Published in final edited form as:

Psychol Rev. 2012 October ; 119(4): 770–788. doi:10.1037/a0029343.

Separating the Contributions of Primary and Unwanted Cues in Psychophysical Studies

Huanping Dai and Christophe Michey

University of Arizona

University of Minnesota

Abstract

A fundamental issue in the design and the interpretation of experimental studies of perception relates to the question of whether the participants in these experiments could perform the perceptual task assigned to them using another feature, or cue, than that intended by the experimenter. An approach frequently used by auditory- and visual-perception researchers to guard against this possibility involves applying random variations to the stimuli across presentations or trials, so as to make the “unwanted” cue unreliable for the participants. However, the theoretical basis of this widespread practice is not well developed. In this article, we describe a two-channel model based on general principles of psychophysical signal-detection theory, which can be used to assess the respective contributions of the unwanted cue and of the primary cue to performance or thresholds measured in perceptual discrimination experiments involving stimulus randomization. Example applications of the model to the analysis of results obtained in representative studies from the auditory- and visual- perception literature are provided. In several cases, the results of the model-based analyses indicate that the effectiveness of the randomization procedure was less than originally assumed by the authors of these studies. These findings underscore the importance of quantifying the potential influence of unwanted cues on the results of psychophysical experiments, even when stimulus randomization is used.

Keywords

psychophysics; signal detection theory; cue-combination; optimal-observer models

Correspondence concerning this manuscript should be addressed to: Huanping Dai, Department of Speech, Language, and Hearing Sciences, University of Arizona, 1131 E. 2nd Street, Tucson, Arizona, 85721. hdai@email.arizona.edu, Phone: (520) 626-5595, Fax: (520) 621-9901.

¹The perceived intensity of tones of constant physical intensity varies across frequency. This is reflected in equal-loudness contours (ELCs), which show physical levels that produce constant loudness across frequency. These variations in ELCs are due to irregularities in the transfer function of the middle ear and, possibly, to other factors (Mauermann, Long, & Kollmeier, 2004; Moore, Glasberg, Low, Cope, & Cope, 2006). According to Moore et al. (2006), even at relatively high sound levels, “irregularities of the order of 5 dB remain in the ELCs, and the level required for equal loudness can change by 5 dB when the frequency is changed by 2% or less.”

²These changes in pitch as a function of level are usually less than 1% for tones with frequencies between 1 and 2 kHz, but can be larger (up to 5%) for tones of lower or higher frequency (Stevens, 1935; Terhardt, 1974; Verschuure & van Meeteren, 1975).

³These thresholds are expressed as ratios of coordinates, I , which are themselves ratios of cone excitations, such that $I = L/(L+M)$. Specifically, the threshold, a , is defined as the ratio of the I coordinate of the test stimulus, I_A , to a reference coordinate, I_0 , such that $a = I_A/I_0$. Since the I coordinates of the two stimuli, A and B, were centered geometrically on the reference, $I_A/I_0 = I_0/I_B$, and the ratio of I values for the two stimuli, I_A/I_B , equaled a^2 .

⁴Although the 2I-2AFC and 1I-2AFC tasks are different, under the assumption that, for the 2I-2AFC tasks, the primary cue is based on the difference between the two stimuli within each trial, and that the unwanted cue is based entirely on the difference between one stimulus (of either interval) and a fixed internal “reference”, the mathematical expressions for PC_c and PC_p for these two tasks are identical.

Psychophysics is traditionally defined as the scientific study of the relation between stimulus and sensation (Gescheider, 1997). Typically, this relation is investigated by varying one physical stimulus parameter (e.g., the contrast in an image, or the intensity of a sound), and measuring how the sensation evoked by the stimulus changes. Ideally, the stimulus parameter being manipulated corresponds to a unique sensory attribute, or perceptual dimension (e.g., loudness). Unfortunately, in practice, it is not uncommon for changes in the value of a single physical parameter to induce changes in more than one aspect of sensation. For instance, a change in the frequency of a pure tone usually affects the perceived frequency (pitch) of the tone, but it can also affect its perceived intensity, or loudness (Mauermann, Long, & Kollmeier, 2004; Moore, Glasberg, Low, Cope, & Cope, 2006).¹ Conversely, changing the physical intensity of a tone usually alters its loudness, but it can also alter its pitch (Stevens, 1935; Terhardt, 1974; Verschuure & van Meeteren, 1975).² Similarly, changes in either the color or the intensity of a visual stimulus can affect both the hue and the brightness of that stimulus (e.g., Purdy, 1931, 1937; Garner & Felfoldy, 1970; Hyman & Well, 1968). The auditory- and visual-psychophysics literatures abound with examples such as these, where a change in one stimulus parameter influences more than one aspect of the sensation or percept.

Such co-dependence of multiple aspects of sensation on a single physical stimulus parameter creates a pervasive problem for psychophysicists, who often seek to study a specific perceptual attribute (e.g., pitch), while excluding others (e.g., loudness and timbre). Indeed, a key point of contention in interpretations of psychophysical data relates to whether participants actually used the stimulus aspect (feature, dimension, or cue) which the experimenter intended, or whether they use a different (unwanted) cue. For example, suppose that in an auditory frequency-discrimination experiment that seeks to measure listeners' ability to discriminate small changes in pitch, the differences in frequency between the tones induce differences, not only in pitch but also in loudness; then, it is unclear whether the results reflect pitch perception, loudness perception, or a combination of the two. The possibility that participants used another cue for task performance than that intended by the experimenter is a major concern for psychophysicists, as it may confound the interpretation of the results and lead to erroneous conclusions. Therefore, a crucially important aspect of the design and interpretation of psychophysical experiments is concerned with eliminating, or at least limiting, the possible influence of "unwanted" cues on performance.

Two approaches have traditionally been used to limit the influence of unwanted cues in studies of perception. The first approach involves adjusting stimulus parameters in an attempt to ensure that the stimuli differ only in the perceptual dimension that is of primary interest to the researcher. For instance, in a pitch discrimination experiment, the intensities of the two tones that the listener must discriminate can be adjusted separately, depending on the frequency of each tone, in such a way that the loudness of the tones remains approximately constant while the frequency difference between them varies (e.g., McDermott, Lech, Kornblum, & Irvine, 1998; Thai-Van, Michey1, Moore, & Collet, 2003). This approach may be referred to as "perceptual equalization." Unfortunately, precisely equalizing the sensations produced by physically different stimuli is often time-consuming, and not always feasible.

A second approach that can be used to limit the influence of an unwanted cue on performance involves applying random perturbations to the stimulus so as to make the unwanted cue unreliable. For instance, in an auditory pitch-discrimination experiment where the researcher is concerned that participants might be able to take advantage of loudness differences induced by frequency changes, the intensity of the stimulus can be randomly varied across presentations in an attempt to neutralize the listener's ability to take advantage

of loudness differences for correct responses in the task. This approach is traditionally referred to as “randomization” or “roving.” It has been used in numerous studies. Examples in the auditory-perception literature include studies of intensity discrimination (e.g., Berliner & Durlach, 1973; Berliner, Durlach, & Braida, 1977; Heinz, Colburn, & Carney, 2001; Dai & Green, 1992; Oxenham & Buus, 2000), pitch discrimination (e.g., Henning, 1966; Henning, Richards, & Lentz, 2005; Micheyl, Divis, Wroblewski, & Oxenham, 2010; Moore & Sek, 1998; Oxenham, Micheyl, & Keebler, 2009), detection in noise (e.g., Hall & Fernandes, 1983; Kidd, Mason, Brantley, & Owen, 1989), spectral-shape discrimination (e.g., Berg & Green, 1990; Dai, Nguyen, & Green, 1996; Drennan & Watson, 2001; Farrar et al., 1987; Kidd et al., 1989), temporal gap detection or discrimination (e.g., Forrest & Green, 1987; Formby & Muir, 1989), amplitude- or frequency-modulation perception (e.g., Furukawa & Moore, 1997; Stellmack, Viemeister, & Byrne, 2006; Strickland & Dhar, 2000), and discrimination of speech tokens (e.g., Holt & Lotto, 2006; Uchanski & Braida, 1998). In visual psychophysics, randomization has been used in studies of contrast discrimination (e.g., Dannemiller & Stephens, 1998; Nachmias, 2002; Lee and Harris, 1996), color discrimination (e.g., Danilova & Mollon, 2010), grating detection (e.g., Nachmias, 1999), binocular-correlation detection (e.g., Cormack & Riddle, 1996), shape and size discrimination (e.g., Regan & Hamstra, 1991, 1992; Morgan, 2005; Nachmias, 2008, 2011), orientation discrimination (e.g., Webster, De Valois, & Switkes, 1989), spatial-frequency discrimination (e.g., Webster et al., 1989), texture discrimination (e.g., Morgan, Chubb, & Solomon, 2008), velocity discrimination (e.g., McKee et al., 1986), letter discrimination (e.g., Murray et al., 2003), vernier discrimination (e.g., Levi, McGraw, & Klein, 2000), discrimination of time of contact, and discrimination of angular expansion (e.g., Regan & Hamstra, 1993). Randomization has also been used, albeit more rarely, in studies of tactile perception (e.g., Yuan et al., 2006).

It is worth pointing out that stimulus randomization also plays a key role in studies of “perceptual independence” (e.g., Garner, 1974, 1988; Garner & Morton, 1969; Lockhead, 1966, 1972; Melara & Marks, 1990a, 1990b; Thomas, 1999, 2003). In this context, it has been proposed that if the perception of a (first) dimension is not impaired by the application of simultaneous random variations to another (second) dimension, the two dimensions are perceptually “independent,” or “separable;” otherwise, the two dimensions are said to be “non-independent,” or “integral” (Garner & Morton, 1969). However, as pointed out by Ashby & Townsend (1986), “perceptual independence” is a very broad concept, and the phrase has been coined in various contexts and with various meanings. In this article, we take a pragmatic approach and leave aside the much-debated issues surrounding the definitions of, and distinctions between, different forms of perceptual dependence or independence, to focus on the practical problem of analyzing data obtained in psychophysical experiments in which performance could be based on two different perceptual cues, and randomization (roving) was used specifically to limit participant’s ability to use one of these cues for task performance.

Considering that the use of randomization as a way of limiting participants’ ability to rely on unintended cues for task performance is quite widespread in psychophysics, it seems especially important for researchers to have a way of answering the following questions. To what extent did the randomization effectively limit participants’ ability to use the unwanted cue for task performance? Was the randomization range sufficiently large to make any residual influence of the unwanted cue negligible? If not, to what extent do the thresholds, proportions of correct responses (PCs), or d' values measured in the experiment reflect sensitivity to the unwanted cue—in addition to, or instead of, sensitivity to the primary cue? Is it possible to separate the contributions of the primary and unwanted cues to the measured thresholds, PCs, or d' values, so that, after correction, these measures reflect sensitivity to the cue of interest only, uncontaminated by sensitivity to the unwanted cue? While empirical

measures of the influence of stimulus roving are undoubtedly important in answering these questions, such measures alone are not sufficient. A theoretical framework is needed, in the context of which empirical data can be interpreted.

This article is structured in two main parts. In the first part, *Theory*, we describe three existing models of the influence of roving on performance in psychophysical experiments; then, we present a new model, which combines advantageous features of these earlier models. In the second part of the article, *Applications*, we apply the model to data obtained in published studies of auditory and visual perception, in which roving was used in an attempt to limit the ability of participants to take advantage of unwanted cues for task performance.

Theory

Review of earlier models of the influence of roving on performance

Durlach, Braida, & Ito (1986) —Durlach et al. (1986) described a model of sensory and decision processes in the discrimination of “multi-dimensional” stimuli. In this context, “multi-dimensional” is used in a broad sense, to describe stimuli that give rise to more than one sensory observation. The example that they considered involved discriminating between tones that contain multiple spectral components, the relative levels of which are adjusted to produce different spectral shapes. In this example, the different observations correspond to the levels of the different components. They assumed that these observations, conveyed via different sensory channels, were contaminated by additive Gaussian noise, and combined linearly to form a decision variable. The result is a fairly general linear model summarized by the following equation,

$$Z = \sum_{i=1}^n w_i \Delta_i, \quad (1)$$

where Δ_i denotes the difference between the two sensory observations evoked successively by the i^{th} stimulus component on the current trial of a 2I-2AFC discrimination task, and w_i denotes the “weight” of the corresponding observation in the observer’s decision. Z is the decision variable, which the observer compares against a fixed criterion in order to select a response.

It can be shown that, for this model, discrimination performance is optimal, in a maximum-likelihood sense, when the vector of weights, $\mathbf{W} = [w_1, \dots, w_n]^T$, equals $\Sigma^{-1}E[\Delta]$, where Σ^{-1} is the inverse of the variance-covariance matrix of the noise added to the observations (van Trees, 1968), $E[\cdot]$ is the mathematical expectation, and $\Delta = [\Delta_1, \dots, \Delta_n]$. In this situation,

$$d' = \sqrt{E[\Delta^T] \Sigma^{-1} E[\Delta]}, \quad (2)$$

where the superscript, T , denotes vector transposition.

