Journal of Mathematical Psychology 73 (2016) 117–139

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

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

Pete R. Jones

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

ARTICLE INFO

*Article history:*
Received 28 May 2015
Received in revised form
5 April 2016
Available online 27 June 2016

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

ABSTRACT

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


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

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

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

http://dx.doi.org/10.1016/j.jmp.2016.04.006

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

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

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

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

***The present paper***

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

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

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

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

## 1. Background theory

### *1.1. Using signal detection theory to measure perceptual sensitivity*

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

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

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

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

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