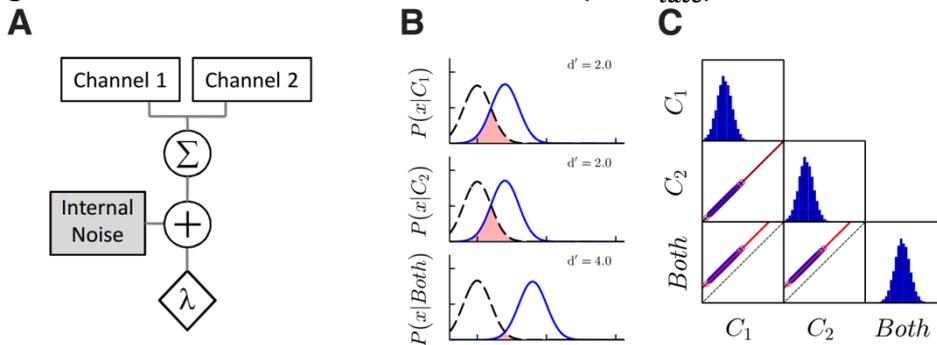
564 which, solving for ρ , yields:

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

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

573 Ultimately, the only robust way to estimate ρ would be to compare the trial-by-trial internal
574 responses to each cue (**Fig 8C**). This would be difficult to achieve behaviorally, however, as it
575 would require observers to make independent judgments of each cue simultaneously. This is
576 conceptually possible, but we know of no study where this has been attempted, and the obvious
577 concern is that the two stimuli/judgments would affect each other, either negatively (interference)
578 or positively (compulsory integration^{26,82}). Any such interaction would be evident by a change in
579 sensitivity relative to the single-cue conditions, and would cause ρ to be over or under estimated,
580 respectively. Alternatively, one might try to use neuroimaging to estimate an observer's internal
581 responses directly⁸³; however, such techniques are still in their infancy, and have yet to be proven
582 robust.

583 §2.5 Linear summation with late noise (Sum_{late})



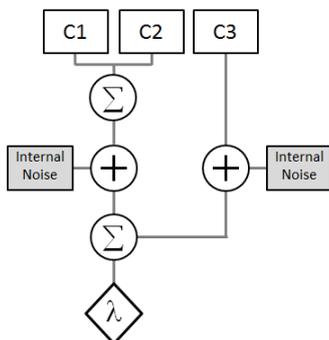
584

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

590 Up until this point, we have assumed that all internal noise arises *early* in the decision-making
 591 process (i.e., before the point at which the internal responses to each sensory cue are integrated).
 592 However, internal noise may also be introduced at a later stage. For example, the circuits that
 593 perform the integration may themselves be subject to random error, or the observer may be
 594 unable to maintain a stable decision criterion^{84,85}, or the measured response of the observer may
 595 itself be stochastic (e.g., in a pointing or reaching task). Thus, Figure 9 shows the opposite
 596 extreme, in which *all* variability is introduced after the sensory signals have been combined (Late
 597 Noise). The decision strategy for late-noise linear-summation is:

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

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



604

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

607 If, as is typically the case in psychophysical paradigms, the amount of external noise is minimal, the
 608 ideal late-noise observer should always weight each cue equally, as each cue provides only signal
 609 (in contrast to the early-noise model, where each cue also contributes additional noise). Ideally,

610 the relative weights should therefore all equal 1, and the weight terms, ω , can simply be omitted.
611 Accordingly, ideal performance is given by:

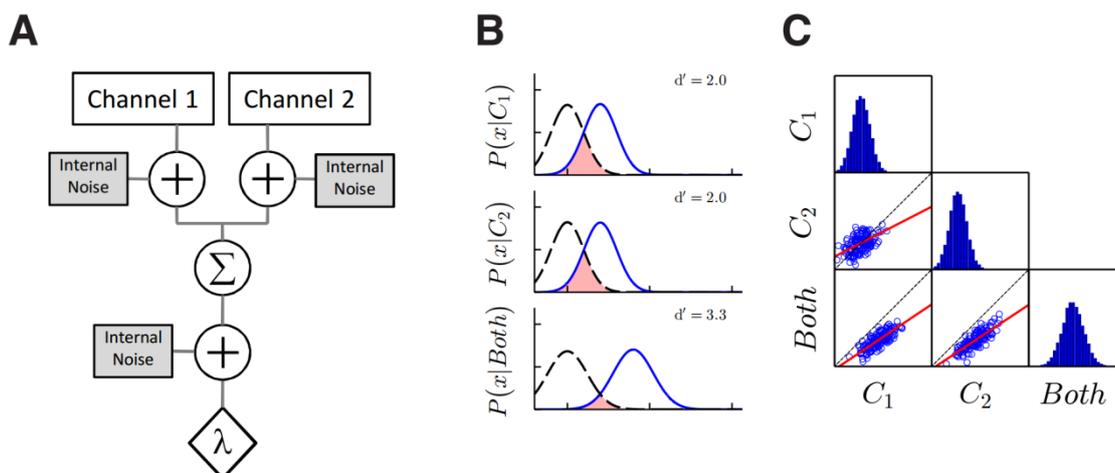
$$d'_{Sum:late} = \frac{1d'_1 + 1d'_2}{1} = d'_1 + d'_2. \quad (2.5.2a)$$

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

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

615 It is clear that the potential gains are far larger than with the models discussed previously. Thus, if
616 all cues are equally informative then ideal sensitivity will increase by a factor of N , versus the \sqrt{N}
617 predicted by the early-noise linear-summation model (Eq 2.3.3). A minority of authors have
618 reported cases of factor-of- N improvements⁴⁰. However, for most tasks the late-noise model is
619 actually logically implausible in its extreme form. For example, it would imply that there was some
620 location in the brain where the two stimuli are represented perfectly (i.e., such that they could be
621 summed together prior to any noise being introduced). Since even the peripheral mechanisms that
622 encode incoming sensory information act in a stochastic manner^{86,87}, it is difficult to imagine
623 where in the brain such a noiseless representation could exist. A more realistic scheme is therefore
624 the hybrid model (Fig 11), in which some noise is early, and some late:

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



625 **Fig 11.** Linear summation with half late noise, and half early noise; same format as Fig 3. The early noise
626 components are independent across cues, so the internal responses to each are only partially correlated (Panel C;
627 Cue₁ vs. Cue₂). The overall increase in sensitivity (Panel B) is an intermediate value, between the lower bound of
628 the early, independent noise model (Fig 5), and the upper bound of the pure late noise model (Fig 9).
629

630 Expected performance in the hybrid model is given by:

$$d'_{Sum:mix} = \frac{\omega_1 d'_1 + \omega_2 d'_2}{\sqrt{\omega_1^2 \sigma_1^2 + \omega_2^2 \sigma_2^2 + \sigma_L^2}} \quad \text{where} \quad \max(|\omega_i|) = 1, [\sigma_i + \sigma_L] = 1. \quad (2.5.4)$$

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

$$d'_{mix} \approx 2^{\frac{1}{1}} d' = N d'. \quad (2.5.5a)$$

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

$$d'_{mix} \approx 2^{\frac{1}{2}} d' = \sqrt{N} d'. \quad (2.5.5b)$$

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

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

642 **Empirical separating early and late noise**

643 Is it possible to determine what proportion of noise is early/late on the basis of sensitivity
644 alone? In principle, one could attempt to infer the ratio of early-to-late noise by comparing
645 observed sensitivity to the ideal. For example, if d' equals **Eq (2.5.5a)** then one might claim
646 that all noise is late, or if d' equals **Eq (2.5.5b)** then one might claim that all noise is early.
647 However, as discussed previously with regards to estimating internal-noise correlation (**§2.4**),
648 the difficulty with this inference-from-efficiency approach is that it requires us to assume that
649 no other factors limit performance. Thus, the level of sensitivity predicted by early noise (**Eq**
650 **2.5.5a**) could also be explained by 100% late noise and suboptimal weights (see **Fig 2B**, dot-
651 dashed green line), or by a mixture of late noise and correlated early noise.

652 Instead then, one possible way to disambiguate early and late noise would be to examine how
653 sensitivity changes with cue asymmetry. Thus, it was noted earlier that multi-cue sensitivity
654 decreases exponentially as the difference in single-cue sensitivity increases (**Fig 2B**, red dotted). In
655 contrast, with late-noise, benefit varies linearly with cue asymmetry (**Fig 2B**, solid line), and
656 various admixtures of early/late noise will fall along intermediate isobars (e.g., **Fig 2B**, black
657 dashed). These differing predictions could be used in principle to differentiate between relative
658 quantities of early vs. late noise. We are not aware of any attempts to perform such an analysis,
659 and most existing datasets would be unsuitable, since experimenters tend to design cues to always
660 be equally useful (since this is when expected gain is maximized --- see **§3.2**). Moreover, such a
661 technique implicitly assumes that the late noise is additive, and does not vary with combined
662 internal response magnitude; an assumption which is itself open to question.