Durlach et al. (1986) considered various specific cases of this general linear decision model. In one case, which is commonly encountered in spectral-shape discrimination studies, one value is drawn at random on each stimulus presentation, and this value is added to the levels of all spectral components in the stimulus. This alters the overall level of the stimulus but leaves its spectral shape, which is by the *relative* levels of the components, unchanged. This form of roving is commonly employed by experimenters to limit listeners’ ability to take advantage of overall level differences between the two stimuli. Durlach et al. modeled this

situation by adding a Gaussian random variable with mean 0 and variance σ_R^2 to the observations, and showed that under such circumstances, sensitivity was given by

$$d' = \sqrt{d_{independent}^2 - \zeta}. \quad (3)$$

In this equation, $d_{independent}^2$ is the squared sensitivity for the case in which the noises added to the n stimulus observations are statistically independent, given by (Green & Swets, 1966),

$$d_{independent}^2 = \sum_{i=1}^n \left(\frac{\Delta_i}{\sigma_i} \right)^2. \quad (4)$$

ζ is the amount by which the addition of a common source of variability across all channels reduces the squared sensitivity below $d_{independent}^2$, and is given by,

$$\zeta = \frac{\sigma_R^2 d_{independent}^2}{1 + \sigma_R^2 \sum_{i=1}^n \sigma_i^{-2}}. \quad (5)$$

Looking at equation 5, it can be seen that unless σ_R^2 is zero, which corresponds to the situation in which no overall-level roving is applied, ζ is larger than zero. Thus, this model predicts a lower d' when overall level is roved than when it is not roved.

Durlach et al.'s (1986) article marked an important step forward in the development of normative models of the influence of stimulus randomization on discrimination performance based on multiple observations. However, this model was developed in the context of intensity discrimination and spectral-shape discrimination studies in which the primary cue (a level difference) and the unwanted cue (also a level difference) occupy the same stimulus dimension. A model of the influence of randomization of one stimulus dimension on discrimination performance based on another dimension was proposed a few years later, in a short note by Durlach et al. (1989).

Durlach, Tan, Macmillan, Rabinovitz, & Braida's (1989) —Durlach et al. (1989) described an extension of a model of the discrimination and identification of fixed unidimensional stimuli (Durlach & Braida, 1969) to situations involving random variations along a “background” dimension. In this model, sensitivity (d') is assumed to be limited by two sources of internal noise: “sensory” noise, and “memory” noise. The phrase “sensory noise” was coined to denote unavoidable noise associated with the sensory processing of the stimulus. Durlach and Braida assumed that this noise had a constant variance, denoted β^2 . Memory noise depends on the mode of processing. In the “trace” mode, the observer compares the sensory representation of the current stimulus against a trace of the previously presented stimulus in sensory (e.g., echoic) memory; the variance of this “trace” noise increases as a function of the delay between the two stimuli being compared. In the context-coding mode, the sensory representation of the current stimulus is compared against representations derived from the ensemble of stimuli presented during the experiment. These representations are presumably stored in longer-term memory than the sensory traces that are used in the trace-coding mode. An example of context-coding involves judging the loudness of a sound relative to an internal reference corresponding, for instance, to the memory of the loudest sound heard so far. Observers are more likely to rely on this coding strategy in experiments in which the set of stimuli is sufficiently small that they can be individually labeled and organized into categories. The variance of this context-coding noise

increases with the range of stimulus values used in the experiment. Specifically, if $\alpha(\cdot)$ denotes the function relating stimulus values to the expected values of the corresponding sensations, and T_{min} and T_{max} denote the smallest and largest stimulus values, then the variance of context-coding noise equals $(GR)^2$, where G is a coefficient, and $R = \alpha(T_{max}) - \alpha(T_{min})$.

In extending this model of unidimensional discrimination of fixed stimuli to situations involving random variations along another dimension, Durlach et al. (1989) suggested that any decrease in sensitivity to differences along the primary dimension (T) due to random variations along the background stimulus dimension (B) could arise, either from an increase in sensory-noise variance (β^2), or from an increase in the coefficient, G , that controls the variance of memory noise in the context coding mode.

Durlach et al. (1989) outlined various ways in which the assumptions underlying this model might be tested and the values of the different unknowns in the model might be estimated. To our knowledge, such empirical tests have not yet been published.

Green (1988) —An often-quoted model of the influence of randomization on discrimination performance was described by Green (1988). The model is considerably simpler than the models of Durlach and colleagues described in the preceding section. Green focused on the case of a uniform (i.e., rectangular) roving distribution, which is the distribution most commonly used by experimenters. He assumed that the “roving range”, i.e., the width of the rectangular distribution, was large relative to the standard deviation of the internal noise for the unwanted cue, so that the contribution of this internal noise to performance was negligible. With this assumption, Green computed the highest PC that can be achieved in a two-interval two-alternative forced-choice (2I-2AFC) discrimination experiment, by an observer who relies exclusively but optimally on an unwanted cue, given that the value of the stimulus parameter corresponding to this unwanted cue is randomized over some range on each stimulus presentation, i.e., each observation interval of the 2I task.

The situation is illustrated schematically in Figure 1. Each rectangle represents the probability density function (PDF) of a random variable, X_u , representing the stimulus parameter associated with the unwanted cue, to which random perturbations are applied. The rectangle delineated by a solid line represents the PDF of X_u conditioned on the presentation of stimulus A. The rectangle delineated by a dashed line represents the PDF of X_u conditioned on the presentation of stimulus B. On each trial of the 2I-2AFC discrimination experiment considered by Green (1988), the two stimuli, A and B, are presented in random order, and the observer’s task is to indicate whether stimulus A was presented first or second. Intuitively, the highest PC that can be achieved in this situation depends on two factors: the “roving range,” D , which corresponds to the width of the rectangular distribution, and the magnitude of the unwanted cue, Δ_u , which corresponds to the horizontal shift between the two rectangles.

Using a geometric argument, Green showed that, for this situation, the highest PC that can be achieved based on the unwanted cue, denoted PC_u , is given by,

$$\begin{aligned} PC_u &= 0.5 + \Delta_u/D - 0.5(\Delta_u/D)^2 & \text{for } \Delta_u \leq D, \\ PC_u &= 1 & \text{for } \Delta_u > D. \end{aligned} \quad (6)$$

Green’s (1988) analysis is often used by experimenters as a guideline for selecting effective randomization ranges in psychophysical studies. By completing the square in equation 6, and solving for D , it is found that,

$$D = \frac{\Delta_u}{1 - \sqrt{2(1 - PC_u)}} \quad (7)$$

Equation 7, or its equivalent for other paradigms (Dai & Kidd, 2009; Dai & Michey, 2010; Michey & Dai, 2009) can be used to determine the roving range (D) that is needed to keep PC_u under a pre-defined level, given the known—or estimated—unwanted-cue size (Δ_u). For example, suppose that in a 2I-2AFC discrimination experiment the experimenter would like to ensure that participants cannot achieve more than 70% of correct responses by relying on an unwanted cue, the size of which is estimated at 1 dB; then, using equation 7, the experimenter can determine that to keep PC_u under 0.7, the roving range, D , must be at least 4.4 times the size of the unwanted cue, Δ_u , i.e., D should be at least to 4.4 dB. With this roving range, if a participant is found to produce significantly more than 70% of correct responses, the experimenter can conclude that the performance of this participant cannot be based solely on the unwanted cue.

There is an important distinction between ruling out the possibility that performance was based *solely* on an unwanted cue, and ruling out the possibility that the unwanted cue contributed significantly to performance. Even when performance is found to exceed PC_u , as determined using equation 6, it does not necessarily follow that the unwanted cue did not contribute significantly to performance. To completely rule out any possible contribution of the unwanted cue to performance, one would have to use an infinitely large roving range. In practice, ensuring that PC_u does not exceed the chance level (i.e., 0.5 in a 2I-2AFC task) by more than a few percentage points usually requires an impractically large roving range. For example, to limit PC_u to under 0.52 in a 2I-2AFC task, D must be about 50 times larger than Δ_u . If the size of the unwanted cue is 1 dB, this implies that the roving range must be at least 50 dB. The application of random perturbations over such a wide range is often impractical, and even in cases where such wide stimulus fluctuations can be applied, they often have undesirable consequences on performance—if only, by distracting the participant. Thus, experimenters usually prefer to avoid using wide roving ranges.

Considering that roving ranges required to ensure that unwanted cues cannot significantly contribute to performance are often impractically large, it seems highly desirable for researchers to have a way of estimating (*post-hoc*) the extent to which unwanted cues could contribute to performance, given the roving range that was actually used in the experiment. Equations 1 and 2 cannot be used for this purpose. A more complete model of the influence of randomization is needed, which takes into account the possible combination, by the observer, of information derived from the primary cue with information derived from the unwanted cue. In addition, the model should allow the researcher to separate the contributions of the two cues, and to predict the performance that can be achieved based on the primary cue alone.

A Two-Channel Model of the Influence of Roving on Performance

As mentioned in the preceding section, Green's (1988) analysis yields an upper bound on performance based *solely* on the unwanted cue, assuming that the participant completely ignores the primary cue. By contrast, the models of Durlach and colleagues (Durlach et al., 1986, 1989), allow for the possibility that the observer combines information from the unwanted cue with information from the primary cue in order to achieve higher performance than can be obtained by using the primary cue alone or the unwanted cue alone. Here, we extend Green's analysis to take into account the latter possibility.

Specifically, we consider a two-channel observer which uses both the primary cue and the unwanted cue for task performance. The observer's processing of the primary cue is assumed to be limited by internal noise, so that its performance is imperfect, as is that of human participants. By contrast, the processing of the unwanted cue is assumed to be unimpeded by internal noise, and to be limited only by the random variability in the external stimulus, i.e., the roving. It is important to stress that this assumption in no way negates the possibility that, for human observers, the processing of the unwanted cue may be significantly limited by internal noise—in addition to external noise. However, our goal, in this work, is to provide researchers with an upper bound on performance based on the unwanted cue, that is, an estimate of the highest PC that an observer can achieve by relying solely, but optimally, on the unwanted cue; this requires assuming that the unwanted cue is processed optimally. In addition, we assume that the primary cue and the unwanted cue are processed independently, i.e., that there is no interaction in the processing of the two cues. Again, we acknowledge that this assumption may not always hold in practice. However, for the applications that we consider in this article, there is no evidence, to our knowledge, to suggest that the perception of the primary cue was significantly impeded (or enhanced) by the perception of the unwanted cue, or *vice versa*. While it would be possible to extend the model described in this article to accommodate interactions in the processing of the two cues, we believe that such an extension is beyond the scope of the present article.

In the remainder of this *Theory* section, we describe these equations which researchers can use to estimate the extent to which performance (PC) measured in an experiment using roving may have been influenced by the unwanted cue—and, consequently, to infer what performance could have been achieved if the participant had relied exclusively on the primary cue. In general, the latter performance is what researchers ultimately seek to determine. We consider two cases, which are commonly encountered in the auditory and visual psychophysics: the case of within-trial (across observation intervals) roving in 2I-2AFC experiment, and the case of across-trial roving in a 2I-2AFC or 1I-2AFC experiment. Here, we focus on building intuition, and provide the equations without proof. In the Appendix we explain how these equations were obtained.

2I-2AFC with within-trial roving—Green's (1988) analysis of the 2I-2AFC experiment with roving of the unwanted cue for each observation interval can be extended by taking into account both the performance based on the primary cue (which we hereafter denote as PC_p) and the performance based on the unwanted cue (PC_u) in a calculation of “combined” performance, PC_c , assuming that the observer combines the two cues optimally to obtain the highest possible performance. The mathematical derivation is provided in the *Appendix*. The solution is,

$$PC_c = (1 - \Delta_u/D)^2 PC_p + (2 - \Delta_u/D)(\Delta_u/D), \quad (8)$$

where D and Δ_u are as defined above, and PC_p is the proportion of correct responses that would be achieved by the observer if it relied exclusively on the primary cue.

Under the assumption that human participants behave as the optimal observer defined by equation 8, the measured proportion of correct responses in an experiment provides an estimate of PC_c , whereas the researcher is interested in estimating PC_p , i.e., the probability of a correct response based on the primary cue only. The terms in equation 8 can be rearranged to obtain PC_p as a function of PC_c , as follows.

$$PC_p = [PC_c - (2 - \Delta_u/D)(\Delta_u/D)] / (1 - \Delta_u/D)^2. \quad (9)$$

Researchers can use equation 9 to estimate PC_p by, firstly, setting PC_c equal to the proportion of correct responses measured in the participant—this is equivalent to replacing PC_c by a point estimate of its value. Secondly, the roving range, D , must be specified; in general, this quantity is defined by the experimenter. Finally, the researcher must know, or be able to estimate, the size of the unwanted cue, Δ_u . In some experimental applications, the actual size of the unwanted cue can be determined fairly precisely by the researcher, based on the parameters of the experiment; however, in some cases, the actual size of the unwanted cue is unknown, and the researcher has no other option than replacing Δ_u with the best available estimate of an upper bound on the actual size of the unwanted cue (see *Applications*).

Figure 2 illustrates the relationship between PC_c and PC_p in a 2I-2AFC experiment. Researchers can use this plot to find the approximate value of PC_p given PC_c , D , and Δ_u . For example, suppose that discrimination thresholds corresponding to $PC_c = 0.79$ were measured in an experiment, with a roving range of $D = 20$ units and an estimated Δ_u of 1 unit—the units of D and Δ_u are irrelevant, but they must be the same. Thus, the ratio Δ_u/D was equal to $1/20 = 0.05$. The line corresponding to this value of Δ_u/D in Figure 2 is the third (dotted) line from the bottom. The intercept of this line gives the sought value for PC_p : 0.77. This is somewhat, but not markedly lower than PC_c , indicating that the roving range was sufficiently wide to ensure that the PC measured in the experiment is not substantially higher than the PC that would have been achieved by a (suboptimal) observer relying solely on the primary cue. However, suppose that instead of being set to 20, the roving range, D , in the same 2I-2AFC experiment was chosen so as to limit PC_u to 0.60. Using Green's (1988) formula (equation 1), $PC_u = 0.60$ requires that D be about 9 times larger than Δ_u , so that Δ_u/D is approximately equal to 0.11. Figure 2 (and equation 8) indicates that, in this situation, PC_p is equal to 0.74. This is 5 percentage points lower than the PC measured in the experiment ($PC_c = 0.79$). Thus, in this example, the roving range was not wide enough, and the contribution of the unwanted cue is not negligible. In psychophysical experiments where PC estimates are often based on hundreds of trials, a difference of 5 percentage points in PC is statistically significant.