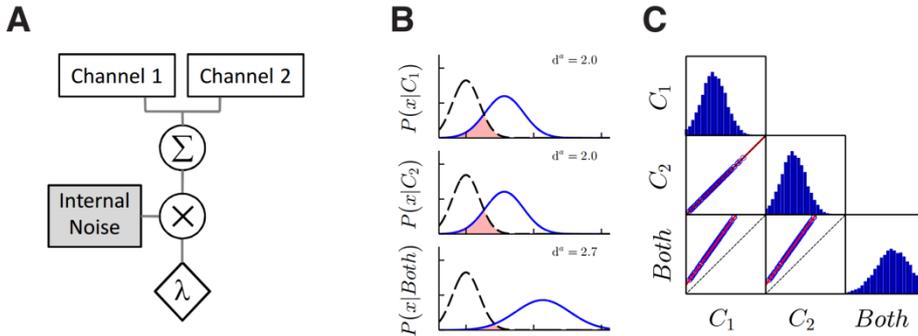
663 It is also important to note that the ratio of early-to-late-noise may not be stationary. For example,
664 it has been well-established that sensory judgements improve with practice (*perceptual*

665 *learning*^{88,89}), suggesting that sensitivity changes over time. This can lead to some interesting
666 predictions regarding how cue-combination strategies may differ between naïve and well-trained
667 observers. For example, it has been shown that some of the learnt improvements in sensitivity can
668 be attributed to reductions in nonstationary bias⁸⁵, which for present purposes can be considered
669 a source of late internal noise^{viii}. Thus, while sensitivity may primarily reflect sensory limitations in
670 well-trained observers (early noise), naïve observers may be more limited by late noise. Such
671 considerations lead to the novel, and somewhat counterintuitive prediction, that naïve observers
672 should gain *more* benefit from having multiple cues than well-practised observers do (i.e., since
673 before practice, decisions will be limited partially by late noise, and so sensitivity should increase
674 by more than the Pythagorean sum of the two cues). In contrast, after practice, sensitivity should
675 be determined by early noise alone, and so follow **Eq (2.3.3)**. If a greater proportion of internal
676 noise were 'late' in naïve observers, then this would also mean that the ideal weights for a naïve
677 and a trained observer would differ when the cues are of unequal utility. Thus, in well-practised
678 observers each cue should be weighted proportional to its reliability (as per the early-noise
679 model), whereas in naïve observers the weights should tend more towards equality, irrespective of
680 any asymmetry (as per the late-noise model). This may be an important consideration when
681 determining whether, for example, observers can learn to optimize their decision weights with
682 practice⁷⁵⁻⁷⁷ (i.e., since an observer could conceivably change their weights while remaining equally
683 efficient/inefficient).

684 Finally, it is worth noting that there is a potential contradiction between the correlated-early-noise
685 model and the late-noise linear-summation model. The late-noise model predicts that when cues
686 are similar, the information should be integrated early (i.e., more noise should be late), and so the
687 benefits of integration should be *greater*. Conversely, it might be argued that when two cues
688 activate similar sensory regions, more of the noise should be correlated across cues, and so the
689 benefits would be *smaller*. The evidence tends to favor the former viewpoint, with the benefits of
690 integration being greater when the two cues are more similar. Thus, when the cues are located in
691 different modalities^{22,90,91} or spatial locations⁹², improvements tend to follow the early-noise
692 predictions (**Eq. 2.3.2**). In contrast, reported improvements for two visual-depth-cues⁴⁰, or two
693 nearby retinal locations⁹², have tended to be greater, and to follow the late-noise prediction (**Eq.**
694 **2.5.2b**).

^{viii} NB: response bias ought to be independent of sensitivity, but, as discussed in Ref~[⁸⁵], the two factors are liable to be 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.

695 **§2.6 Linear summation with multiplicative noise ($Sum_{early/late}^{multi}$)**



696
697 **Fig 12.** Linear summation with multiplicative noise; same format as Fig 3. In the example shown here, all internal
698 noise is assumed to be late (Panel A). Since in this model internal noise magnitude increases as a function of signal
699 strength, internal noise magnitude is greater in the Signal Present condition (Panel B; solid blue line) than in the
700 Signal Absent condition (Panel B; dashed black line). This makes d^a a more appropriate metric of sensitivity than
701 d' . Furthermore, internal noise magnitude will increase more as N cues increases (Panel B; bottom). As a result,
702 sensitivity in the compound-cue condition will increase by less than that predicted by the additive late noise
703 model (Fig 9). In this case, the multiplicative factor is relatively low ($\gamma = 1.19$), so the overall increase in sensitivity
704 is similar to the linear summation model with early, independent noise (Fig 5). If γ were greater, the increase in
705 sensitivity would be smaller, and could even become negative.

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

710 The first point to note is that the effect of multiplicative noise depends on whether the internal
711 noise is early or late. In the Early Noise model, the multiplicative gain is applied before the signals
712 have been summed;

$$R_{Sum:multi}^{early} = \begin{cases} \text{'Signal Present'} & \text{if } (S_1 + N_1\gamma^{S_1} + S_2 + N_2\gamma^{S_2}) > \lambda, \\ \text{'Signal Absent'} & \text{otherwise} \end{cases}, \quad (2.6.2)$$

713 In both the single-cue and combined-cue condition, the strength of the sensory signal (and thus the
714 amount of multiplication) is identical at the point when internal noise is introduced. The fact that
715 the noise is multiplicative therefore has no effect on the predicted benefit of integration, which
716 remains unchanged from the additive early noise model (Eq 2.3.3). The caveat to this, however, is
717 that d' is no longer a valid measure of sensitivity when internal noise is multiplicative. Thus, recall
718 that for a single parameter, d' , to fully constrain sensitivity, it is necessary to assume that the signal
719 and noise distributions have equal variance^{43,93} (see Eq 1.1.3). Equal variance cannot be the case
720 with multiplicative noise, since, tautologically, the signal magnitude will be greater in the 'signal'
721 condition than in the 'noise' condition ($\gamma^S > \gamma^0$). Instead, recalling Eq (1.1.2), sensitivity must
722 therefore be estimated using the more general sensitivity measure, d^a :

$$d_i^a = \frac{\mu_{signal} - 0}{\sqrt{\frac{1}{2}(\sigma_{signal}^2 + \sigma_{noise}^2)}} \neq \frac{\mu_{signal}}{\sqrt{\frac{1}{2}(\sigma^2 + \sigma^2)}}. \quad (2.6.2)$$

723 As discussed extensively elsewhere^{42,43,93}, d^a can be estimated empirically, using the equation:

$$d^a = \frac{\sqrt{2}a}{\sqrt{1 + b^2}}. \quad (2.6.3)$$

724 where a and b are, respectively, the intercept and slope of an observed Receiver Operator

725 Characteristic^{42,43,94} (ROC; see **Fig 13A**). Failure to use the appropriate measure of sensitivity may
726 lead to spurious variations in apparent sensitivity, depending on where precisely the observer
727 places their criterion.

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

$$d_{Sum:multi}^{a\ early} = \frac{\omega_1 d_1^a + \omega_2 d_2^a}{\sqrt{\omega_1^2 \sigma_1^2 + \omega_2^2 \sigma_2^2}} = \sqrt{d_1^{a2} + d_2^{a2}}. \quad (2.6.4)$$

730 With d^a , ideal weights remain proportional to the sensitivity values of the two cues, as per **Eq**
731 **(2.3.2)**. Thus, as with d' , the ideal linear-weighted sum of two equally useful cues will cause d^a to
732 improve by $\sqrt{2}$ (41%), and this benefit will diminish exponentially as one cue is made less
733 informative (**Fig 2B**). In short then, if the multiplicative noise is *early*, then the practical method of
734 computing sensitivity differs from the additive case, but predicted change in sensitivity remains
735 invariant.

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

$$R_{Sum:multi}^{late} = \begin{cases} \text{'Signal Present'} & \text{if } (S_1 + S_2 + N_L \gamma^{(S_1+S_2)}) > \lambda, \\ \text{'Signal Absent'} & \text{otherwise} \end{cases}, \quad (2.6.2)$$

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

$$d_{Sum:multi}^{a\ late} = \frac{\omega_1 d_1^a + \omega_2 d_2^a}{\frac{\sqrt{[\sigma_L \gamma^{(d_1^a + d_2^a)}]^2 + \sigma_L^2}}{2}} \neq \frac{\omega_1 d_1^a}{\sigma_L \gamma^{d_1^a}} + \frac{\omega_2 d_2^a}{\sigma_L \gamma^{d_2^a}} \quad (2.6.5)$$

$$\neq \omega_1 d_1^a + \omega_2 d_2^a,$$

743 When $\gamma = 1$, **Eq (2.6.5)** reduces to the additive late noise model, and sensitivity in the multi-cue
744 condition will equal the arithmetic sum of the individual sensitivities (**Eq 2.5.2a**). When $\gamma < 1$, the
745 integration is *supralinear* (sensitivity will increase at a rate greater than predicted by the late noise
746 model). When $\gamma > 1$, the integration is *sublinear* (sensitivity will increase at a rate less than that
747 predicted by the late noise model). Notably though, even if the gain parameter, γ , were known, the
748 unobservable DV parameters μ and σ would also have to be known in order to predict ideal
749 performance in a combined-cue model (i.e., rather than the sensitivity ratio μ/σ , indexed by d').

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

756 Traditionally, however, it has been common to assume that levels of multiplicative noise are
757 negligibly small. In part, this reflects mathematical convenience. However, there are also empirical
758 arguments both for and against the presence of multiplicative noise. In favor of multiplicative noise
759 being present, there is converging evidence from psychophysics^{11,95,96} and neurophysiology⁹⁷⁻¹⁰¹.

760 For example, Weber’s law^{ix} is often taken to indicate the presence of a limiting noise source that
 761 increases with stimulus strength. Similarly, single cells are often shown to exhibit Poisson-like
 762 processes, with spike-rate variability increasing as a function of mean firing rate. *Prima facie*, these
 763 appear good evidence of multiplicative noise in decision-making. However, in the context of cue-
 764 combination, the arguments are misleading. Weber’s law is typically observed over large (order of
 765 magnitude) stimulus ranges. In contrast, performance around threshold, which we are principally
 766 interested in, may be approximately, locally linear^x. Furthermore, decisions are likely to be driven
 767 by populations of neurons, rather than individual cells. In that case, it follows from the Central
 768 Limit Theorem that the total product of the individual, Poisson-like process will rapidly converge
 769 to a Gaussian as the number of neurons increases⁴.