Across-trial roving—When roving is applied only once on each trial, the equations for PC_c , PC_u , and PC_p are different from those described above. Two cases arise commonly in practice. The first case involves a 2I-2AFC task, and the experimenter wants to ensure that participants compare an aspect of the stimulus (which corresponds to the primary cue) across the two observation intervals within a trial—as opposed to comparing a single stimulus (e.g., the first presented stimulus on a trial) against an internal “reference”, which usually refers to a long-term memory trace of the standard stimulus. To discourage the participants from adopting the latter strategy, the experimenter applies roving to the stimulus parameter of primary interest (e.g., in an auditory pitch-discrimination experiment, frequency) *across* trials. Applying roving *within* trials would obviously not make sense in this context because the roving is applied to the stimulus parameter of *primary* interest, so that a wide rove applied between the first and second observation interval would hamper participants' ability to rely on the primary cue. Examples of psychophysical studies falling in this category include studies by Ahissar et al. (2006) and Jesteadt & Bilger (1974), which we analyze in the *Applications* section. A second case of interest involves a 1I-2AFC task. In this context, the experimenter would like to ensure that the participants are comparing different aspects or features of a stimulus within each trial, as opposed to comparing one of these aspects to some internal reference. Examples of psychophysical studies falling in this category include studies by Danilova & Mollon (2010), Jesteadt & Bilger (1974), and Webster et al. (1990), which we analyze in the *Applications* section.

Although these two cases involve different tasks, it turns out that the mathematical expressions for PC_c , PC_u , and PC_p for these two cases are identical (see *Appendix*). PC_u is given by

$$\begin{aligned} PC_u &= 0.5 + 0.5 (\Delta_u/D) \text{ for } \Delta_u \leq D, \\ PC_u &= 1 \text{ for } \Delta_u > D, \end{aligned} \quad (10)$$

and the relationship between PC_p and PC_c is described by the following two equations,

$$PC_c = (1 - \Delta_u/D) PC_p + (\Delta_u/D), \quad (11)$$

and,

$$PC_p = [PC_c - (\Delta_u/D)] / (1 - \Delta_u/D). \quad (12)$$

In the following *Applications* section, we illustrate the use of equations 8-12 to analyze data obtained in various type of auditory and visual psychophysics experiments in which roving was used to limit participants' ability to rely on an unwanted cue for task performance.

Applications

In this section, we illustrate how the model and equations described above can be used, firstly, to estimate the extent to which thresholds or performance measures in psychophysical experiments may have been influenced by an unwanted cue despite the use of roving, and secondly, to obtain "corrected" performance (or threshold) measures, in which the estimated influence of the unwanted cue has been removed. These "corrected" measures provide an estimate of the performance (or of the thresholds) that would have been obtained if the participants had ignored the unwanted cue and relied solely on the primary cue—which is, usually, what the researcher is ultimately interested in determining.

To illustrate the wide range of applicability of the approach, we consider several examples drawn from the visual- and auditory-perception literatures. These include experiments in which the primary and unwanted cues corresponded to the same physical dimension, as well as experiments in which the two cues corresponded to different physical dimensions. Moreover, experiments in which roving was used to limit participants' ability to take advantage of a long-term memory representation of the standard stimulus are also considered.

In general, the analyses described below follow two main steps. Firstly, the performance (or threshold) that the observer could achieve by relying solely on the unwanted cue is estimated. This first step usually requires determining the size of the unwanted cue (Δ_u) relative to the roving range (D), then substituting the result for Δ_u/D into equation 6 (for experiments involving 2I-2AFC tasks with within-trial roving) or equation 10 (for experiments involving across-trial roving) to obtain PC_u —the performance that could be achieved, in theory, by relying solely on the unwanted cue. If it is found that PC_u exceeds the PC measured (or targeted) in the experiment, then the possibility that the participants could perform the task based solely on the unwanted cue cannot be rejected, and further analysis is moot. If PC_u is lower than the measured (or targeted) PC, we then move to the second step of the analysis. In this second step, the estimated influence of the unwanted cue is "removed" from the performance measured in the experiment to yield a corrected estimate of performance, PC_p , which reflects performance based solely on the primary cue. To this aim, the measured (or targeted) PC is used as an estimate of PC_c and substituted for the latter into equation 9 (for 2I-2AFC tasks) or equation 12 (for 1I-2AFC tasks) to obtain PC_p .

While these are the general steps, the details of the analysis depend on the specifics of the considered application. In some applications, rather than measuring PC, the experimenters measured thresholds corresponding to a predefined PC, or sensitivity (d'). In such cases, it is necessary to know—or, when such knowledge is not available, to make assumptions concerning—the relationships between PC and other relevant quantities, such as d' and the sizes of the primary and unwanted cues. For the applications considered in this article, the assumption of an approximately linear relationship between d' and the size of the primary and unwanted cues is usually supported by empirical data, and the relationship between d' and PC is predicted by the equal-variance Gaussian SDT model (Green & Swets, 1966).

Primary and Unwanted Cues on the Same Physical Dimension

Under this heading, we consider experiments in which the unwanted cue and the primary cue occupy the same physical dimension, and are expressed in the same physical units. In this situation, the size of the unwanted cue can often be computed directly; it does not need to be estimated by the researcher. This simplifies the analysis, and yields more precise estimates of the respective contributions of the primary and unwanted cues to performance than can be obtained in situations in which the primary and unwanted cues occupy different physical dimensions. Situations of the latter type are dealt with in a subsequent section.

Shape Discrimination in Vision—Our first example of application is an experiment by Nachmias (2008) concerning the discrimination of the shape of geometric figures. Here, we focus on one of the conditions tested in this experiment. In this condition, the observers had to discriminate between a square and a rectangle. The shape of such figures is defined by the aspect ratio, $a = h/w$, where h and w are the height and the width of the figure, respectively. For a square, by definition, $a = 1$, while for a rectangle, $a \neq 1$. Thus, the cue of primary interest in this experiment was a change in aspect ratio. However, changes in aspect ratio were necessarily accompanied by changes in h or in w . To discourage observers from taking advantage of the latter (unwanted) cue, Nachmias varied the height and the width of the figures randomly across observation intervals within each trial over a range of 40% ($\pm 20\%$) around the height of the “standard” stimulus (a square), while keeping the aspect ratio of the two figures (square and rectangle) as intended. The results that were obtained in this condition can be found in Figure 2 of Nachmias’ article. These results are expressed as Weber fractions for aspect ratio. In this context, the Weber fraction corresponds to the change in aspect ratio which the observer could discriminate correctly on 76% of the trials, divided by the aspect ratio of the standard stimulus. According to the results shown in Figure 2 of Nachmias’ article, the mean Weber fraction for aspect ratio was approximately equal to 5.3%, or 0.22 dB. Thus, observers needed a 0.22 dB increase in aspect ratio, on average, to obtain approximately 76% of correct responses in the shape-discrimination task.

Here, we address the following questions: (i) what was the average size of the unwanted cue, at threshold, in this condition? (ii) was the roving range used by Nachmias sufficiently large to ensure that observers could not obtain a PC of 0.76 using solely the unwanted cue in this condition? (iii) assuming that the human observers made optimal use of both cues during the experiment, what PC would they have obtained, and what change in aspect ratio would they have needed to achieve a PC of 0.76, had they relied solely on the primary cue?

We start by determining the size of the unwanted cue at threshold, Δ_u . Since in this experiment the unwanted cue was a change in height, the size of the unwanted cue at threshold is equal to the change in height corresponding to the Weber fraction for aspect ratio, 0.22 dB. Thus, $\Delta_u = 0.22$ dB. Next, we use equation 6 to compute, PC_u , the proportion of correct responses that the observer could have obtained by using solely the unwanted cue. To this aim, we must compute Δ_u/D , the size of the unwanted cue relative to the roving

range, D . As mentioned above, Nachmias used a 40% ($\pm 20\%$) around the height of the standard stimulus. In absolute terms, this corresponds ± 0.2 centered on a value of 1, i.e., from 0.8 to 1.2, or $10\log_{10}(1.2/0.8) = 1.76$ dB. Thus, $\Delta_u/D = 0.22/1.76 = 0.13$. Substituting this value for Δ_u/D into equation 6 we obtain, $PC_u = 0.62$. Since this value is smaller than the PC targeted in the experiment (0.76), we conclude that the size of the unwanted cue was too small for the observers to achieve the targeted PC based on the unwanted cue alone.

We now have all the information needed to estimate the quantity of primary interest, PC_p , using equation 12. After substituting 0.76 for PC_c , 0.22 for Δ_u , and 1.76 for D into this equation, we find: $PC_p = 0.69$. Note that this value is about 7 percentage points smaller than PC_c (0.76). Assuming that the function relating d' to aspect ratio in dB is approximately linear (Nachmias, 2006), we can estimate that, to obtain 76% of correct responses using solely the primary cue (a change in aspect ratio), the observers in this experiment would have needed a change in aspect ratio of about 7.9%. This is approximately 49% larger than the mean Weber fraction measured by Nachmias (which, as mentioned above, was approximately equal to 5.3%). This first example illustrates the importance of taking into account the potential influence of unwanted cues in studies of visual-shape discrimination.

Color Discrimination—Our second example is an experiment on parafoveal color discrimination by Danilova & Mollon (2010). In this experiment, the primary cue was a change in hue produced by a change in the relative amount of excitation of the three types of cones. These excitations are denoted S , M , and L , for the blue, green, and red cones respectively. In the absence of roving, however, discrimination could be based on changes in the excitation of just one type of cone; specifically, the red cones. To limit the ability of their observers to rely on the latter cue rather than on changes in hue, the authors applied random variations to the green and red components, so that the total excitation of the green and red cones ($L+M$) varied randomly across trials. The roving range was equal to 10% ($\pm 5\%$), or approximately 0.435 dB. The question is whether this roving range was sufficient to ensure that the observers in this experiment could not obtain 79.4% of correct responses (the level targeted in this experiment).

To answer this question, we need to determine, firstly, the size of the unwanted cue at threshold, Δ_u . In this experiment, the unwanted cue at threshold corresponds to the change in L at threshold. The smallest and largest changes in L at threshold that were measured by Danilova & Mollon in three observers can be found in Figure 4 of their article. These thresholds were comprised between 1.001 and 1.004. Since these thresholds are already expressed as ratios, and since the unwanted cue is related linearly to the primary cue, the size of the unwanted cue at threshold corresponds to the square of the measured threshold.³ Thus, when expressed in dB, Δ_u was comprised between $10\log_{10}(1.001^2) = 0.0087$ and $10\log_{10}(1.004^2) = 0.0347$. Next, we compute Δ_u/D . As mentioned above, this study used a 10% roving range which, when expressed as a ratio in dB, equals $10\log_{10}(1.05/0.95) = 0.435$. Thus, Δ_u/D was comprised between $0.0087/0.435 = 0.02$ and $0.0347/0.435 = 0.08$. Substituting these Δ_u/D values into equation 10 (the equation for the single-interval case, consistent with the use of a single-interval spatial 2AFC task in this experiment), we find that for this experiment, PC_u was comprised between 0.510 and 0.540. Similarly, after substituting the two Δ_u/D values into equation 12, we find that PC_p was comprised between 0.78 and 0.79. Thus, for the lowest threshold (1.001), PC_u (0.510) is close to chance (0.500), and PC_p (0.790) is practically indistinguishable from the PC targeted in the experiment (0.794). Even for the highest threshold (1.004), PC_u (0.540) is only slightly above chance, and PC_p (0.776) is only slightly below the targeted PC (0.794). Thus, for these observers, the 10% roving range was sufficient to effectively rule out the possibility that the unwanted cue significantly influenced the thresholds that were measured in this experiment.

Discrimination of Spectral Shape in Audition—The third example of application is an auditory-perception experiment by Mason, Kidd, Hanna, & Green (1984). This experiment is concerned with the ability of human listeners ability to discriminate the spectral shape of sounds, an important topic of research commonly known as “profile analysis” in auditory psychophysics (see Green, 1988, for a review). In a typical profile-analysis experiment, the listener must discriminate two sounds that differ in their spectral shape. For example, one sound—referred to as the “standard” tone— has a “flat” spectrum, while the other sound—referred to as the “comparison” tone—contains a spectral peak (Figure 3). One way for the listener to discriminate between the standard and comparison tones involves comparing the levels of the components in each sound: in the comparison tone, the level of the component corresponding to the peak is higher than the level of the other components; the difference in level between the components is the primary cue, Δ_P . However, if the absolute level of the components does not vary across stimulus presentations, the listener can discriminate between the standard and comparison tones by comparing the overall levels of the two sounds (since the comparison tone contains a higher-level component, its overall level is greater than that of the standard) or, by comparing the levels of corresponding components across the two sounds; in the example of Figure 3, the observer can perform the task simply by comparing the level of the middle component across the two sounds. The difference in the level of the middle component across the standard and comparison sounds is the unwanted cue, Δ_U . In the example illustrated in Figure 3, this cue has the same magnitude as the primary cue, Δ_P . To limit listeners’ ability to take advantage of the unwanted cue, profile-analysis researchers usually randomize the overall level of the sounds on each presentation. They reason that, as long as the roving range, D , is sufficiently large—compared to the smallest level difference that listeners can reliably discriminate—the influence of the unwanted cue on performance should be negligible. However, this influence is rarely quantified. Here, we illustrate how the model described above can be used to, firstly, estimate the influence of unwanted absolute-level cues in profile-analysis experiments, and secondly, to “partial out” this influence from the measured thresholds, so that the results reflect specifically the discrimination of spectral shape.