770 In contrast, the evidence that multiplicative noise is of relatively minor concern is more robust. For
 771 example, additivity can be assessed empirically by constructing ROC curves. As shown in **Figure**
 772 **13A**, when integration is linear (additive noise), the curve will have unit slope when plotted on
 773 Gaussian-transformed coordinates. In contrast, sublinear and supralinear conditions produce
 774 shallower or greater slopes, respectively. Such curves do not tend to be measured in studies of cue
 775 integration, but have been studied on a range of more basic sensory judgment tasks, where they
 776 tend to have approximately unit slope (albeit with some substantial variation, e.g., 0.5 to 2.0¹⁰²).
 777 Similarly, the effects of multiplicative noise should also be apparent in psychometric functions¹⁰³
 778 (**Fig 13B**). For example, a sublinear process (greater noise in the combined condition) would
 779 result in an asymmetric function with a suppressed asymptote at high stimulus magnitudes
 780 (‘saturation’; purple triangles). This is not typically observed in the types of tasks used in cue-
 781 integration studies, again allowing us to discount a substantial source of multiplicative noise.

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

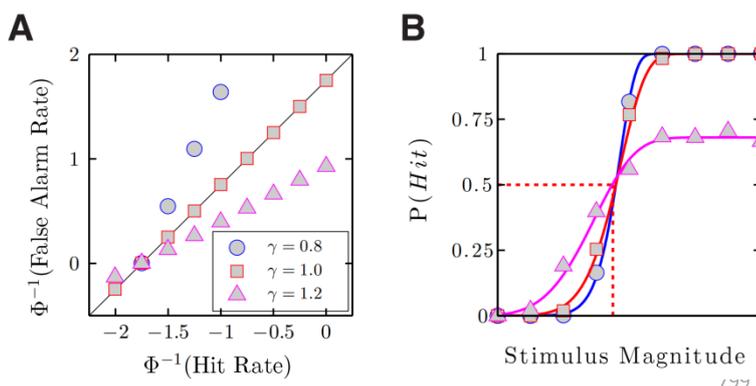


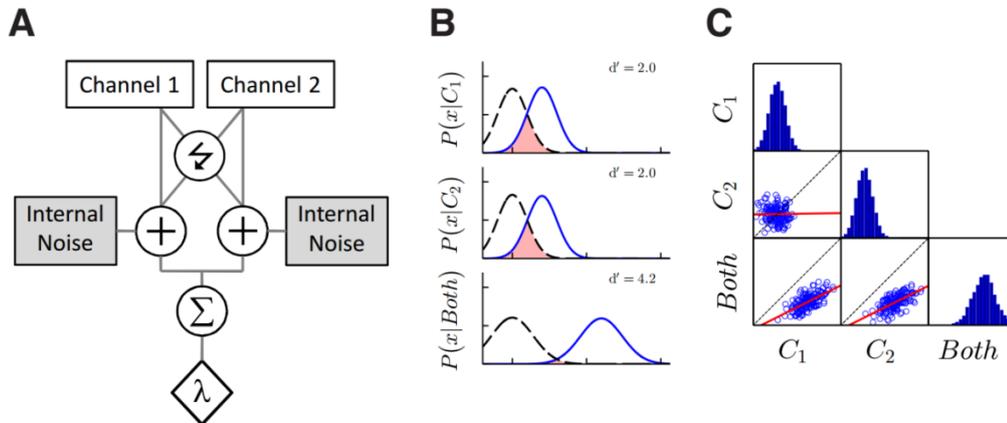
Fig 13. Schematic tests of multiplicative noise, using **(A)** ROC analysis, and **(B)** Psychometric analysis. When internal noise is sublinear (purple triangles), internal noise is proportionally greater at higher stimulus levels, resulting in a smaller improvement in sensitivity than in the linear condition (red square). The supralinear (blue circle) condition results in the opposite asymmetry (relatively less noise at higher signal levels). As detailed elsewhere^{42,43}, an ROC curve

800 constructed by measuring hit rate and false alarm rates as the criterion, λ (see **Fig 1**), is systematically shifted
 801 (e.g., by varying the benefit/cost of Hits/Misses). The psychometric function shown here plots proportion of hits,
 802 which is expected to increase monotonically as a function of stimulus magnitude. See body text for details.

^{ix} Weber’s law states that the Just Noticeable Difference between two stimuli is proportional to their magnitude

^x 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^[146]: “*Theoretical and Experimental Analysis of Second Choices*”). However, as discussed by Solomon (2007)¹⁴⁷, 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.

803 §2.7 Superadditivity



804 **Fig 14.** Superadditive cue-combination schema; same format as Fig 3. Some putative mechanism is assumed to
805 increase the internal response to each cue, when both are observed simultaneously (Panel A). As a result,
806 sensitivity in the combined condition can increase by more than the sum of the individual cue conditions: $d'_{1+2} >$
807 $(d'_1 + d'_2)$.
808

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

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

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

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

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

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

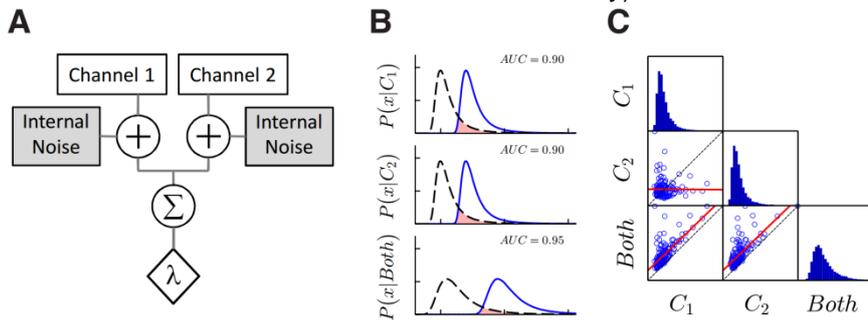
819 How plausible is superadditivity? The notion of superadditivity is made credible, *prima facie*, by
820 studies of physiology. For example, neurons in the Superior Colliculus^{3,104,105} and Superior
821 Temporal Sulcus^{106,107} have been shown to fire more than twice as often when presented with
822 corresponding information from two modalities (e.g., sight and sound), versus either in isolation.
823 This has been argued to reflect the linear summation of membrane potentials¹⁰⁸ followed by a
824 static nonlinearity (threshold) in spike generation¹⁰⁹.

825 However, as with multiplicative noise, it may be misleading to draw inferences regarding system-
826 level decision-making from the dynamics of single-cell, and studies at the behavioral level have
827 seldom reported superadditivity. Moreover, there is doubt over the extent to which, even in
828 physiology, superadditivity generalizes beyond situations where both inputs are very weak /
829 subthreshold¹¹⁰. Nonetheless, the possibility of superadditivity cannot be ruled out completely,
830 and so caution is advised when attempting to infer decision efficiency from observations of
831 sensitivity alone. For example, based on sensitivity alone, it is impossible to distinguish an early-
832 noise linear-summation observer with ideal weights, from a superadditive observer with

833 suboptimal weights.

834

§2.8 Non-Gaussian noise models ($Sum_{early/late}^{non-norm}$)



835

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, ν , increases). Due to this asymmetry, the non-parametric ‘Area Under Curve’ metric is a more appropriate measure of sensitivity than d' (Panel B).

836

837

838

839

840

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 Figure 15B.

841

842

843

844

What effect do deviations from normality have on expected cue-combination behavior? The answer is: surprisingly little (see Ref~[81]). In theory, deviations from normality are a substantial complication. For example, as shown in Figure 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.

845

846

847

848

849

850

851

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 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.

852

853

854

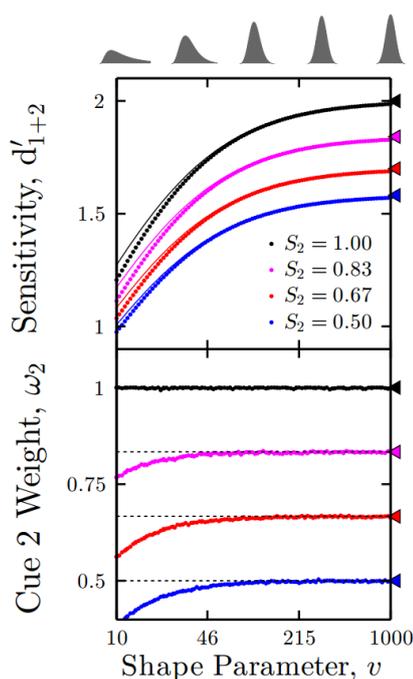


Fig 16. Means of Monte Carlo simulations, showing how sensitivity (top) and ideal weights (bottom) vary when the skew of the internal noise distribution is varied, and also as the utility of the second cue (different colored curves) is varied. Each data point (100 points per curve, per level of skew) was estimated by mean-averaging over 100 independent simulations, using 2M trials per simulation (80B trials total). The signal magnitude, μ , of Cue₁ was fixed at 1.0 (i.e., $d'_1 \approx 1.0$, though, as discussed in §2.6, d' is not technically a valid measure of sensitivity when internal noise is not additive and Gaussian). For Cue₂, μ varied from 0.5—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).