Before the analysis of Mason et al.’s (1984) data can be presented, the following important points must be mentioned. Mason et al.’s study involved a 2I-2AFC task. The spectra of the stimuli were similar to those depicted in Figure 3a. Three roving ranges (D) were tested: 10, 20, and 40 dB. For each roving range, discrimination thresholds were measured using an adaptive procedure that tracked a PC of 0.707 on the psychometric function. The mean thresholds that were obtained in the experiment (Figure 1 in Mason et al., 1984) are re-plotted (as triangles) in Figure 4. In the auditory profile-analysis literature, it is customary to express thresholds as “signal-to-standard ratios.” Denoting as P the sound pressure of the standard (or reference) stimulus, and as ΔP the difference in pressure between the standard and comparison stimuli, the signal-to-standard ratio, T , equals $20\log_{10}(\Delta P/P)$. Given a standard-to-noise ratio, T , the corresponding level difference, Δ (in dB), is obtained as: $20\log_{10}(1+10^{T/20})$; conversely, T equals $20\log_{10}(10^{\Delta/20} - 1)$.

The first step in our analysis seeks to determine whether the participants in Mason et al.’s experiment could achieve PC = 0.707 (the PC targeted in this experiment) using solely the unwanted cue. To answer this question, we compute PC_U by applying equation 6 to the thresholds measured by Mason et al. (triangles in Figure 4). We find: $PC_U = 0.603$ for the 10-dB roving range, $PC_U = 0.584$, for the 20-dB roving range, and $PC_U = 0.548$ for the 40-dB roving range. These values are all well below 0.707. Therefore, we conclude that the listeners’ performance cannot be entirely accounted for by assuming that they relied *solely* on the unwanted cue. Consistent with this, the thresholds that we estimated based on the assumption that the listeners relied solely on the unwanted cue—these thresholds, which are

shown as open squares in Figure 4, were computed by setting $PC_u = 0.707$ in equation 6, then solving for Δ_u , and transforming the resulting Δ_u values into signal-to-standard ratios as explained above—are larger than the thresholds measured by Mason et al.

While these results indicate that Mason et al.'s listeners did not rely *solely* on the unwanted cue, they do not imply that the listeners did not use the unwanted cue *at all*. Thus, it is important to assess the relative contributions of the primary cue and of the unwanted cues to the measured performance, and to the measured thresholds. This is the aim of the second step of the analysis. This second step includes multiple sub-steps. Firstly, the threshold signal-to-standard ratios that were measured by Mason et al. (triangles in Figure 4) are converted into level differences (in dB) using the formula given above: $\Delta = 20\log_{10}(1+10^{T/20})$. Secondly, $\Delta_p = 20\log_{10}(1+10^{T/20})$, D (= 10, 20, or 40 dB), and PC_c (= 0.707) are substituted into equation 9 to obtain PC_p . Thirdly, the size of the primary cue needed to achieve $PC = 0.707$ using solely the primary cue is computed; to avoid confusion between this quantity and Δ_p , we denote it as, $\Delta_p^{0.707}$. Computing this quantity requires knowing how d' is related to the signal size (Δ) for the considered perceptual task. For spectral-shape discrimination, the relationship between d' and Δ is approximately linear (Dai, 1994; Dai & Green, 1992; Raney, Richards, Onsan, & Green, 1989). Accordingly, $\Delta_p^{0.707}$ can be computed as,

$$\Delta_p^{0.707} = \Delta_p \left(0.77 / d'_p{}^{PC_p} \right) \quad 13$$

where, as mentioned above, $\Delta_p = 20\log_{10}(1+10^{T/20})$, and $d'_p{}^{PC_p}$ is found using the equal-variance Gaussian SDT model for the 2I-2AFC paradigm $d' = \sqrt{2}\Phi^{-1}(PC)$ (Green & Swets, 1966). The constant, 0.77, is the value of d' corresponding to $PC = 0.707$ according to this same model. Lastly, the $\Delta_p^{0.707}$ value computed using equation 13 is transformed into a signal-to-standard ratio using the formula, $20\log_{10}(\Delta_p^{0.707} - 1)$. The result is the threshold predicted by the model for discrimination based solely on the primary cue, after removing the estimated contribution of the unwanted cue.

The predicted thresholds, which were computed in this way for each of the three roving ranges tested by Mason et al. (10, 20, and 40 dB), are shown as filled circles in Figure 4. It is important to remember that these predicted thresholds are based on the assumption that, in the actual experiment, the listeners used the primary and unwanted cues optimally. Thus, the measured thresholds (triangles) are assumed reflect the optimal use of both cues, while the predicted thresholds (circles) reflect optimal use of the primary cue alone. Although we cannot determine certainly the actual contribution of the unwanted cue with certainty, we can conclude that, under the model assumptions stated above, listeners' true thresholds for the discrimination of spectral shape in Mason et al.'s study are inside the range delimited by the triangles and the circles in Figure 4.

Finally, consider the predicted thresholds (circles) in Figure 4. For the 10- and 20-dB roving ranges, these thresholds are 4.6 and 3.5 dB higher, respectively, than the thresholds measured by Mason et al. (triangles). Such differences are larger than the typical within-listener standard deviation of measurements in studies of this type: 1-2 dB (e.g., Green, 1988). This indicates that the 10- and 20-dB roving ranges were not large enough to make the influence of the unwanted cue negligible. For the 40-dB roving range, the difference between the corrected thresholds and the measured thresholds is smaller (about 1.7 dB), and comparable to the typical variability of repeated measurements within the same listener in

this type of study—which, as stated above, is usually between 1 and 2 dB. Thus, for practical purposes, the 40-dB range was sufficiently large to neutralize the unwanted cue.

It is worth making one final observation concerning this example. If the primary cue were processed completely independently from the unwanted cue, the corrected thresholds, which reflect sensitivity to the primary cue only, should not depend on the roving range. The finding that they do (as indicated by Figure 4) indicates that, in this experiment, roving had an adverse effect on the processing of the primary cue. A possible explanation for this outcome is that roving was applied to the same dimension as the primary cue: sound level. It could be that, when trying to detect differences in level between simultaneous components within complex tones, human listeners are unable to completely disregard differences in overall level between the tones. Models that assume that the observer processes the primary cue optimally, and ignores the unwanted cue completely, cannot explain this effect.

Auditory Intensity Discrimination—The fourth example of application is an intensity-discrimination experiment by Dai & Green (1992). In this experiment, two pairs of tones differing in frequency were presented sequentially to the listener. As illustrated schematically in Figure 3b, the two tones in one of the two pairs had the same level, while the two tones in the other pair had a different level. The two pairs were presented in random order, and the listener's task was to indicate which pair contained the two tones having different levels. The primary cue was the level difference between these two tones, denoted Δ_p . This difference was varied adaptively using a maximum-likelihood procedure, which tracked the point corresponding to $PC = 0.94$ on the psychometric function. However, in the absence of any random variation in the levels of the tones in the first and second pairs, the listeners could in principle perform the task without ever comparing the levels of the two tones in each pair; they could identify the pair containing tones with different levels based on the fact that this pair contained a higher-level tone than the other pair. In fact, most listeners find it considerably easier to compare the levels of tones having the same frequency than to compare the levels of tones that differ widely in frequency. Therefore, the experimenters had a good reason to think that the listeners would prefer to rely on across-pair, same-frequency comparisons than on within-pair, across-frequency comparisons. Thus, in this experiment, the unwanted cue was a difference in level across the two pairs of tones. We denote the magnitude of this difference as Δ_u . In an attempt to prevent the listeners from relying on across-pair comparisons, they added a random quantity to the levels of the two tones in the first pair, and another random quantity to the levels of the two tones in the second pair. These two random quantities were independently drawn from a uniform distribution with a range (D) of 20 dB (± 10 dB).

The main results of this experiment are shown in Figure 5. The values along the x-axis indicate the frequency of the second tone in the pairs; the frequency of the first tone was fixed at 1000 Hz. The triangles show the mean discrimination thresholds that were measured by Dai & Green (1992), re-plotted from Figure 2 in their article. Note that these thresholds increased as the frequency of the second tone increased. As discussed by Dai & Green, this effect suggests that the listeners were indeed comparing the levels of the tones in each pair: as the frequency of the second tone increased—while the frequency of the first tone remained fixed—the two tones in each pair became more distant in frequency, and it became more difficult for listeners to compare their levels. However, important questions, which were not addressed by Dai & Green, are: (i) To what extent could the listener's performance in this task be influenced by the unwanted cue, in addition to the primary cue? (ii) What thresholds could the listeners obtain by relying solely on the unwanted cue? (iii) To what extent could the unwanted cue contribute to the thresholds that were measured in this experiment? (iv) How large are thresholds for across-frequency level discrimination (the perceptual ability of primary interest in this study), after the potential contribution of same-

frequency, across-pair comparisons (the unwanted perceptual strategy) has been partialled out?

The analysis of these data follows the same general steps as the preceding example. These steps are not reiterated here. Instead, the results of the analysis are presented directly. The “corrected” thresholds, which were obtained after removing the estimated influence of the unwanted cue from the measured thresholds, are shown as circles in Figure 5. For a reason explained below, corrected thresholds could only be computed for the three conditions in which the frequency of the second tone was either equal, or very close, to the frequency of the first tone—the data points corresponding to these three conditions are the three data points closest to the dip of the V-shaped pattern. However, note that even for the two smallest non-zero frequency separations—which correspond to the first two points on either side of the dip of the V-shaped pattern—the corrected thresholds were already as high as the thresholds that could theoretically be achieved by relying *solely* on the unwanted cue, which are shown as squares. For all other frequency-separation conditions, the thresholds measured by Dai & Green (1992) can be explained entirely, or almost entirely, by assuming that listeners relied solely on the unwanted strategy, i.e., comparisons between tones of the same frequency across the two pairs. This is why, for these conditions, “corrected” thresholds (after the estimated contribution of the unwanted cue was partialled out) could not be estimated, or were too high to be shown on this plot.

Based on these results, we conclude that the 20-dB roving range used by Dai & Green (1992) may not have been sufficient to convincingly rule out the possibility that the performance of the listeners in their study was significantly influenced by the unwanted cue, and that the thresholds that were measured by these authors using relatively large frequency separations may substantially over-estimate the ability of human listeners to compare the levels of sequentially presented tones with widely different frequencies. Thus, this analysis underscores the importance of quantifying the potential influence of unwanted cues in studies of perception.

Roving to Prevent the Use of an Internal Standard—Another type of empirical data that can be analyzed using the approach described above are data from experiments in which roving is used to limit the participants’ ability to benefit from the use of a fixed reference stored in long-term memory. For instance, in a 2I-2AFC discrimination experiment with an unchanging standard stimulus, participants can in principle memorize the standard, and then perform the task by comparing the first presented stimulus to the stored memory representation of the standard to decide whether it is the standard or comparison stimulus; they no longer need to compare the two stimuli presented on a trial. To limit listeners’ ability to take advantage of this unwanted strategy, experimenters can rove the standard stimulus, and then adjust the comparison stimulus based on the current standard so as to maintain a consistent difference between the two stimuli, which corresponds to the primary cue. We consider two examples of experiments of this type.

The first experiment was published in the journal *Nature Neuroscience* by Ahissar, Lubin, Putter-Katz, & Banai (2006). These authors measured thresholds for auditory frequency discrimination in dyslexic and control (age-matched, non-dyslexic) children. In one condition of the experiment, the frequency of the standard (lower-frequency) tone was fixed at 1000 Hz across a block of trials. In the other condition, the standard frequency was roved over a 400-Hz range (from 1000 to 1400 Hz) across trials. In the latter condition, the frequency of the comparison (higher-frequency) tone was varied adaptively to track the threshold corresponding to $PC = 0.707$ on the psychometric function. The results showed that, for the control group, thresholds were significantly higher (i.e., worse) in the roving-standard condition than in the fixed-standard condition. By contrast, in the test group, there

was no benefit of keeping the standard frequency fixed across trials; in fact, in the dyslexic children, thresholds were significantly worse in the fixed-standard than in the roved-standard condition. Ahissar and colleagues interpreted these results as an indication that the dyslexic children have difficulties forming and/or retrieving from memory stable perceptual representations (“anchors”) of repeatedly presented stimuli; thus, these children are forced to rely on stimulus comparisons and on echoic memory, even in conditions in which other children, who do not have dyslexia, benefit from using representations stored in long-term memory. However, one may wonder whether the roving range that was used in this study (400 Hz) was sufficiently large, relative to the thresholds that were measured in the test group (about 300 Hz on average in the fixed-standard condition, and about 400 Hz on average in the roving-standard condition), to warrant the assumption that the roving effectively prevented listeners from using an “anchor” for correct task performance.

The analysis is straightforward. Given that the mean threshold in the roving-standard condition was approximately equal to 300 Hz, or 30% of the standard frequency (1000 Hz), and that the roving range was equal to 400 Hz, $\Delta_u/D = 0.75$. Substituting this value into equation 10, we find that PC_u (for the considered 2I-2AFC task) was equal to 0.88. Thus, in theory, the listeners of the test group could achieve 88% correct in the roving-standard condition using the unwanted cue alone. Since 88% is larger than 70.7% (the targeted point on the psychometric function in the considered study), we conclude that the roving range was not sufficiently wide to rule out the possibility that the dyslexic children in Ahissar et al.’s (2006) study did, in fact, rely entirely on a fixed anchor, even in the roving-standard condition.