§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 §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 I). 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^{111,112}, but to the max-DV and *N-look* (probability summation) models²⁶, 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 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 I). 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

920 performance (i.e., only a minority of which involve giving zero-weight to any single cue).
921 Thus, changes in sensitivity cannot reveal the number of cues an observer has used to make
922 their decision (although, as some authors have noted^{113,114}, sensitivity can be quantified in
923 terms of the 'effective' number of cues that would be required to produce the observed
924 performance, assuming binary --- 0.0 or 1.0 --- weights)

925 **§3.2 Why multi-cue sensitivity may differ between observers**

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

938 A third alternative is that the two observers do *not* differ in sensitivity, and that the apparent
939 difference is a statistical artifact. This may happen for either of two reasons. Firstly, when
940 assumptions of either normality or additivity are breached, then d' ceases to be an appropriate
941 measure of sensitivity. This can cause apparent sensitivity to differ spuriously between
942 tests/observers, depending on where they place their criterion. Such differences would be
943 eradicated by using an appropriate, non-parametric measure, such as d^a or AUC. Secondly, as
944 shown in **Table I**, the amount of measurement error associated with estimates of sensitive are
945 non-trivial. This can lead to Type I (false difference reported) or Type II (true difference missed)
946 errors, either of which can lead to misleading conclusions as to how observers compare in terms of
947 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., bias^{21,115} or response time^{116,117}). 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~[38]. 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.

§4.1 Relative weights

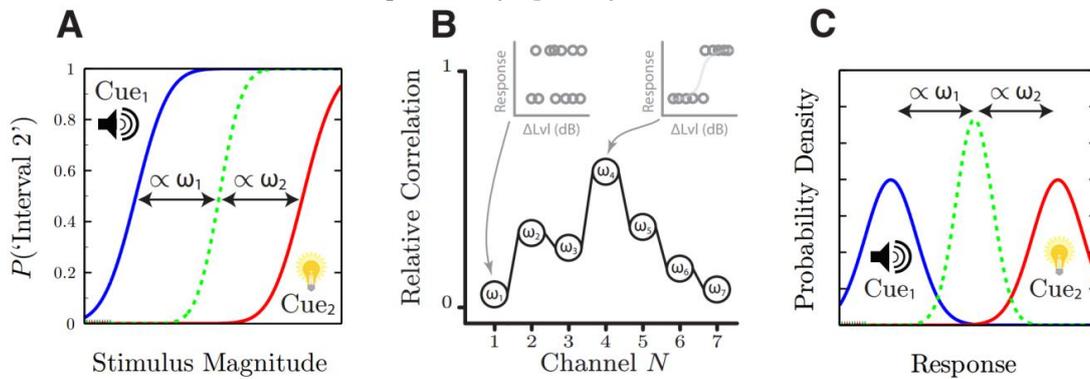
In §3.1 we saw that sensitivity alone is a relatively poor indicator of how efficiently an observer is performing, or of what their underlying decision strategy is. As a result, researchers interested in human decision-making are increasingly looking to quantify relative decision weights, ω . These indicate 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 paradigm¹¹⁸), in which case the effect is to laterally shift the psychometric function leftwards/rightwards in favor of the more weighted cue (Fig 17A). 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^{64,119-121}). 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 17B). 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)¹²².

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

994 strategy can be evidenced by making the response measure continuous, thereby allowing weights
 995 to be estimated within a single trial. For example, if asked to point at the location of a conflicting
 996 flash-beep compound, then the relative weight given to the visual and auditory cues can be
 997 observed directly from a single response. An alternating observer will produce a bimodal
 998 distribution of responses, while the observer who maintains constant weightings will produce a
 999 Gaussian distribution of responses (Fig 17C).



1000
 1001 **Fig 17** Two methods of computing relative decision weights. **(A)** Cue Conflict Paradigm. To the extent that the
 1002 Point of Subjective Equality [PSE] is shifted laterally towards the Point of Objective Equality [POE] in Cue₁, the
 1003 observer can be inferred to be relying upon (giving weight to) Cue₁. To the extent that the PSE is shifted towards
 1004 the POE in Cue₂, the observer is giving weight to Cue₂ (see **Appendix D**). **(B)** Reverse Correlation Paradigm. The
 1005 relative magnitude of correlation between the trial-by-trial variations in Cue-level and observer-response,
 1006 indicates the degree to which the observer attends to (weights) that Cue. **(C)** Continuous Cue Conflict Paradigm. A
 1007 modified version of (A), in which the dependent variable is continuously distributed, allowing weights to be
 1008 inferred from a single trial. If the decision strategy varies across trials, a Gaussian mixture model will be observed.
 1009 If the decision strategy is constant, then a Gaussian distribution will be observed, with a standard deviation
 1010 determined by the observer's sensitivity (as per the slope of the green dashed curve in A).

1011 **Why measure relative weights**

1012 Irrespective of precisely how they are measured, relative weight coefficients may be of interest for
 1013 two main reasons. The first is that, when used in combination with sensitivity measurements, they
 1014 can help to constrain the number of possible decision models. Thus, while many models may
 1015 predict observed sensitivity given *some* combination of cues (see **Table I**), generally only a small
 1016 subset of models are consistent with both a particular level of sensitivity, d' , and a particular set of
 1017 relative weights, ω (**Fig 18**). Therefore, by independently measuring both d' and ω , the underlying
 1018 decision model (and thus efficiency) can be inferred empirically, without the need to make a series
 1019 of debatable assumptions.

1020 Unfortunately, while this approach is elegant in principle, there is a practical difficulty. Namely, the
 1021 amount of measurement error typically associated with estimates of both d' and ω means that
 1022 their union may fail to adequately constrain the range of possible models (**Fig 18**). Thus, without
 1023 uncommonly precise measures of d' and ω , it remains impossible to say with certainty precisely
 1024 which model underlies performance, and therefore whether a given level of sensitivity was more
 1025 or less close to ideal. Researchers looking to ensure that they have sufficient data to distinguish
 1026 between rival hypotheses are encouraged to inspect the sampling distributions of d' given in
 1027 **Appendix B**.

1028 Furthermore, the potential for an experiment to be underpowered increases as expected benefit
 1029 decreases --- e.g., either because cues are sub-optimally weighted (**Fig 18B**), or because the
 1030 observer is relative insensitive to the second cue (**Fig 18C**; see also **Fig 2**). In either case, the
 1031 difference in predicted sensitivity is further diminished between models. In this light, it is perhaps
 1032 unsurprising that so many studies have found humans to act consistent with a 'Bayesian (early-
 1033 noise) ideal observer'²²⁻²⁸, as with only two cues it would require an acutely poor strategy, or an

1069 that the noise is correlated or multiplicative^{xi}, the ideal weights will vary in more complex ways. As
1070 discussed, some of these assumptions are more likely to be valid than others. Notably, deviations
1071 from normality will not tend to substantially affect the optimal weightings, though may affect the
1072 taking of measurements on a practical level¹²¹.

1073 In principle then, neither measurements of d' or ω alone are sufficient to estimate an observer's
1074 efficiency. Both are required to constrain the underlying model of decision making, and neither are
1075 intelligible without knowledge of the underlying model. There are, however, two exceptions. The
1076 first is when all cues are equally useful. In that case almost all decision strategies predict equal
1077 weights, so there is no need to commit to any specific model. The second occurs when external
1078 noise is introduced to the stimuli by the experimenter. Such noise is generally assumed to be
1079 exclusively early in effect^{xii}, can be specified as being independent and Gaussian, and when it is
1080 great enough in magnitude will swamp the decision-process, making any internal noise negligible.
1081 Given these assumptions, the ideal weights are guaranteed to be those predicted by the early-noise
1082 model (i.e., directly proportional to reliability, Eq 2.3.4), and so the efficiency of the weights can be
1083 meaningfully considered independent of observed sensitivity.

1084 §4.2 Response Times

1085 Throughout the present paper, we have assumed that the observer's goal is always to minimize
1086 response error. However, observers may also wish to optimize other properties of the decision
1087 making process, such as *response time*. A reduction in response time may thus constitute an
1088 entirely separate reason to attend to a second cue, and a second cue may thereby confer benefit
1089 even when accuracy/sensitivity in the single-cue conditions is near ceiling.

1090 Many authors have used response time data to make inferences regarding compound-signal
1091 decision making^{116,117,124–126}, and a full exposition of this literature is beyond the scope of the
1092 present work. However, it is worth noting that many commonalities exist between the approaches
1093 used to study response times, and the models of sensitivity in Section §2.

1094 In brief, classic response-time theory posits that noisy sensory data is accumulated over time until
1095 a criterion is reached (a 'drift-diffusion' process¹²⁷), at which point a decision is made. Given this
1096 framework, authors typically advance three alternative hypotheses, illustrated in Fig 19, for how
1097 observers respond to two cues:

- 1098 1. The observer responds based upon a single cue only. In this case, response times will be
1099 no quicker than response times in the faster single cue condition. This is equivalent to
1100 the 1-look model for response-accuracy presented in §2.1.
- 1101 2. The observer responds based upon whichever random walk reaches threshold first. In
1102 this case, response times will, by chance, tend to be faster than in either single-cue
1103 condition, but should never exceed the sum of both (Miller's Bound; Fig 19, black line).
1104 This represents probability summation of response times, and is equivalent to the 2-
1105 look accuracy model presented in §2.2.
- 1106 3. Evidence from both cues 'coactivate' some central decision-making process (i.e.,
1107 evidence from both is accumulated together within a single random walk). In this case,

^{xi} or, more generally, to the extent that the system is non-linear through superadditivity, or in some other way

^{xii} 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)¹⁴⁸. Evidence for strategy changes (for some, but not all types of noise) can be found in works by Allard and colleagues^{149–151}.

1108 response times in the two-cue condition may reduce by more than the sum of the single-
1109 cue conditions. (Typically, no parametric predictions are made, though are possible if
1110 assumptions are made regarding the underlying error distribution and sampling rate.)
1111 In its non-parametric form, this hypothesis is a superset of all the models presented in
1112 §2.3 onwards.

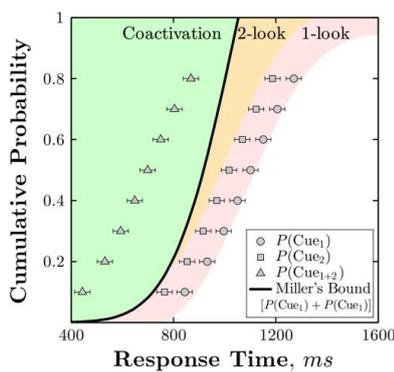


Fig 19. Example response-time analysis. Markers show simulated single-cue performance (squares, circles), and two-cue performance (triangles). Shaded regions show predicted two-cue performance for three models of decision making. The solid black line represents Miller's Bound (or: 'the Race Model Bound'), which is defined as the sum of the two single-cue cumulative probability density functions (for details, see Refs~[128,129]). 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.

1125 In short, decision-models of response times have typically attempted to address the same basic
 1126 questions as decision-models of sensitivity, and many studies of response times neatly parallel
 1127 those of sensitivity (see, for example, Ref~[26] for a study of sensitivity that explicitly attempts to
 1128 disambiguate between the three hypotheses shown graphically in **Figure 19**). In doing so, studies
 1129 of response times also suffer from the same theoretical caveats (e.g., failure to exceed Miller's
 1130 Bound does not rule out more powerful 'coactivation' models, unless one assumes independent
 1131 noise and unlimited processing capacity), as well as many of the same practical limitations (e.g.,
 1132 relatively small differences in effect size between models). As a dependent measure, response
 1133 times are also complicated by their relative volatility. Thus, researchers often find it necessary to
 1134 exclude as outliers empirical estimates below and/or above an arbitrary threshold. Such trimming
 1135 can in turn introduce non-trivial artefacts into the data, which, if not corrected for, can lead to
 1136 fallacious conclusions¹³⁰. For practical reasons, many authors therefore to prefer measures of
 1137 sensitivity over response times.

1138 Recently, however, there has been interest in response time data, for two, related reasons. Firstly,
 1139 because many sensory decisions intuitively represent a *trade-off* between speed and accuracy. This
 1140 is most obviously the case in situations where the stimulus duration is under the participant's
 1141 control, but trade-offs may even occur with brief stimulus presentations, given that evidence
 1142 continues to be accumulated even after the physical input is removed (e.g., as evidenced by
 1143 physiological recordings¹³¹, and by backward masking¹³²). Thus, an observer may choose to spend
 1144 less time gathering evidence at the cost of decreased accuracy, or may deliberate longer to improve
 1145 accuracy. A number of models have been proposed to account for speed/accuracy trade-off (see
 1146 Refs~[133-135]), and these can, for example, reveal highly efficient decision-process even in
 1147 situations where response accuracy in the multi-cue condition did not improve significantly¹³⁵.

1148 Secondly, response times may be of interest because they can provide additional information to
 1149 complement sensitivity metrics. Thus, a recurring conclusion of the present work is that it can be
 1150 remarkably difficult to distinguish between competing models of decision making, based on
 1151 accuracy alone (e.g., see final column of **Table I**). Often, however, differences in accuracy are
 1152 comorbid with differences in response time. For example, age-related sensory decline is
 1153 characterized by both slower and less accurate responses. Combining both response times and
 1154 accuracy into a single unified measure – such as in the non-parametric 'integration coefficient'
 1155 proposed by Townsend and Altieri (2012)^{136,137} – might therefore provide a more sensitive test of
 1156 cue-combination ability. Currently, however, the data are lacking to conclusively validate such a
 1157 compound measure.

1158 Acknowledgments

1159 The author thanks Joshua Solomon for comments on an earlier draft. This work was supported by
 1160 the Special Trustees of Moorfields Eye Hospital, and by the NIHR Biomedical Research Centre at
 1161 Moorfields Eye Hospital and the UCL Institute of Ophthalmology.

References

1163

- 1164 1. Sahani, M. & Whiteley, L. in *Sens. cue Integr.* (Trommershauser, J., Kording, K. & Landy, M. S.) 82–100
1165 (Oxford University Press, 2011). doi:10.1093/acprof:oso/9780195387247.003.0001
- 1166 2. DeVries, S. H. Bipolar cells use kainate and AMPA receptors to filter visual information into separate
1167 channels. *Neuron* **28**, 847–856 (2000).
- 1168 3. Stein, B. E. & Stanford, T. R. Multisensory integration: current issues from the perspective of the single
1169 neuron. *Nat. Rev. Neurosci.* **9**, 255–266 (2008).
- 1170 4. Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat.*
1171 *Neurosci.* **9**, 1432–1438 (2006).
- 1172 5. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat.*
1173 *Rev. Neurosci.* **7**, 358–366 (2006).
- 1174 6. Pouget, A., Deneve, S. & Duhamel, J.-R. A computational perspective on the neural basis of multisensory
1175 spatial representations. *Nat. Rev. Neurosci.* **3**, 741–747 (2002).
- 1176 7. Deneve, S., Latham, P. E. & Pouget, A. Efficient computation and cue integration with noisy population
1177 codes. *Nat. Neurosci.* **4**, 826–831 (2001).
- 1178 8. Macpherson, E. A. & Middlebrooks, J. C. Listener weighting of cues for lateral angle: the duplex theory of
1179 sound localization revisited. *J. Acoust. Soc. Am.* **111**, 2219–2236 (2002).
- 1180 9. Hillis, J. M., Watt, S. J., Landy, M. S. & Banks, M. S. Slant from texture and disparity cues: Optimal cue
1181 combination. *J. Vis.* **4**, 967–992 (2004).
- 1182 10. Jones, P. R., Shub, D. E., Moore, D. R. & Amitay, S. Reduction of internal noise in auditory perceptual
1183 learning. *J. Acoust. Soc. Am.* **133**, 970–981 (2013).
- 1184 11. Swets, J. A. Multiple observations of signals in noise. *J. Acoust. Soc. Am.* **31**, 514–521 (1959).
- 1185 12. Alexander, J. M. & Lutfi, R. A. Sample discrimination of frequency by hearing-impaired and normal-hearing
1186 listeners. *J. Acoust. Soc. Am.* **123**, 241–253 (2008).
- 1187 13. Juni, M. Z., Gureckis, T. M. & Maloney, L. T. Effective integration of serially presented stochastic cues. *J. Vis.*
1188 **12**, 1–16 (2012).
- 1189 14. Blake, R. & Fox, R. The psychophysical inquiry into binocular summation. *Percept. Psychophys.* **14**, 161–185
1190 (1973).
- 1191 15. Blake, R., Sloane, M. & Fox, R. Further developments in binocular summation. *Percept. Psychophys.* **30**, 266–
1192 276 (1981).
- 1193 16. Langhans, A. & Kohlrausch, A. Differences in auditory performance between monaural and diotic
1194 conditions. I: Masked thresholds in frozen noise. *J. Acoust. Soc. Am.* **91**, 3456–3470 (1992).
- 1195 17. Geldard, F. A. & Sherrick, C. E. Multiple cutaneous stimulation: The discrimination of vibratory patterns. *J.*
1196 *Acoust. Soc. Am.* **37**, 797–801 (1965).
- 1197 18. Fetsch, C. R., Turner, A. H., DeAngelis, G. C. & Angelaki, D. E. Dynamic reweighting of visual and vestibular
1198 cues during self-motion perception. *J. Neurosci.* **29**, 15601–15612 (2009).
- 1199 19. Zacharias, G. L. & Young, L. R. Influence of combined visual and vestibular cues on human perception and
1200 control of horizontal rotation. *Exp. Brain Res.* **41**, 159–171 (1981).
- 1201 20. Stein, B., Meredith, M., Huneycutt, W. & McDade, L. Behavioral indices of multisensory integration:
1202 orientation to visual cues is affected by auditory stimuli. *Cogn. Neurosci. J.* **1**, 12–24 (1989).
- 1203 21. Landy, M. S., Maloney, L. T., Johnston, E. B. & Young, M. Measurement and modeling of depth cue
1204 combination: In defense of weak fusion. *Vision Res.* **35**, 389–412 (1995).
- 1205 22. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion.
1206 *Nature* **415**, 429–433 (2002).
- 1207 23. Knill, D. C. & Saunders, J. A. Do humans optimally integrate stereo and texture information for judgments of
1208 surface slant? *Vision Res.* **43**, 2539–2558 (2003).
- 1209 24. Fetsch, C. R., Pouget, A., DeAngelis, G. C. & Angelaki, D. E. Neural correlates of reliability-based cue
1210 weighting during multisensory integration. *Nat. Neurosci.* **15**, 146–154 (2012).
- 1211 25. Alais, D. & Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* **14**,
1212 257–262 (2004).
- 1213 26. Nardini, M., Bedford, R. & Mareschal, D. Fusion of visual cues is not mandatory in children. *Proc. Natl. Acad.*
1214 *Sci. U. S. A.* **107**, 17041–17046 (2010).
- 1215 27. Gori, M., Del Viva, M., Sandini, G. & Burr, D. C. Young children do not integrate visual and haptic form
1216 information. *Curr. Biol.* **18**, 694–698 (2008).
- 1217 28. Nardini, M., Jones, P., Bedford, R. & Braddick, O. Development of cue integration in human navigation. *Curr.*
1218 *Biol.* **18**, 689–693 (2008).
- 1219 29. Laurienti, P. J., Burdette, J. H., Maldjian, J. A. & Wallace, M. T. Enhanced multisensory integration in older