The second experiment was performed by Jesteadt & Bilger (1974). These authors measured PC for the discrimination of tones that differed in level by relatively small amounts (between 0.75 and 2.25 dB) in four trained listeners. Here, we re-analyze the results of one of these listeners (subject #2), who performed generally less well than the other listeners—suggesting that this listener was most influenced by the rove. The authors tested two roving ranges (D): small (4 dB), and large (40 dB). The d' values that they measured in these two roving conditions are listed in the third column of Table 1. The first two columns indicate the level difference between the two tones that the listener had to discriminate, and the roving range. The last two columns show the results of a re-analysis of these data using equations 10 and 12. The computed PC values were converted to d' values using $d'\sqrt{2}\Phi^{-1}(PC)$ (Green & Swets, 1966). The column labeled “ d'_u ” contains estimates of the d' that could in principle be achieved by relying solely on absolute level, that is, by comparing the level of the tones to an internal reference, rather than comparing directly the levels of the two tones presented on a trial. As can be seen, for the large (40 dB) roving range, the obtained d'_u values were relatively small, indicating that this roving range effectively prevented the listener from using the unwanted listening strategy. This was expected, because the large roving range was considerably (about 17 times) larger than the level differences that the listeners had to discriminate. By contrast, for the small (4 dB) roving range, the predicted d'_u values were relatively large, being equal to approximately 40% of the d' values that were reported by Jesteadt & Bilger (1974). Thus, for this small roving range, a substantial portion of the measured sensitivity can be accounted for by comparisons between the absolute levels of the tones and a fixed internal reference. It is possible that the listener used a mixed strategy, in which he/she compared the levels of the two tones on some trials, and compared the absolute levels of the tones against a fixed internal reference on other trials. Finally, the column labeled “ d'_p ” shows estimates of sensitivity based on the primary cue alone. Consistent with our previous observation that with the small (4-dB) roving range the contribution of the unwanted cue was relatively large, d'_p was markedly smaller than the d' values reported by Jesteadt & Bilger (1974), being equal to 71-74% of

this value. For the 40-dB roving range, the calculated d'_p values were not markedly lower than the d' reported by Jesteadt & Bilger.

This example illustrates how the model described here can be used to obtain quantitative estimates of sensitivity to the cue of primary interest. In situations where the assumptions of the model are reasonable, the calculated d'_p values provide a lower bound on sensitivity to the primary cue. Under such circumstances, a reasonable conclusion is that true sensitivity to the primary cue was in the range bounded below by d'_p , and above by d' .

Primary and Unwanted Cues on Different Dimensions

In some studies, the unwanted cue that the experimenter seeks to eliminate occupies a different physical dimension than the primary cue. In this section, we consider two examples of such studies, one involving auditory perception, and the other involving visual perception.

Frequency Discrimination with Intensity Roving—Frequency is a fundamental property of sounds. Therefore, numerous studies have been devoted to characterizing the ability of human listeners to discriminate the frequency of tones. One complication in the interpretation of the results of these studies stems from the fact that, as the frequency of a tone changes, the loudness of the tone can also change, even if the physical intensity of the tone remains constant (see Footnote 1 for an explanation of the origin of this effect). Loudness changes that are associated with frequency changes provide a potential cue for task performance in frequency-discrimination experiments. This cue is usually undesirable for experimenters, who are specifically interested in the ability of listeners to perceive differences in pitch (the subjective correlate of frequency), rather than in the ability of listeners to perceive differences in loudness. To limit listeners' ability to take advantage of this unwanted loudness cue, experimenters usually rove the level of tones in frequency-discrimination studies (e.g., Dai, Nguyen, & Green, 1995; Emmerich, Ellermeier, & Butensky, 1989; Henning, 1966; Moore & Glasberg, 1989).

The results of frequency-discrimination experiments with roving-level tones can be analyzed using an approach similar to that used in the previous examples. First, the size of the unwanted cue, Δ_u , must be determined. Since in this type of application, the size of the unwanted cue is not related directly and simply to the primary cue, it must be estimated. The change in loudness induced by a change in frequency may be small or large, depending on various factors—including, not only the magnitude of the frequency change, but also, the intensity of the tones relative to the absolute threshold of hearing, the standard frequency, and even the types of headphones used. To ensure that the conclusions of our analysis do not depend critically on the specifics of the particular experiment being considered, we equate the size of the unwanted cue with the smallest detectable difference in level (in dB) between two pure tones. By “smallest detectable difference,” we mean specifically, the difference that can be discriminated correctly at a given PC. Since the two example studies considered in this section both measured thresholds corresponding to PC = 0.707 in a 2I-2AFC task, we define the “smallest detectable difference” as the difference yielding PC = 0.707 in a 2I-2AFC task. Data in the psychoacoustic literature indicate that this is approximately equal to 1 dB (Green, 1988). Accordingly, we set: $\Delta_u = 1$ dB. We stress that this is an estimate. If the smallest difference in loudness that a listener can detect turns out to be less than 1 dB, the contribution of the unwanted cue may be larger than indicated in the analyses presented below.

Having estimated the size of the unwanted cue, Δ_u , this value can be substituted, along with the roving range, D , and $PC_c (= PC = 0.707)$, into equation 6 to obtain PC_u , and into equation 9 to obtain PC_p . For this application, we also estimated the frequency-discrimination thresholds corresponding to PC_u and PC_p , so that the results of our analyses

could be compared directly with the thresholds measured in the two example studies. To this aim, we assumed that $\log(d')$ was approximately equal to $\log(\Delta f/f)$, where f denotes the frequency of the standard tone, and Δf denotes the frequency difference between the standard tone and the comparison tone. The assumed relationship between d' and PC was as predicted by the equal-variance Gaussian SDT model for the 2I-2AFC paradigm (Green & Swets, 1966).

The analysis described above was applied to two published sets of data on pure-tone frequency discrimination with roving (Moore & Glasberg, 1989; Dai et al., 1995). The mean frequency-discrimination thresholds that were measured in the presence of intensity roving in these two studies are re-plotted as Weber fractions ($\Delta f/f$) on a logarithmic scale in Figure 6, along with the “corrected” thresholds, i.e., the thresholds corresponding to PC_p . The results can be summarized as follows. For Dai et al.’s (1995) study, the difference between the corrected thresholds and the measured thresholds is relatively small, indicating that the 20-dB roving range that was used in that study was sufficiently large to eliminate almost completely the unwanted loudness cue. By contrast, for Moore & Glasberg’s (1989) study, the corrected thresholds are markedly higher than the measured thresholds, indicating that the 6-dB roving range used by these authors was too small to rule out the possibility that the listeners took advantage of the unwanted loudness cue.

This example illustrates two important points. The first point is that the “corrected” thresholds computed using the above-described method must not be regarded as precise estimates of the thresholds that would have been obtained if the participants had relied exclusively on the primary cue. Rather, the “corrected” thresholds provide an estimate of how large the thresholds for the discrimination of the primary cue could be, assuming that the participants made optimal use of the unwanted cue—in addition to the primary cue. This is a subtle, but very important, distinction. If the corrected thresholds were precise estimates of thresholds for frequency discrimination, the observation that they differ markedly between the two studies would be troubling, because thresholds for the discrimination of the primary cue—after the influence of the unwanted cue has been removed—should be at least as similar across studies as the measured thresholds. If the listeners did not use the unwanted cue optimally, the “true” thresholds for the discrimination of the primary cue could be closer to the measured thresholds than the “corrected” thresholds. Unless the researcher has access to additional information, which allows him/her to determine the extent to which the participant *actually* relied on the unwanted cues, all that can be concluded based on the analysis described above is that thresholds for the discrimination of the primary cue only can fall anywhere between the measured thresholds and the corrected thresholds. In fact, if the participants did not use the unwanted cue at all, the measured thresholds truly do provide unbiased estimates of the “true” thresholds for the discrimination of the primary cue.

Interestingly, Moore & Glasberg (1989) and Dai et al. (1995) also measured thresholds in the absence of intensity roving in the same listeners. These thresholds were not markedly lower than those measured with roving: on average, the thresholds measured without roving were only a factor of 1.15 smaller than the thresholds measured with roving. This was the case despite the wide roving range (20 dB) used by Dai et al.. This suggests that the listeners in these two studies made little use of loudness cues. The slight threshold elevation caused by roving could be due to the random intensity variations being distracting rather than to listeners using loudness cues. This is the second important point illustrated by this example: when the analysis leads to the conclusion that the possibility that participants made use of the unwanted cue cannot be ruled out, this should not be interpreted as evidence that the participants did use the unwanted cue.

Spatial Frequency Discrimination in Vision—Our final example of application is an experiment by Webster et al. (1990). In this experiment, the authors were primarily interested in the ability of observers to discriminate spatial frequency. However, two spatial gratings that have the same *physical* contrast do not necessarily have the same *perceived* contrast, if their spatial frequencies differ. Therefore, in the same way as measures of auditory frequency discrimination can be confounded by perceived-intensity cues, the perception of differences in spatial frequency (the primary cue) can be influenced by perceived-contrast cues (the unwanted cue). To limit the ability of their observers to take advantage of the latter cue, Webster et al. roved the contrast of luminance gratings from trial to trial in a 1I-2AFC task. Here, we examine whether the rove was sufficiently wide to rule out the possibility that the unwanted cue had a significant impact on performance.

Following the general strategy outlined above, we start by estimating the size of the unwanted (contrast) cue, Δ_u . The processing of spatial frequencies in the visual system may be modeled as involving a bank of overlapping bandpass filters in the spatial-frequency domain. Thus, a change in spatial frequency can be discriminated based on a change in contrast at the output of a single internal filter. Out of all the filters in the bank, the one for which the change in spatial frequency (away from the standard frequency) produces the largest change in contrast is not the filter that is centered on the standard frequency, but the filter for which the slope measured at the standard frequency is the steepest. It is assumed that the observer's perception of the unwanted contrast cue is based on the output of this filter. Based on behavioral data (e.g., Henning, 1988; Sachs et al., 1971), we estimated the slope of the filter to be 0.7 log units in contrast ratio per octave; this means that a 1-octave change in spatial frequency produces a change of 0.7 log units in contrast ratio at the output of the filter. The mean Weber fraction of the two observers who discriminated spatial frequencies in the presence of the rove in Webster et al.'s study was equal to 0.04. This corresponds to a ratio of 1.04, or 0.057 octaves. We must double this number to account for the fact that observers in Webster et al.'s experiment actually had to discriminate stimuli, the spatial frequencies of which differed by a ratio equal to twice the ratio of the spatial frequency of each stimulus to the standard frequency. The change in contrast ratio at the output of the selected filter was approximately equal to 0.114 (the difference in spatial frequency between the two stimuli in octaves) times 0.07 (the filter slope), which equals 0.082 log units. This value is our estimate of Δ_u .

To determine PC_u , we simply substitute the Δ_u value computed in the previous step (0.082) and the roving range, D , into equation 10. Webster et al. (1990) used a uniform rove on contrast with a range of 0.15 log units. Thus, $D = 0.15$, and $\Delta_u/D = 0.082/0.15 = 0.547$. The result is: $PC_u = 0.774$. This value exceeds the PC that was targeted in experiment ($PC = 0.75$). Thus, we find that for this study, as for the study by Ahissar et al. (2006) described above, the roving range was not sufficiently wide to rule out the possibility that participants could perform the task at the targeted PC level by relying entirely on the unwanted cue. This example further underscores the potential importance of estimating the contribution of unwanted cues in psychophysical studies.

General Discussion

In the two preceding sections, we have outlined an approach based on an SDT model for separating the contributions of primary and unwanted cues to performance, sensitivity (d'), or thresholds in psychophysical experiments; we have illustrated the applicability of this approach to a wide range of experimental situations, using examples from the visual- and auditory-perception literatures. In this section, we address the following questions. Firstly, how do the approach and model described in this article differ from those proposed by Green (1988) and Durlach et al. (1986, 1989), which were described in *Theory*? Secondly,

how does this work relate to the broader theoretical framework of multidimensional detection theory? Thirdly, what are the limitations of the model, and how can it be extended to overcome these limitations?

Differences between the Present, Green's, and Durlach et al.'s Models

As explained in *Theory*, Green's (1988) often-cited analysis and its more recent extensions (Dai & Kidd, 2009; Dai & Michey1, 2010) allows researchers to determine the highest percentage of correct responses that can be obtained by an "ideal" observer relying exclusively on an unwanted cue to which roving is applied. Although this type of analysis represents an important first step, it is not sufficient. In general, the best performance that can be achieved by relying solely on the unwanted cue is not what the researcher is ultimately interested in; rather, the quantity of interest is the performance, sensitivity, or threshold that can be achieved by relying solely on the primary cue. When an unwanted cue is present, estimating this quantity requires a model of the processing of *both* the unwanted cue and the primary cue. From this point of view, the two-channel model described in this article differs importantly from the model assumed in Green's (1988) analysis or in recent extensions of this analysis (Dai & Kidd, 2009; Dai & Michey1, 2010).

The two-channel model described above may be regarded as a specific case of Durlach et al.'s (1986) multi-channel model. However, while Durlach et al. (1986) sought to formulate a general model, we opted instead to focus on the simplest case: only one primary cue and one unwanted cue. To our knowledge, all of the published studies in which roving has been used to limit the use of unwanted cues fall into this category. Moreover, one advantage of focusing on the simplest case is that it allowed more in-depth analyses than would have been possible otherwise. As the examples presented in the *Applications* section illustrate, even when only one primary cue and one unwanted cue are involved, sophisticated analyses involving several steps can be required in order to estimate the influence of the unwanted cue and estimate performance, sensitivity, or thresholds for the primary cue alone.