- 1220 adults. *Neurobiol. Aging* **27**, 1155–1163 (2006).
- 1221 30. Bates, S. L. & Wolbers, T. How cognitive aging affects multisensory integration of navigational cues.
1222 *Neurobiol. Aging*
- 1223 31. Alexander, J. M. & Lutfi, R. A. Informational masking in hearing-impaired and normal-hearing listeners:
1224 Sensation level and decision weights. *J. Acoust. Soc. Am.* **116**, 2234–2247 (2004).
- 1225 32. Moro, S. S., Harris, L. R. & Steeves, J. K. E. Optimal Audiovisual Integration in People with One Eye.
1226 *Multisens. Res.* **27**, 173–188 (2014).
- 1227 33. Garcia, S. Spatial navigation with a new sensory signal: Visuomotor procesing in blind patients with a
1228 retinal prosthesis. *Prep.*
- 1229 34. Putzar, L., Gondan, M. & Röder, B. Basic multisensory functions can be acquired after congenital visual
1230 pattern deprivation in humans. *Dev. Neuropsychol.* **37**, 697–711 (2012).
- 1231 35. Landry, S. P., Guillemot, J.-P. & Champoux, F. Temporary Deafness Can Impair Multisensory Integration A
1232 Study of Cochlear-Implant Users. *Psychol. Sci.* 0956797612471142 (2013).
- 1233 36. Berg, B. G. Observer efficiency and weights in a multiple observation task. *J. Acoust. Soc. Am.* **88**, 149–158
1234 (1990).
- 1235 37. Fetsch, C. R., DeAngelis, G. C. & Angelaki, D. E. Bridging the gap between theories of sensory cue integration
1236 and the physiology of multisensory neurons. *Nat. Rev. Neurosci.* **14**, 429–442 (2013).
- 1237 38. Trommershauser, J., Kording, K. & Landy, M. S. *Sensory cue integration*. (Oxford University Press, 2011).
- 1238 39. Landy, M. S., Banks, M. S. & Knill, D. C. in *Sens. cue Integr.* (Trommershauser, J., Kording, K. & Landy, M. S.)
1239 (Oxford University Press, 2011). doi:10.1093/acprof:oso/9780195387247.003.0001
- 1240 40. Saarela, T. P. & Landy, M. S. Combination of texture and color cues in visual segmentation. *Vision Res.* **58**,
1241 59–67 (2012).
- 1242 41. Green, D. M. & Swets, J. A. in 1–455 (Krieger, 1974).
- 1243 42. Macmillan, N. A. & Creelman, C. D. in 1–495 (Lawrence Erlbaum Associates, 2005).
- 1244 43. Wickens, T. D. in 114–118 (Oxford University Press (USA), 2002).
- 1245 44. Machilsen, B. & Wagemans, J. Integration of contour and surface information in shape detection. *Vision Res.*
1246 **51**, 179–186 (2011).
- 1247 45. Persike, M. & Meinhardt, G. Cue combination anisotropies in contour integration: The role of lower spatial
1248 frequencies. *J. Vis.* **15**, 17 (2015).
- 1249 46. Ban, H., Preston, T. J., Meeson, A. & Welchman, A. E. The integration of motion and disparity cues to depth
1250 in dorsal visual cortex. *Nat. Neurosci.* **15**, 636–643 (2012).
- 1251 47. Dekker, T. M. *et al.* Late development of cue integration is linked to sensory fusion in cortex. *Curr. Biol.* **25**,
1252 2856–2861 (2015).
- 1253 48. Gu, Y., Angelaki, D. E. & DeAngelis, G. C. Neural correlates of multisensory cue integration in macaque
1254 MSTd. *Nat. Neurosci.* **11**, 1201–1210 (2008).
- 1255 49. Ernst, M. O. in *Hum. body Percept. from Insid. out* (Knoblich, G.) (Oxford University Press, 2006).
- 1256 50. Javel, E. & Viemeister, N. F. Stochastic properties of cat auditory nerve responses to electric and acoustic
1257 stimuli and application to intensity discrimination. *J. Acoust. Soc. Am.* **107**, 908–921 (2000).
- 1258 51. Soderquist, D. R. & Lindsey, J. W. Physiological noise as a low-frequency masker: The cardiac cycle. *J.*
1259 *Acoust. Soc. Am.* **50**, 143 (1971).
- 1260 52. Tanner, J. W. P. & Birdsall, T. G. Definitions of d'_{\prime} and n as psychophysical measures. *J. Acoust. Soc.*
1261 *Am.* **30**, 922–928 (1958).
- 1262 53. Treisman, M. Combining information: Probability summation and probability averaging in detection and
1263 discrimination. *Psychol. Methods* **3**, 252–265 (1998).
- 1264 54. Gureckis, T. & Love, B. Learning in noise: Dynamic decision-making in a variable environment. *J. Math.*
1265 *Psychol.* **53**, 180–193 (2009).
- 1266 55. Macmillan, N. A. & Creelman, C. D. Response bias: Characteristics of detection theory, threshold theory, and
1267 “nonparametric” indexes. *Psychol. Bull.* **107**, 401–413 (1990).
- 1268 56. Nadarajah, S. & Kotz, S. Exact distribution of the max/min of two gaussian random variables. *Very Large*
1269 *Scale Integr. Syst. IEEE Trans.* **16**, 210–212 (2008).
- 1270 57. Lavie, N., Hirst, A., de Fockert, J. W. & Viding, E. Load theory of selective attention and cognitive control. *J.*
1271 *Exp. Psychol. Gen.* **133**, 339–354 (2004).
- 1272 58. Repovš, G. & Baddeley, A. The multi-component model of working memory: Explorations in experimental
1273 cognitive psychology. *Neuroscience* **139**, 5–21 (2006).
- 1274 59. Luciana, M. & Nelson, C. A. The functional emergence of prefrontally-guided working memory systems in
1275 four-to eight-year-old children. *Neuropsychologia* **36**, 273–293 (1998).
- 1276 60. Jones, P. R., Moore, D. R. & Amitay, S. Development of auditory selective attention: Why children struggle to
1277 hear in noisy environments. *Dev. Psychol.* (2014).

- 1278 61. Hasher, L. & Zacks, R. T. Working memory, comprehension, and aging: A review and a new view. *Psychol.*
1279 *Learn. Motiv.* **22**, 193–225 (1988).
- 1280 62. Miller, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing
1281 information. *Psychol. Rev.* **63**, 81–97 (1956).
- 1282 63. Cowan, N. The magical mystery four how is working memory capacity limited, and why? *Curr. Dir. Psychol.*
1283 *Sci.* **19**, 51–57 (2010).
- 1284 64. Gorea, A., Belkoura, S. & Solomon, J. A. Summary statistics for size over space and time. *J. Vis.* **14**, 22 (2014).
- 1285 65. Dakin, S. C. Information limit on the spatial integration of local orientation signals. *JOSA A* **18**, 1016–1026
1286 (2001).
- 1287 66. Solomon, J. A. Visual discrimination of orientation statistics in crowded and uncrowded arrays. *J. Vis.* **10**,
1288 19 (2010).
- 1289 67. Solomon, J. A., Morgan, M. & Chubb, C. Efficiencies for the statistics of size discrimination. *J. Vis.* **11**, 13
1290 (2011).
- 1291 68. Abbott, L. F. Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Res. Bull.* **50**,
1292 303–304 (1999).
- 1293 69. Angelaki, D. E., Gu, Y. & DeAngelis, G. C. Multisensory integration: psychophysics, neurophysiology, and
1294 computation. *Curr. Opin. Neurobiol.* **19**, 452–458 (2009).
- 1295 70. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
- 1296 71. Yuille, A. L. & Buelhoff, H. H. in *Bayesian Approaches to Percept.* (Knill, D. & Richards, W.) (Cambridge
1297 University Press, 1996).
- 1298 72. Jacobs, R. A. What determines visual cue reliability? *Trends Cogn. Sci.* **6**, 345–350 (2002).
- 1299 73. Pouget, A., Dayan, P. & Zemel, R. S. Inference and computation with population codes. *Annu. Rev. Neurosci.*
1300 **26**, 381–410 (2003).
- 1301 74. Knill, D. C. & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation.
1302 *TRENDS Neurosci.* **27**, 712–719 (2004).
- 1303 75. Li, R. W., Levi, D. M. & Klein, S. A. Perceptual learning improves efficiency by re-tuning the
1304 decision 'template' for position discrimination. *Nat. Neurosci.* **7**, 178–183 (2004).
- 1305 76. Jones, P. R., Moore, D. R., Shub, D. & Amitay, S. Learning to detect a tone in unpredictable noise.
- 1306 77. Gold, J. M., Sekuler, A. B. & Bennett, P. J. Characterizing perceptual learning with external noise. *Cogn. Sci.*
1307 **28**, 167–207 (2004).
- 1308 78. Sweeny, T. D. & Whitney, D. Perceiving Crowd Attention Ensemble Perception of a Crowd's Gaze. *Psychol.*
1309 *Sci.* **25**, 1903–1913 (2014).
- 1310 79. Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity
1311 with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
- 1312 80. Salinas, E. & Sejnowski, T. J. Correlated neuronal activity and the flow of neural information. *Nat. Rev.*
1313 *Neurosci.* **2**, 539–550 (2001).
- 1314 81. Oruç, I., Maloney, L. T. & Landy, M. S. Weighted linear cue combination with possibly correlated error.
1315 *Vision Res.* **43**, 2451–2468 (2003).
- 1316 82. Hillis, J. M., Ernst, M. O., Banks, M. S. & Landy, M. S. Combining sensory information: mandatory fusion
1317 within, but not between, senses. *Science (80-.)*. **298**, 1627–1630 (2002).
- 1318 83. Tickle, H., Speekenbrink, M., Tsetsos, K., Michael, E. & Summerfield, C. Near-optimal integration of
1319 magnitude in the human parietal cortex. *J. Cogn. Neurosci.* **Epub**, 1–15 (2016).
- 1320 84. Kubovy, M. & Healy, A. F. The decision rule in probabilistic categorization: What it is and how it is learned.
1321 *J. Exp. Psychol. Gen.* **106**, 427–446 (1977).
- 1322 85. Jones, P. R., Shub, D. E., Moore, D. R. & Amitay, S. The role of response bias in perceptual learning. *J. Exp.*
1323 *Psychol. Learn. Mem. Cogn.* **in press**, in press
- 1324 86. Field, G. D. & Rieke, F. Mechanisms regulating variability of the single photon responses of mammalian rod
1325 photoreceptors. *Neuron* **35**, 733–747 (2002).
- 1326 87. Jaramillo, F. & Wiesenfeld, K. Mechano-electrical transduction assisted by Brownian motion: a role for noise
1327 in the auditory system. *Nat. Neurosci.* **1**, 384–388 (1998).
- 1328 88. Wright, B. A. & Fitzgerald, M. B. Different patterns of human discrimination learning for two interaural
1329 cues to sound-source location. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12307–12312 (2001).
- 1330 89. Fine, I. & Jacobs, R. A. Comparing perceptual learning across tasks: A review. *J. Vis.* **2**, 190–203 (2002).
- 1331 90. Battaglia, P. W., Jacobs, R. A. & Aslin, R. N. Bayesian integration of visual and auditory signals for spatial
1332 localization. *JOSA A* **20**, 1391–1397 (2003).
- 1333 91. van Beers, R. J., Wolpert, D. M. & Haggard, P. When feeling is more important than seeing in sensorimotor
1334 adaptation. *Curr. Biol.* **12**, 834–837 (2002).
- 1335 92. Redmond, T., Zlatkova, M. B., Vassilev, A., Garway-Heath, D. F. & Anderson, R. S. Changes in Ricco's Area

- with background luminance in the S-Cone pathway. *Optom. Vis. Sci.* **90**, 66–74 (2013).
- 1337 93. Macmillan, N. A. & Creelman, C. D. Triangles in ROC space: History and theory of “nonparametric”
1338 measures of sensitivity and response bias. *Psychon. Bull. Rev.* **3**, 164–170 (1996).
- 1339 94. Swets, J. A. The relative operating characteristic in psychology. *Science (80-.)*. **182**, 990–1000 (1973).
- 1340 95. McGill, W. J. & Goldberg, J. P. A study of the near-miss involving Weber’s law and pure-tone intensity
1341 discrimination. *Attention, Perception, Psychophys.* **4**, 105–109 (1968).
- 1342 96. Watson, C. S., Rilling, M. E. & Bourbon, W. T. Receiver-operating characteristics determined by a
1343 mechanical analog to the rating scale. *J. Acoust. Soc. Am.* **36**, 283–288 (1964).
- 1344 97. Young, E. D. & Barta, P. E. Rate responses of auditory nerve fibers to tones in noise near masked threshold.
1345 *J. Acoust. Soc. Am.* **79**, 426–442 (1986).
- 1346 98. Teich, M. C. & Khanna, S. M. Pulse-number distribution for the neural spike train in the cat’s auditory nerve.
1347 *J. Acoust. Soc. Am.* **77**, 1110–1128 (1985).
- 1348 99. Winter, I. M. & Palmer, A. R. Intensity coding in low-frequency auditory-nerve fibers of the guinea pig. *J.*
1349 *Acoust. Soc. Am.* **90**, 1958–1967 (1991).
- 1350 100. Goris, R. L. T., Movshon, J. A. & Simoncelli, E. P. Partitioning neuronal variability. *Nat. Neurosci.* **17**, 858–865
1351 (2014).
- 1352 101. Tolhurst, D. J., Movshon, J. A. & Thompson, I. D. The dependence of response amplitude and variance of cat
1353 visual cortical neurones on stimulus contrast. *Exp. brain Res.* **41**, 414–419 (1981).
- 1354 102. Swets, J. A. Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychol. Bull.*
1355 **99**, 100–117 (1986).
- 1356 103. Tyler, C. W. & Chen, C.-C. Signal detection theory in the 2AFC paradigm: attention, channel uncertainty and
1357 probability summation. *Vision Res.* **40**, 3121–3144 (2000).
- 1358 104. Stanford, T. R., Quessy, S. & Stein, B. E. Evaluating the operations underlying multisensory integration in
1359 the cat superior colliculus. *J. Neurosci.* **25**, 6499–6508 (2005).
- 1360 105. Perrault Jr, T. J., Vaughan, J. W., Stein, B. E. & Wallace, M. T. Superior colliculus neurons use distinct
1361 operational modes in the integration of multisensory stimuli. *J. Neurophysiol.* **93**, 2575–2586 (2005).
- 1362 106. Stevenson, R. A. & James, T. W. Audiovisual integration in human superior temporal sulcus: inverse
1363 effectiveness and the neural processing of speech and object recognition. *Neuroimage* **44**, 1210–1223
1364 (2009).
- 1365 107. Werner, S. & Noppeney, U. Superadditive Responses in Superior Temporal Sulcus Predict Audiovisual
1366 Benefits in Object Categorization. *Cereb. Cortex* **20**, 1829–1842 (2010).
- 1367 108. Skalióra, I., Doubell, T. P., Holmes, N. P., Nodal, F. R. & King, A. J. Functional topography of converging visual
1368 and auditory inputs to neurons in the rat superior colliculus. *J. Neurophysiol.* **92**, 2933–2946 (2004).
- 1369 109. Holmes, N. P. & Spence, C. Multisensory integration: space, time and superadditivity. *Curr. Biol.* **15**, R762–
1370 R764 (2005).
- 1371 110. Morgan, M. L., DeAngelis, G. C. & Angelaki, D. E. Multisensory integration in macaque visual cortex depends
1372 on cue reliability. *Neuron* **59**, 662–673 (2008).
- 1373 111. Petrini, K., Remark, A., Smith, L. & Nardini, M. When vision is not an option: children’s integration of
1374 auditory and haptic information is suboptimal. *Dev. Sci.* **17**, 376–387 (2014).
- 1375 112. Weisswange, T. H., Rothkopf, C. A., Rodemann, T. & Triesch, J. Bayesian cue integration as a developmental
1376 outcome of reward mediated learning. *PLoS One* **6**, e21575 (2011).
- 1377 113. Dakin, S. C., Mareschal, I. & Bex, P. J. Local and global limitations on direction integration assessed using
1378 equivalent noise analysis. *Vision Res.* **45**, 3027–3049 (2005).
- 1379 114. Manning, C., Dakin, S. C., Tibber, M. S. & Pellicano, E. Averaging, not internal noise, limits the development
1380 of coherent motion processing. *Dev. Cogn. Neurosci.* **in press**, to appear (2014).
- 1381 115. Scarfe, P. & Hibbard, P. B. Statistically optimal integration of biased sensory estimates. *J. Vis.* **11**, 12 (2011).
- 1382 116. Hershenson, M. Reaction time as a measure of intersensory facilitation. *J. Exp. Psychol.* **63**, 289–293 (1962).
- 1383 117. Grice, G. R., Canham, L. & Boroughs, J. M. Combination rule for redundant information in reaction time
1384 tasks with divided attention. *Percept. Psychophys.* **35**, 451–463 (1984).
- 1385 118. De Gelder, B. & Bertelson, P. Multisensory integration, perception and ecological validity. *Trends Cogn. Sci.*
1386 **7**, 460–467 (2003).
- 1387 119. Lutfi, R. A. Correlation coefficients and correlation ratios as estimates of observer weights in multiple-
1388 observation tasks. *J. Acoust. Soc. Am.* **97**, 1333–1334 (1995).
- 1389 120. Dai, H. & Micheyl, C. Psychophysical reverse correlation with multiple response alternatives. *J. Exp. Psychol.*
1390 *Hum. Percept. Perform.* **36**, 976–993 (2010).
- 1391 121. Richards, V. M. & Zhu, S. Relative estimates of combination weights, decision criteria, and internal noise
1392 based on correlation coefficients. *J. Acoust. Soc. Am.* **95**, 423–434 (1994).
- 1393 122. Roach, N. W., Heron, J. & McGraw, P. V. Resolving multisensory conflict: a strategy for balancing the costs