Lastly, as mentioned in *Theory*, the present work also builds upon some of the ideas outlined in Durlach et al. (1989). However, our approach differs from theirs in one important respect. Durlach et al. (1989) proposed to model the deleterious influence of randomization on discrimination performance as an increase in the variance of the internal noise associated with the representation of the primary cue. This reflects an assumption that the application of roving to the unwanted cue makes the primary cue less detectable, or less discriminable, than it would have been in the absence of such roving. By contrast, in the model described in the present work, it is assumed that the primary cue and the unwanted cue are processed optimally and independently, i.e., the processing of the unwanted cue does not interfere with the processing of the primary cue. Assumptions of optimality and of statistical independence play a key role in psychophysical SDT because they provide a guideline for the development of normative observer models, the predictions of which provide a baseline against which the performance of human observers can be compared (e.g., Green & Swets, 1966). As additional empirical data concerning the influence of stimulus randomization on perception become available, it will be important to test and, if necessary, relax these assumptions. In the final section of this discussion, we suggest ways in which the model described above could be amended to include interaction mechanisms and sources of sub-optimality in the processing of the primary and unwanted cues.

Relationship with Multidimensional Detection Theory and Cue Combination Models

How the perception of a multidimensional sensory stimulus relates to the perception of the individual dimensions, or components, of this stimulus is a fundamental question in psychophysics. The question has inspired numerous empirical studies and motivated the

development of theories such as multidimensional detection theory (MDT; Ashby & Townsend, 1986; Macmillan & Creelman, 2005). Studies of “cue combination,” or “perceptual integration,” and the models that have been formulated to analyze them (see, e.g., Fine & Jacobs, 1999; Jacobs, 1999, 2002; Lu & Doshier, 2000), fall into this category. So do many studies of selective and divided attention (e.g., Broadbent, 1958; Dai & Buus, 1991; Kahneman, 1973; Norman, 1976; Lu, Liu, & Doshier, 2000), of the effects of stimulus uncertainty (e.g., Davis & Graham, 1981; Green, 1961; Pelli, 1985), or of perceptual integration in time (e.g., Dai & Wright, 1996; Viemeister & Wakefield, 1991), frequency (e.g., Buus et al., 1986), or space (e.g., Shiffrin, McKat, & Shaffer, 1976). Many of these studies rely critically on comparisons of performance, sensitivity, or thresholds measured in conditions in which participants attend selectively to—or have access to only—one aspect (dimension or cue) of the stimulus, and performance, sensitivity, or thresholds measured in conditions in which the same participants attend—or have access—to two or more dimensions or cues.

The work described in this article provides another illustration of the ramifications of this fundamental question. Here, we focused on the problem of inferring performance, sensitivity, or thresholds for a stimulus dimension or cue of interest based on performance, sensitivity, or thresholds measured in experiments where the cue of interest was accompanied by another (unwanted) cue, and the participant’s ability to take advantage of the latter was limited by the application of random variations to the stimulus dimension corresponding to the unwanted cue. As mentioned above, the model that we formulated for this purpose is based on the general framework provided by psychophysical SDT, probability theory, and statistical decision theory, which also forms the basis of MDT and of cue-combination models (Ashby & Townsend, 1986; Macmillan & Creelman, 2005).

One difference between this work and other theoretical contributions in the area of MDT (e.g., Ashby & Townsend, 1986), which have usually remained quite general and abstract, is that, here, we deliberately abandoned generality by focusing exclusively on the simplest case—i.e., only two dimensions or cues—and by making normative assumptions—i.e., optimality and independence in the processing of the two cues—at the outset. One advantage of this approach is that it allowed us to analyze—or reanalyze—empirical data, which could not have been analyzed using more general models without making additional assumptions. For example, while MDT provides a general framework for modeling relationships between sensitivity for a single stimulus dimension and combined sensitivity for multiple dimensions, applications of this general framework to the analysis of experimental results is often limited by the lack of empirical data needed to determine, e.g., the nature and degree of any statistical dependence in the processing of the dimensions or cues. Although we are not advocating the blind reliance on assumptions of independence and optimality, we believe that, until sufficient empirical data are available to make it possible for researchers to precisely specify the structure and parameters of multidimensional SDT models, these normative assumptions will provide a useful starting point, and a baseline, for analyzing the results of psychophysical experiments.

Limitations and Possible Extensions of the Model

The relative simplicity of the two-channel model comes at the expense of its generality. In particular, the model in its current form still cannot be used to analyze the results of experiments in which more than one unwanted cue, or more than one primary cue, is present. Dealing with such situations requires a model containing more than two channels, such as the multi-channel model of Durlach et al. (1986). However, one problem with complex models is that, by construction, they require more assumptions and more free parameters than simpler models, and that more empirical data are needed to test the validity of these assumptions and to evaluate these model parameters.

Another limitation of the model described above is that it does not allow for the possibility of interactions in the processing of the primary and unwanted cues: the two cues are assumed to be processed independently, and optimally. As mentioned above, the assumptions of optimal and independent processing are commonly made in the psychophysical SDT literature (e.g., Green & Swets, 1966), as they provide a natural starting point for the development of normative observer models. However, these assumptions may not always be warranted; there may be situations in which the application of random variations to the unwanted cue interferes with the ability to process the primary cue. One way to model such interference effects in the processing of the primary and unwanted cues is to assume that roving applied to the unwanted cue increases the variance of the internal noise corresponding to the primary cue (Durlach et al., 1989). The model described in this article can be amended to allow for this possibility. However, this implies adding one parameter—corresponding to the amount or factor by which the variance of the internal noise corresponding to the primary cue increases due to the roving. The value of this parameter may not always be unambiguously identifiable based on the data at hand.

Despite these limitations, as illustrated in *Applications*, the range of applicability of the simple two-channel model described in this article is quite wide. It applies to situations in which the primary and unwanted cues share the same dimension as well as situations in which these cues occupy different dimensions. It also applies to situations in which the unwanted cue is explicit (e.g., a difference in level between two physical stimuli), as well as to situations in which it is latent (e.g., a difference between a physical stimulus and a long-term memory representation, or perceptual “anchor”).

Summary

In this article, we have outlined a theoretical framework for analyzing and interpreting the results of psychophysical experiments in which the stimulus attribute (feature, dimension, or “cue”) that is of primary interest to the researcher is not the only cue available to the participant, and randomization (“roving”) is used to limit the participant’s ability to rely on an “unwanted” cue for task performance. The wide range of applications of this theoretical framework was illustrated using a variety of examples drawn from the auditory and visual psychophysics literatures. In several cases, the re-analyses of existing data using the model described in this article have revealed that the roving range that was used in the considered studies was not sufficiently large to warrant the conclusion that the participants could not benefit significantly from relying on the unwanted cue. In some cases, the possibility could not be ruled out that the measured performance was actually based *entirely* on the unwanted cue. We believe that these results underscore the importance of quantifying the potential influence of unwanted cues on performance (or thresholds) measured in psychophysical experiments *even when roving is used*. We hope that the approach outlined in this article will facilitate and promote this practice among perception researchers.

Acknowledgments

This research was supported by the University of Arizona and the National Institutes of Health, National Institute on Deafness and other Communication Disorders Grant R01 DC 05216. We are grateful to the Associate Editor, Dr. Zhong-Lin Lu, for guiding us through the revisions, and we thank the reviewers for their helpful comments and suggestions.

Appendix

In this appendix, we explain how equations 8-12 were derived. As a reminder, the aim is to derive formulas for PC based solely on a primary cue, PC_p , and for PC based solely on an unwanted cue, PC_u , given the PC measured (or targeted) in a 2I-2AFC discrimination

experiment, PC_c . The latter is assumed to reflect performance based on a combination of the primary cue and the unwanted cue. Specifically, following Green (1988), and consistent with our goal of determining an upper bound on performance based on the unwanted cue, we assume that the internal representation of the unwanted cue, Y_u , is noiseless and that the processor of this cue is ideal; therefore, as Green (1988), we equate the psychological quantity, Y_u , with the corresponding physical variable, X_u , to which the experimenter applies a random rove. The internal representation of the primary cue, Y_p , is assumed to be contaminated by internal noise, so that performance of the observer based on the primary cue is imperfect—as is the case for human observers. Finally, as mentioned in the main text, it is assumed that the primary and unwanted cues are processed independently of each other, so that Y_p and X_u are statistically independent. From the assumption of independence of Y_p and X_u , it follows that the joint probability density function (PDF) of Y_p and X_u , $h(X_u, Y_p)$, equals the product of the marginal PDFs of Y_p and X_u , i.e., $f(X_u)g(Y_p)$.

The variable X_u has a uniform distribution with a range of D . The PDF of X_u depends on which of the two stimuli that the observer must discriminate, A and B , is presented (as illustrated in Figure 1). The PDF of X_u conditioned on the occurrence of stimulus A , $f_{X_u|A}(X_u|A) = 1/D$ if $0 < X_u < D$, and $f_{X_u|A}(X_u|A) = 0$ elsewhere. The PDF of X_u conditioned on the occurrence of stimulus B is $f_{X_u|B}(X_u|B) = 1/D$ if $\Delta_u < X_u < D + \Delta_u$, where Δ_u is the size of the unwanted cue, and $f_{X_u|B}(X_u|B) = 0$ elsewhere.

To determine the PC that can be achieved by an observer making optimal use of the unwanted cue, PC_c , we must integrate the joint PDF of X_u and Y_p conditioned on the occurrence of stimulus A or of stimulus B , $h([X_u, Y_p]|A)$, and $h([X_u, Y_p]|B)$, over suitably chosen regions. For an ideal observer, these regions are bounded by equal-likelihood-ratio contours. However, the computations are greatly simplified by recognizing that, on a given trial (corresponding to the presentations of stimulus A and stimulus B in random order), either both observations of X_u fall within the range over which the two uniform PDFs, $f_{X_u|A}(X_u|A)$ and $f_{X_u|B}(X_u|B)$, overlap, i.e., the region bounded below by $X_u = \Delta_u$ and above by $X_u = D$ in Figure 1, or at least one of the observation falls outside of this range. In the first case, the unwanted cue provides no useful information for task performance. Therefore, the optimal observer must ignore the unwanted cue completely, and respond based on the primary cue only. As a result, for this case, the observer's performance equals PC_p . In the second case, the unwanted cue provides perfect information and the observed can achieve a PC of 1 by relying solely on the unwanted cue.

In a 2I-2AFC task, the observer makes two observations of the primary cue and two observations of the unwanted on each trial—one observation of each cue per observation interval. On each of these observation intervals, the probability that X_u falls into the range comprised between $X_u = \Delta_u$ and $X_u = D$ equals $(1 - \Delta_u/D)$. Assuming that the observations obtained in the two intervals are statistically independent, the probability that X_u falls into the range comprised between $X_u = \Delta_u$ and $X_u = D$ for both observation intervals equals $(1 - \Delta_u/D)^2$. The probability that X_u falls outside of the indicated range on at least one of the two observation intervals is equal to $1 - (1 - \Delta_u/D)^2 = (2 - \Delta_u/D)(\Delta_u/D)$. Thus, for a virtual observer following the strategy described above, PC_c in the 2I-2AFC task is given by,

$$PC_c = (1 - \Delta_u/D)^2 PC_p + (2 - \Delta_u/D)(\Delta_u/D) \quad (A1)$$

From equation A1, PC_p can be expressed as,

$$PC_p = [PC_c - (2 - \Delta_u/D)(\Delta_u/D)] / (1 - \Delta_u/D)^2 \quad (A2)$$

Equations A1 and A2 appear in the main text as equations 8 and 9, respectively.

Another situation of interest, besides within-trial (i.e., across-interval) roving in 2I-2AFC task, involves across-trial roving in a 2I-2AFC or 1I-2AFC tasks.⁴ The derivations for this case follow the same steps as those described above. The only difference is that, when roving is applied only once per trial, the probability that X_u falls into the range comprised between $X_u = \Delta_u$ and $X_u = D$ is equal to $(1 - \Delta_u / D)$, and the probability that the X_u falls outside of this range is simply the complement, Δ_u / D . As a result, for this case, PC_c is given by,

$$PC_c = (1 - \Delta_u / D) PC_p + (\Delta_u / D), \quad (A3)$$

From equation A3, PC_p can be expressed as

$$PC_p = [PC_c - (\Delta_u / D)] / (1 - \Delta_u / D). \quad (A4)$$

Equations A3 and A4 appear in the main text as equations 11 and 12, respectively.

The equation for PC_u was similarly derived. When X_u falls into the overlapped range, the proportion correct is at chance (0.5). When the X_u falls into a non-overlapped range, the proportion correct is perfect (1). Thus,

$$\begin{aligned} PC_u &= \Delta_u / D + 0.5 (1 - \Delta_u / D) = 0.5 + 0.5 (\Delta_u / D) \quad \text{for } \Delta_u \leq D, \\ PC_u &= 1 \quad \text{for } \Delta_u > D. \end{aligned} \quad (A5)$$

Equation A5 appear in the main text as equation 10.