- and benefits of audio-visual integration. *Proc. R. Soc. B Biol. Sci.* **273**, 2159–2168 (2006).
123. Jones, P. R., Moore, D. R., Shub, D. E. & Amitay, S. Learning to detect a tone in unpredictable noise. *J. Acoust. Soc. Am.* **135**, EL128–EL133 (2014).
124. Hugenschmidt, C. E., Hayasaka, S., Peiffer, A. M. & Laurienti, P. J. Applying capacity analyses to psychophysical evaluation of multisensory interactions. *Inf. Fusion* **11**, 12–20 (2010).
125. Schröger, E. & Widmann, A. Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology* **35**, 755–759 (1998).
126. Otto, T. U. & Mamassian, P. Noise and correlations in parallel perceptual decision making. *Curr. Biol.* **22**, 1391–1396 (2012).
127. Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).
128. Miller, J. Divided attention: Evidence for coactivation with redundant signals. *Cogn. Psychol.* **14**, 247–279 (1982).
129. Ulrich, R., Miller, J. & Schröter, H. Testing the race model inequality: an algorithm and computer programs. *Behav. Res. Methods* **39**, 291–302 (2007).
130. Rach, S., Diederich, A., Steenken, R. & Colonius, H. The race model inequality for censored reaction time distributions. *Attention, Perception, Psychophys.* **72**, 839–847 (2010).
131. Schall, J. D. Neural basis of deciding, choosing and acting. *Nat. Rev. Neurosci.* **2**, 33–42 (2001).
132. Breitmeyer, B. G. & Ogmen, H. Recent models and findings in visual backward masking: A comparison, review, and update. *Percept. Psychophys.* **62**, 1572–1595 (2000).
133. Ratcliff, R. & Smith, P. L. A comparison of sequential sampling models for two-choice reaction time. *Psychol. Rev.* **111**, 333–367 (2004).
134. Palmer, J., Huk, A. C. & Shadlen, M. N. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vis.* **5**, 1 (2005).
135. Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E. & Pouget, A. Optimal multisensory decision-making in a reaction-time task. *Elife* **3**, e03005 (2014).
136. Townsend, J. T. & Altieri, N. An Accuracy--Response Time Capacity Assessment Function that Measures Performance against Standard Parallel Predictions. *Psychol. Rev.* **119**, 500–516 (2012).
137. Altieri, N., Townsend, J. T. & Wenger, M. J. A measure for assessing the effects of audiovisual speech integration. *Behav. Res. Methods* **46**, 406–415 (2014).
138. Miller, J. The sampling distribution of d' . *Attention, Perception, Psychophys.* **58**, 65–72 (1996).
139. Jones, P. R., Kalwarowsky, S., Braddick, O. J., Atkinson, J. & Nardini, M. Optimizing the rapid measurement of detection thresholds in infants. *J. Vis.* **under revi**, under review
140. Ossmy, O. *et al.* The timescale of perceptual evidence integration can be adapted to the environment. *Curr. Biol.* **23**, 981–986 (2013).
141. Usher, M. & McClelland, J. L. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol. Rev.* **108**, 550–592 (2001).
142. Grimmett, G. R. & Stirzaker, D. R. *Probability and random processes*. (Oxford Univ. Press, 2003).
143. Shadlen, M. N. & Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
144. Ecker, A. S. *et al.* State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**, 235–248 (2014).
145. Jesteadt, W., Nizami, L. & Schairer, K. S. A measure of internal noise based on sample discrimination. *J. Acoust. Soc. Am.* **114**, 2147–2157 (2003).
146. Swets, J. A., Tanner Jr, W. P. & Birdsall, T. G. Decision processes in perception. *Psychol. Rev.* **68**, 301–340 (1961).
147. Solomon, J. A. Intrinsic uncertainty explains second responses. *Spat. Vis.* **20**, 45–60 (2007).
148. Pelli, D. G. in *Vis. Coding Effic.* (Blakemore, C.) 3–24 (Cambridge University Press, 1990).
149. Allard, R. & Cavanagh, P. Crowding in a detection task: External noise triggers change in processing strategy. *Vision Res.* **51**, 408–416 (2011).
150. Allard, R. & Cavanagh, P. Different processing strategies underlie voluntary averaging in low and high noise. *J. Vis.* **12**, 1–12 (2012).
151. Allard, R. & Faubert, J. Zero-dimensional noise is not suitable for characterizing processing properties of detection mechanisms. *J. Vis.* **13**, 25 (2013).

1450 **Appendix A: Simulated sensitivity given various cue combination strategies**

1451 **Listings 1-3** provide MATLAB code for simulating the various cue-combination strategies
1452 described in the present paper. In each case, observed performance is computed (as per with a
1453 human observer), and compared to predicted sensitivity given single-cue performance. Relative
1454 weights are set to their ideal values for each strategy, and, when making predictions, knowledge of
1455 other key parameters (e.g., amount of multiplicative gain) is assumed, as specified. **Listing 4** shows
1456 an example output.

1457 ****Listings 1—4 found a separate file: listings_all.pdf****

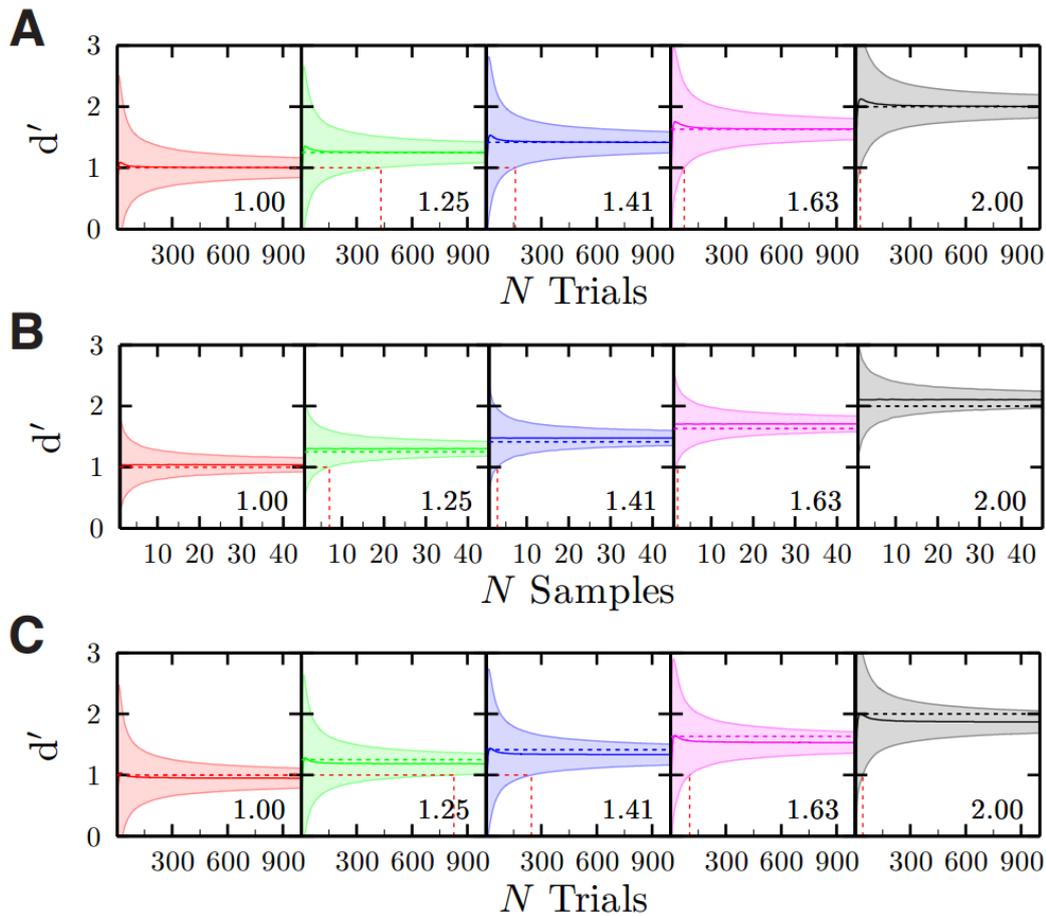
1458

1459 **Appendix B: Sample error in empirical estimates of d'**

1460 Monte Carlo simulations were used to estimate sampling distributions for different values of d' ,
1461 given varying numbers of trials (**Fig S1A**), or varying numbers of 50 trials 'blocks' (**Fig S1B**). The
1462 simulated observer was an ideal SDT observer with additive Gaussian noise, the standard
1463 deviations of which was manipulated to determine true sensitivity.

1464 With small numbers of trials, estimates of d' were highly variable and tended, on average, to be
1465 overestimated (**Fig S1A**, solid lines). Expected estimates of d' became unbiased after
1466 approximately 100 trials (see also ref [99]¹³⁸). Sampling error also decreases exponentially with
1467 small numbers of trials, such that, for example, it would require approximately 150 trials to
1468 distinguish a d' of 1.41 from a d' of 1.00, on 95% of occasions (red vertical dashed lines; i.e., perfect
1469 early-noise integration). It follows that when averaging across multiple blocks of 50 trials (**Fig**
1470 **S1B**), mean d' tends to be overestimated. However, this bias is relatively slight, and approximately
1471 150 trials (3 blocks of 50 trials) was again sufficient to distinguish a d' of 1.41 from a d' of 1.00, on
1472 95% of occasions.

1473 Note, however, that these idealized simulations are a profound simplification. For example, **Fig S1C**
1474 shows data analogous to **Fig S1A**, when a lapse rate of 5% was further introduced to simulate
1475 observers blinking, coughing, or otherwise becoming distracted on a small proportion of trials (i.e.,
1476 the simulated observer responded randomly with a probability of 5%). The result is that d' tended
1477 to be *underestimated*, such that, for example, almost 300 trials were required to reliably
1478 distinguish a d' of 1.41 from a d' of 1.00. Additional complications such as response bias and
1479 nonstationary variations in inattentiveness¹³⁹ may add further noise/inaccuracy to empirical
1480 measurements of sensitivity, and may mean that even more trials are required to differentiate
1481 models.



1482 **Fig S1.** The sampling distribution of d' , as a function of (A) Number of trials within a single observer
 1483 (half signal, half noise), (B) Number of observers performing 50 trials each (mean-averaged), (C)
 1484 Number of trials within a single observer (as per A), but with a simulated lapse rate of 5%. Solid lines
 1485 and shaded regions show the mean estimate of d' and the 95% CI ($\mu \pm 1.96 \times 1$ SD), computed from
 1486 10,000 Monte Carlo simulations. True sensitivity increases left-to-right across panels, and is given
 1487 graphically by the horizontal dashed lines, and numerically in the bottom-right of each panel. The true
 1488 value from the $d' = 1$ condition has been extended rightwards across all panels to indicate when
 1489 estimated sensitivity would be expected to differ significantly from 1 in 95% of experiments.

1490 **Appendix C: Algorithmic pseudocode**

1491 **Listing 5** shows example pseudocode for how a 1-look, 2-look, or early-noise integration strategy
1492 would be implemented. Note that the 1-look and 2-look strategies are markedly less taxing,
1493 computationally. For further discussion, see [§2.3](#).

1494 ****Listing 5 found a separate file: listings_all.pdf****

1495 **Appendix D: Psychometric simulations**

1496 **Listing 6** shows how the sensitivity estimates in **Fig 18** were computed.

1497 ****Listing 6 found a separate file: listings_all.pdf****