References

- Ahissar M, Lubin Y, Putter-Katz H, Banai K. Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*. 2006; 9(12):1558–1564. doi:10.1038/nn1800.
- Ashby FG, Townsend JT. Varieties of perceptual independence. *Psychological Review*. 1986; 93(2): 154–179. doi:10.1037/0033-295X.93.2.154. [PubMed: 3714926]
- Berg BG, Green DM. Spectral weights in profile listening. *Journal of the Acoustical Society of America*. 1990; 88:758–766. doi:10.1121/1.399725. [PubMed: 2212301]
- Berliner JE, Durlach NI. Intensity perception. IV. Resolution in roving-level discrimination. *Journal of the Acoustical Society of America*. 1973; 53(5):1270–1287. doi:10.1121/1.1913465. [PubMed: 4712555]
- Berliner JE, Durlach NI, Braida LD. Intensity perception. VII. Further data on roving-level discrimination and the resolution and bias edge effects. *Journal of the Acoustical Society of America*. 1977; 61(6):1577–1585. doi:10.1121/1.381471. [PubMed: 893804]
- Broadbent, DE. *Perception and communication*. Pergamon; London: 1958. doi.org/10.1037/10037-000
- Buus S, Schorer E, Florentine M, Zwicker E. Decision rules in detection of simple and complex tones. *Journal of the Acoustical Society of America*. 1986; 80:1646–1657. doi.org/10.1121/1.394329. [PubMed: 3794070]
- Cormack LK, Riddle RB. Binocular correlation detection with oriented dynamic random-line stereograms. *Vision Research*. 1996; 36:2303–2310. doi:10.1016/0042-6989(95)00289-8. [PubMed: 8776495]
- Dai H. Signal-frequency uncertainty in spectral-shape discrimination: psychometric functions. *Journal of the Acoustical Society of America*. 1994; 96(3):1388–1396. doi:10.1121/1.410283. [PubMed: 7963004]
- Dai H. On suppressing unwanted cues via randomization. *Perception & Psychophysics*. 2008; 70(7): 1379–1382. doi:10.3758/PP.70.7.1379. [PubMed: 18927020]
- Dai H, Buus S. Effect of gating the masker on frequency-selective listening. *Journal of the Acoustical Society of America*. 1991; 89:1816–1818. doi: 10.1121/1.401015. [PubMed: 2045588]

- Dai H, Green DM. Auditory intensity perception: successive versus simultaneous, across-channel discriminations. *Journal of the Acoustical Society of America*. 1992; 91(5):2845–2854. doi: 10.1121/1.402965. [PubMed: 1629477]
- Dai H, Kidd GJ. Limiting unwanted cues via random rove applied to the yes-no and multiple-alternative forced choice paradigms. *Journal of the Acoustical Society of America*. 2009; 126(2):EL62–67. doi:10.1121/1.3175882. [PubMed: 19640017]
- Dai H, Michey1 C. On the Choice of Adequate Randomization Ranges for Limiting the Use of Unwanted Cues in Same-Different, Dual-Pair, and Oddity Tasks. *Attention, Perception & Psychophysics*. 2010; 72(2):538–547. doi: 10.3758/APP.72.2.538.
- Dai H, Nguyen QT, Green DM. A two-filter model for frequency discrimination. *Hearing Research*. 1995; 85(1-2):109–114. doi:10.1016/0378-5955(95)00036-4. [PubMed: 7559165]
- Dai H, Nguyen QT, Green DM. Decision rules of listeners in spectral-shape discrimination with or without signal-frequency uncertainty. *Journal of the Acoustical Society of America*. 1996; 99:2298–2306. doi.org/10.1121/1.415417. [PubMed: 8730076]
- Dai H, Wright BA. The lack of frequency dependence of threshold for short tones in continuous broadband noise. *Journal of the Acoustical Society of America*. 1996; 100:467–472. doi.org/10.1121/1.415859.
- Danilova MV, Mollon JD. Parafoveal color discrimination: A chromaticity locus of enhanced discrimination. *Journal of Vision*. 2010; 10(1):1–9. doi:10.1167/10.1.4. [PubMed: 20143897]
- Dannemiller JL, Stephens BR. Contrast gain control in psychophysical contrast discrimination. *Perception & Psychophysics*. 1998; 60(7):1153–1163. doi:10.3758/BF03206165. [PubMed: 9821777]
- Davis ET, Graham N. Spatial frequency uncertainty effects in the detection of sinusoidal gratings. *Vision Research*. 1981; 21:705–712. doi.org/10.1016/0042-6989(81)90079-1. [PubMed: 7293001]
- Drennan WR, Watson CS. Sources of variation in profile analysis. II. Component spacing, dynamic changes, and roving level. *Journal of the Acoustical Society of America*. 2001; 110(5 Pt 1):2498–2504. doi:10.1121/1.1408311. [PubMed: 11757939]
- Durlach NI, Braida LD. Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*. 1969; 46(2):372–383. doi:10.1121/1.1911699. [PubMed: 5804107]
- Durlach NI, Braida LD, Ito Y. Towards a model for discrimination of broadband signals. *Journal of the Acoustical Society of America*. 1986; 80(1):63–72. doi:10.1121/1.394084. [PubMed: 3745667]
- Durlach NI, Tan HZ, Macmillan NA, Rabinowitz WM, Braida LD. Resolution in one dimension with random variations in background dimensions. *Perception & Psychophysics*. 1989; 46(3):293–296. doi:10.3758/BF03208094. [PubMed: 2771623]
- Emmerich DS, Ellermeier W, Butensky B. A re-examination of the frequency discrimination of random-amplitude tones, and a test of Henning's modified energy-detector model. *Journal of the Acoustical Society of America*. 1989; 85:1653–1659. doi:10.1121/1.397953.
- Farrar CL, Reed CM, Ito Y, Durlach NI, Delhorne LA, Zurek PM, et al. Spectral-shape discrimination. I. Results from normal-hearing listeners for stationary broadband noises. *Journal of the Acoustical Society of America*. 1987; 81(4):1085–1092. doi:10.1121/1.394628. [PubMed: 3571725]
- Fine I, Jacobs RA. Modeling the combination of motion, stereo, and vergence angle cues to visual depth. *Neural Computation*. 1999; 11(6):1297–1330. doi:10.1162/089976699300016250. [PubMed: 10423497]
- Formby C, Muir K. Effects of randomizing signal level and duration on temporal gap detection. *Audiology*. 1989; 28(5):250–257. doi:10.3109/00206098909081630. [PubMed: 2803111]
- Forrest TG, Green DM. Detection of partially filled gaps in noise and the temporal modulation transfer function. *Journal of the Acoustical Society of America*. 1987; 82:1933–1943. doi: 10.1121/1.395689. [PubMed: 3429731]
- Furukawa S, Moore BC. Dependence of frequency modulation detection on frequency modulation coherence across carriers: effects of modulation rate, harmonicity, and roving of the carrier frequencies. *Journal of the Acoustical Society of America*. 1997; 101(3):1632–1643. doi: 10.1121/1.418147. [PubMed: 9069630]
- Garner, WR. *The processing of information and structure*. Erlbaum; Potomac, MD: 1974.

- Garner WR. Facilitation and interference with a separable redundant dimension in stimulus comparison. *Perception & Psychophysics*. 1988; 44(4):321–330. doi:10.3758/BF03210413. [PubMed: 3226879]
- Garner WR, Felfoldy GL. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*. 1970; 1:225–241. doi.org/10.1016/0010-0285(70)90016-2.
- Garner WR, Morton J. Perceptual independence: definitions, models, and experimental paradigms. *Psychological Bulletin*. 1969; 72(4):233–259. doi:10.1037/h0028024. [PubMed: 5355535]
- Gescheider, GA. *Psychophysics: the fundamentals*. 3rd ed.. Lawrence Erlbaum Associates; Mahwah, N.J.: 1997.
- Green DM. Detection of auditory sinusoids of uncertain frequency. *Journal of the Acoustical Society of America*. 1961; 33:897–903. doi.org/10.1121/1.1908839.
- Green, DM. *Auditory intensity discrimination*. Oxford University Press; New York: 1988. Profile analysis.
- Green DM, Mason CR, Kidd G Jr. Profile analysis: critical bands and duration. *Journal of the Acoustical Society of America*. 1984; 75(4):1163–1167. doi:10.1121/1.390765. [PubMed: 6725764]
- Green, DM.; Swets, JA. *Signal Detection Theory and Psychophysics*. Krieger; New York: 1966.
- Hall JW, Fernandes MA. The effect of random intensity fluctuation on monaural and binaural detection. *Journal of the Acoustical Society of America*. 1983; 74:1200–1203. doi: 10.1121/1.390023. [PubMed: 6643842]
- Heinz MG, Colburn HS, Carney LH. Evaluating auditory performance limits: II. One-parameter discrimination with random-level variation. *Neural Computation*. 2001; 13(10):2317–2338. [PubMed: 11571000]
- Henning GB. Frequency discrimination of random amplitude tones. *Journal of the Acoustical Society of America*. 1966; 39:336–339. doi:10.1121/1.1909894. [PubMed: 5908437]
- Henning BG. Spatial-frequency tuning as a function of temporal frequency and stimulus motion. *Journal of the Optical Society of America*. 1988; A5:1362–1373. doi:10.1364/JOSAA.5.001362. [PubMed: 3171732]
- Henning GB, Richards VM, Lentz JJ. The effect of diotic and dichotic level-randomization on the binaural masking-level difference. *Journal of the Acoustical Society of America*. 2005; 118(5): 3229–3240. doi:10.1121/1.2047167. [PubMed: 16334902]
- Holt LL, Lotto AJ. Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*. 2006; 119:3059–3071. doi:10.1121/1.2188377. [PubMed: 16708961]
- Hyman R, Well A. Perceptual separability and spatial models. *Perception & Psychophysics*. 1968; 3:161–165. doi.org/10.3758/BF03212721.
- Jacobs RA. Optimal integration of texture and motion cues to depth. *Vision Research*. 1999; 39(21): 3621–3629. doi:10.1016/S0042-6989(99)00088-7. [PubMed: 10746132]
- Jacobs RA. What determines visual cue reliability? *Trends in the Cognitive Sciences*. 2002; 6(8):345–350. doi:10.1016/S1364-6613(02)01948-4.
- Jesteadt W, Bilger RC. Intensity and frequency discrimination in one- and two-interval paradigms. *Journal of the Acoustical Society of America*. 1974; 55(6):1266–1276. doi:10.1121/1.1914696. [PubMed: 4846721]
- Kahneman, D. *Attention and effort*. Prentice-Hall; Englewood Cliff, NJ: 1973.
- Kidd G Jr, Mason CR, Brantley MA, Owen GA. Roving-level tone-in-noise detection. *Journal of the Acoustical Society of America*. 1989; 86(4):1310–1317. doi:10.1121/1.398745. [PubMed: 2808906]
- Lee B, Harris J. Contrast transfer characteristics of visual short-term memory. *Vision Research*. 1996; 36:2159–2166. doi:10.1016/0042-6989(95)00271-5. [PubMed: 8776482]
- Levi DM, McGraw PV, Klein SA. Vernier and Contrast Discrimination in central and peripheral vision. *Vision Research*. 2000; 40:973–988. doi:10.1016/S0042-6989(99)00225-4. [PubMed: 10720667]

- Levitt H. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*. 1971; 49:467–477. doi:10.1121/1.1912375. [PubMed: 5541744]
- Lockhead GR. Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*. 1966; 72(1):95–104. doi:10.1037/h0023319. [PubMed: 5967738]
- Lockhead GR. Processing dimensional stimuli: a note. *Psychological Review*. 1972; 79(5):410–419. doi:10.1037/h0033129. [PubMed: 5076869]
- Lu Z-L, Doshier BA. Spatial attention: Different mechanisms for central and peripheral temporal precues? *Journal of Experimental Psychology: Human Perception and Performance*. 2000; 26:1534–1548. doi.org/10.1037/0096-1523.26.5.1534. [PubMed: 11039483]
- Lu Z-L, Liu CQ, Doshier BA. Attention mechanisms for multi-location first- and second-order motion perception. *Vision Res*. 2000; 40:173–86. doi.org/10.1016/S0042-6989(99)00172-8. [PubMed: 10793895]
- Macmillan, NA.; Creelman, CD. *Detection theory: A user's guide*. 2nd ed.. Erlbaum; Mahwah, NJ: 2005.
- Maddox, WT. Perceptual and decisional separability. In: Ashby, FG., editor. *Multidimensional models of perception and cognition*. Erlbaum; Hillsdale NJ: 1992. p. 147-180.
- Maddox WT. Separating perceptual processes from decisional processes in identification and categorization. *Perception & Psychophysics*. 2001; 63(7):1183–1200. doi:10.3758/BF03194533. [PubMed: 11766943]
- Maddox WT, Dodd JL. Separating perceptual and decisional attention processes in the identification and categorization of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory & Cognition*. 2003; 29(3):467–480. doi:10.1037/0278-7393.29.3.467.
- Mason CR, Kidd G Jr, Hanna TE, Green DM. Profile analysis and level variation. *Hearing Research*. 1984; 13(3):269–275. doi:10.1016/0378-5955(84)90080-7. [PubMed: 6735933]
- Mauermann M, Long GR, Kollmeier B. Fine structure of hearing threshold and loudness perception. *Journal of the Acoustical Society of America*. 2004; 116(2):1066–1080. doi:10.1121/1.1760106. [PubMed: 15376673]
- McDermott HJ, Lech M, Kornblum MS, Irvine DRF. Loudness perception and frequency discrimination in subjects with steeply sloping hearing loss: Possible correlates of neural plasticity. *Journal of the Acoustical Society of America*. 1998; 104(4):2314–2325. doi:10.1121/1.423744. [PubMed: 10491696]
- McKee SP, Silverman GH, Nakayama K. Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research*. 1986; 26:609–619. doi: 10.1016/0042-6989(86)90009-X. [PubMed: 3739236]
- Melara RD, Marks LE. Hard and soft interacting dimensions: differential effects of dual context on classification. *Perception & Psychophysics*. 1990a; 47(4):307–325. doi:10.3758/BF03210870. [PubMed: 2345683]
- Melara RD, Marks LE. Interaction among auditory dimensions: timbre, pitch, and loudness. *Perception & Psychophysics*. 1990b; 48(2):169–178. doi:10.3758/BF03207084. [PubMed: 2385491]
- Micheyl C, Dai H. Likelihood ratio, optimal decision rules, and relationship between proportion correct and d' in the dual-pair AB-vs-BA identification paradigm. *Perception & Psychophysics*. 2009; 71(6):1426–1433. doi:10.3758/APP.71.6.1426.
- Micheyl C, Divis K, Wroblewski DM, Oxenham AJ. Does fundamental-frequency discrimination measure virtual pitch discrimination? *Journal of the Acoustical Society of America*. 2010; 128(4): 1930–1942. doi:10.1121/1.3478786. [PubMed: 20968365]
- Moore BC, Sek A. Discrimination of frequency glides with superimposed random glides in level. *Journal of the Acoustical Society of America*. 1998; 104(1):411–421. doi:10.1121/1.423297. [PubMed: 9670533]
- Moore BCJ, Glasberg BR. Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation. *Journal of the Acoustical Society of America*. 1989; 86:1722–1732. doi:10.1121/1.398603.
- Moore BCJ, Glasberg BR, Low KE, Cope T, Cope W. Effects of level and frequency on the audibility of partials in inharmonic complex tones. *Journal of the Acoustical Society of America*. 2006; 120(2):934–944. doi:10.1121/1.2216906. [PubMed: 16938981]

- Morgan MJ. The visual computation of 2-D area by human observers. *Vision Research*. 2005; 45:2564–2570. doi:10.1016/j.visres.2005.04.004. [PubMed: 15904946]
- Morgan M, Chubb C, Solomon JA. A ‘dipper’ function for texture discrimination based on orientation variance. *Journal of Vision*. 2008; 8(11):1–8. doi:10.1167/8.11.9.
- Murray RF, Beutter BR, Eckstein MP, Stone LS. Saccadic and perceptual performance in visual search tasks. II. Letter discrimination. *Journal of the Optical Society of America*. 2003; A20:1356–1370. doi:10.1364/JOSAA.20.001356. [PubMed: 12868640]
- Nachmias J. How is a grating detected on a narrowband noise masker? *Vision Research*. 1999; 39:1133–1142. doi:10.1016/S0042-6989(98)00166-7. [PubMed: 10343831]
- Nachmias J. Contrast discrimination with and without spatial uncertainty. *Vision Research*. 2002; 42:41–48. doi:10.1016/S0042-6989(01)00254-1. [PubMed: 11804630]
- Nachmias J. Visual discrimination with a real or virtual standard. *Vision Research*. 2006; 46:2456–2464. doi:10.1016/j.visres.2006.01.029. [PubMed: 16530243]
- Nachmias J. Judging spatial properties of simple visual figures. *Vision Research*. 2008; 48:1290–1296. doi:10.1016/j.visres.2008.02.024. [PubMed: 18407313]
- Nachmias J. Shape and size discrimination compared. *Vision Research*. 2011; 51:400–407. doi:10.1016/j.visres.2010.12.007. [PubMed: 21172373]
- Norman, DA. *Memory and Attention*. 2nd ed. Wiley; New York: 1976.
- Oxenham AJ, Buus S. Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions. *Journal of the Acoustical Society of America*. 2000; 107(3):1605–1614. doi:10.1121/1.428445. [PubMed: 10738814]
- Oxenham AJ, Michey C, Keebler MV. Can temporal fine structure represent the fundamental frequency of unresolved harmonics? *Journal of the Acoustical Society of America*. 2009; 125(4):2189–2199. doi:10.1121/1.3089220. [PubMed: 19354395]
- Pelli DG. Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America*. 1985; 2(A):1508–1532. doi:10.1364/JOSAA.2.001508. [PubMed: 4045584]
- Peterson WW, Birdsall TG, Fox WC. The theory of signal detectability. *Transactions IRE Professional Group on Information Theory, PGIT-4*. 1954:171–212. doi:10.1109/TIT.1954.1057460.
- Purdy DM. Spectral hue as a function of intensity. *American Journal of Psychology*. 1931; 43(4):541–559. doi:10.2307/1415157.
- Purdy DM. The Bezold-Brücke phenomenon and contours for constant hue. *American Journal of Psychology*. 1937; 49(2):313–315. doi:10.2307/1416302.
- Raney JJ, Richards VM, Onsan ZA, Green DM. Signal uncertainty and psychometric functions in profile analysis. *Journal of the Acoustical Society of America*. 1989; 86(3):954–960. doi:10.1121/1.398730. [PubMed: 2794248]
- Regan D, Hamstra S. Shape discrimination for motion-defined and contrast-defined form: Squareness is special. *Perception*. 1991; 20:315–336. doi:10.1068/p200315. [PubMed: 1762874]
- Regan D, Hamstra S. Shape discrimination and the judgement of perfect symmetry: Dissociation of shape from size. *Vision Research*. 1992; 32:1845–1864. doi:10.1016/0042-6989(92)90046-L. [PubMed: 1287983]
- Regan D, Hamstra S. Dissociation of discrimination thresholds for time to contact and for rate of angular expansion. *Vision Research*. 1993; 33:447–462. doi:10.1016/0042-6989(93)90252-R. [PubMed: 8503195]
- Shiffrin RM, McKay DP, Shaffer WD. Attending to forty-nine spacial positions at once. *Journal of Experimental Psychology: Human Perception and Performance*. 1976; 2:14–22. doi.org/10.1037/0096-1523.2.1.14. [PubMed: 1262795]
- Stellmack MA, Viemeister NF, Byrne AJ. Discrimination of depth of sinusoidal amplitude modulation with and without roved carrier levels. *Journal of the Acoustical Society of America*. 2006; 119(1):37–40. doi:10.1121/1.2133576. [PubMed: 16454261]
- Stevens SS. The relation of pitch to intensity. *Journal of the Acoustical Society of America*. 1935; 6:150–154. doi:10.1121/1.1915715.

- Strickland EA, Dhar S. An analysis of quasi-frequency-modulated noise and random-sideband noise as comparisons for amplitude-modulated noise. *Journal of the Acoustical Society of America*. 2000; 108(2):735–742. doi:10.1121/1.429606. [PubMed: 10955640]
- Terhardt E. Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*. 1974; 55:1061–1069. doi:10.1121/1.1914648. [PubMed: 4833699]
- Thai-Van H, Michey1 C, Moore BC, Collet L. Enhanced frequency discrimination near the hearing loss cut-off: a consequence of central auditory plasticity induced by cochlear damage? *Brain*. 2003; 126(Pt 10):2235–2245. doi:10.1093/brain/awg228. [PubMed: 12847078]
- Thomas RD. Assessing sensitivity in a multidimensional space: Some problems and a definition of a general d' . *Psychonomic Bulletin & Review*. 1999; 6:224–238. doi.org/10.3758/BF03212328. [PubMed: 12199209]
- Thomas RD. Further considerations of a general d' in multidimensional space. *Journal of Mathematical Psychology*. 2003; 47:220–224. doi.org/10.1016/S0022-2496(02)00029-9.
- Uchanski RM, Braida LD. Effects of token variability on our ability to distinguish between vowels. *Perception & Psychophysics*. 1998; 60(4):533–543. doi:10.3758/BF03206044. [PubMed: 9628988]
- van Trees, HL. *Detection, Estimation, and Modulation Theory*. Wiley; New York: 1968.
- Verschuure J, van Meeteren AA. The effect of intensity on pitch. *Acustica*. 1975; 32:33–44.
- Viemeister NF, Wakefield GH. Temporal integration and multiple looks. *Journal of the Acoustical Society of America*. 1991; 90:858–865. doi.org/10.1121/1.401953. [PubMed: 1939890]
- Webster MA, De Valois KK, Switkes E. Orientation and spatial-frequency discrimination for luminance and chromatic gratings. *Journal of the Optical Society of America*. 1990; A7:1034–1049. doi:10.1364/JOSAA.7.001034. [PubMed: 2362227]
- Yuan H, Reed CM, Durlach NI. Temporal onset-order discrimination through the tactual sense: Effects of frequency and site of stimulation. *Journal of the Acoustical Society of America*. 2006; 120:375–385. doi: 10.1121/1.2204452. [PubMed: 16875233]

Figure 1.

Schematic representation of the stimulus-conditional probability density functions (PDFs) of the stimulus variable corresponding to the unwanted cue, X_u , to which uniform roving is applied. The solid line labeled $f_{X_u|A}(X_u|A)$ shows the conditional PDF of X_u given that stimulus A was presented. The dashed line labeled $f_{X_u|B}(X_u|B)$ shows the conditional PDF of X_u given that stimulus B was presented. The two PDFs have the same range, D , and only differ by a horizontal shift, equal to the size of the unwanted cue, Δ_u .

Figure 2.

Predicted relationship between PC_p and PC_c for different relative unwanted-cue sizes (Δ_u/D), in the presence of (uniform) roving. Knowing the size of the unwanted cue (Δ_u) and the roving range (D), researchers can use this figure to determine the value of PC_p as a function of PC_c . (See text for details.)

Figure 3.

Schematic illustration of the stimuli used in the profile-analysis study by Mason et al. (1984) and in the intensity-discrimination study by Dai & Green (1992). a. In Mason et al.'s study, listeners had to discriminate a "standard" complex tone with a "flat" spectrum, i.e., equal-amplitude frequency components, from a "signal + standard" complex tone in which one of the components (here, the center component) was added a tone of the same frequency (the "signal"), resulting in an increased amplitude for that component. To limit listeners' ability to perform the task based on changes in the absolute level of the center component, the overall level of the complexes was roved on each stimulus presentation. As a result, the "signal + standard" stimulus could have a lower overall level than the "standard" stimulus, as illustrated here. b. In Dai & Green's (1992) experiment, listeners were presented on each trial with two successive pairs of tones that had the same or different frequencies (here, 200 and 1000 Hz), and they had to indicate the pair in which the first tone had a higher level. To limit listeners' ability to perform the task reliably based on level differences *between* the two pairs, level was roved across pairs, so that both tones in the pair containing the louder tone (the second pair shown) could actually have a lower level than the tones in the other pair (the first pair shown).

Figure 4.

Re-analysis of Mason et al.'s (1984) data on spectral-shape discrimination as a function of roving range. The thresholds measured by Mason et al. are shown as triangles. As is customary in the auditory "spectral-profile analysis" literature, these thresholds are expressed as signal-to-standard ratios, in dB (see the main text for details). The error bars show one standard deviation of the mean across listeners. The thresholds that could theoretically be achieved by relying solely on the unwanted cue (overall level differences) are shown as squares. The "corrected" thresholds for the spectral-shape discrimination, after eliminating the possible contribution of the unwanted cue, are shown as circles.

Figure 5.

Re-analysis of Dai & Green's (1992) data on auditory intensity discrimination. In that study the listeners compared the level of a 1000 Hz tone with the level of a second tone. The stimuli that were used in this experiment are illustrated schematically in Figure 3b. The thresholds that were measured originally by Dai & Green are plotted as a function of the frequency of the second tone, and are shown as triangles. The horizontal dashed line shows the threshold that could theoretically be achieved by relying solely on the unwanted cue. The filled circles show the thresholds that could theoretically be achieved by relying solely on the primary cue; these thresholds were computed as described in the main text. The grey upward-pointing arrows near the top of the plot indicate that the estimated thresholds were too large to be displayed on the plot. The dark upward-pointing arrows indicate conditions in which the predicted thresholds based on the primary cue alone could not be computed, because performance could be accounted for using solely the unwanted cue. All of the thresholds shown in this figure are expressed as signal-to-standard ratios in dB.

a

b

Figure 6.

Re-analysis the data of Moore & Glasberg (1989, panel 6a) and Dai et al. (1995, panel 6b) on auditory frequency discrimination. These plots show frequency-discrimination thresholds for pure tones as a function of the frequency of the standard. The thresholds that were reported in the original studies are shown as triangles. The circles show the thresholds that were predicted, using the analysis described in the text, assuming that the observers used only the primary cue.

Table 1

Re-analysis of the results obtained from an individual listener (Subject #2) in a 2I-2AFC intensity-discrimination experiment by Jesteadt & Bilger (1974).

Δ_p	D	d'	d'_u	d'_p
0.75 dB	4 dB	0.87	0.34	0.62
1.50 dB	4 dB	1.67	0.69	1.24
1.50 dB	40 dB	1.67	0.07	1.64
2.25 dB	40 dB	1.68	0.10	1.63

The first column indicates the level difference (in dB) between the two tones that were presented on a trial, and which the listeners had to discriminate. The second column indicates the range of level roving (in dB). The third column shows the d' values reported by Jesteadt & Bilger (1974). The last two columns show the results of calculations using the model described in this article: in the fourth column, d'_u refers to the sensitivity that could theoretically be achieved by using the unwanted strategy of comparing the level of each tone to a fixed internal reference stored in long-term memory; in the last column, d'_p refers to the sensitivity to the primary cue, i.e., level differences between the two tones presented on a trial. Small differences between d' and d'_p indicate that the unwanted cue had little influence on performance, so that performance based on the primary cue only was almost as large as the performance measured; by contrast, large deviations between d' and d'_p indicate that performance based on the primary cue alone cannot account for the measured performance, indicating that the latter was probably influenced substantially by the unwanted cue. It is worth pointing out that d'_u and d'_p can equal, but never exceed d' .