Decision-variable correlation

Stephen Sebastian

Wilson S. Geisler

An extension of the signal-detection theory framework is described and demonstrated for two-alternative identification tasks. The extended framework assumes that the subject and an arbitrary model (or two subjects, or the same subject on two occasions) are performing the same task with the same stimuli, and that on each trial they both compute values of a decision variable. Thus, their joint performance is described by six fundamental quantities: two levels of intrinsic discriminability ($d'$), two values of decision criterion, and two decision-variable correlations (DVCs), one for each of the two categories of stimuli. The framework should be widely applicable for testing models and characterizing individual differences in behavioral and neurophysiological studies of perception and cognition. We demonstrate the framework for the well-known task of detecting a Gaussian target in white noise. We find that (a) subjects’ DVCs are approximately equal to the square root of their efficiency relative to ideal (in agreement with the prediction of a popular class of models), (b) between-subjects and within-subject (double-pass) DVCs increase with target contrast and are greater for target-present than target-absent trials (rejecting many models), (c) model parameters can be estimated by maximizing DVCs between the model and subject, (d) a model with a center–surround template and a specific (modest) level of position uncertainty predicts the trial-by-trial performance of subjects as well as (or better than) presenting the same stimulus again to the subjects (i.e., the double-pass DVCs), and (e) models of trial-by-trial performance should not include a representation of internal noise.

Extended signal-detection theory framework

The workhorse behavioral paradigm in many areas of human and animal research (e.g., perception, memory, decision making) is the two-alternative identification task. On each trial of an experiment, the subject is presented with a stimulus drawn from one of two categories, $a$ and $b$, and is required to identify the category. The category identified by the subject (the subject’s response) can be represented with capital letters $A$ and $B$. Thus, on each trial there are four possibilities: The subject responds $B$ when the category is $b$ ($B|b$), the subject responds $B$ when the category is $a$ ($B|a$), the subject responds $A$ when the category is $a$ ($A|a$), or the subject responds $A$ when the category is $b$ ($A|b$). We note that the common two-alternative forced-choice task can be regarded as a special case where category $a$ is a pair of stimuli in one spatial or temporal order and category $b$ is a pair of stimuli in the opposite order.

Signal-detection theory (SDT) is the standard theoretical framework for interpreting the data measured in the two-alternative identification task (e.g., Green & Swets, 1974; Wickens, 2001). The theory assumes that the end result of all the processing in the organism, prior to the behavioral response, can be represented with a single, probabilistic decision variable that is described by one normal distribution if the stimulus is from category $b$ and another normal distribution if the stimulus is from category $a$. The behavioral response is assumed to be generated by comparing the value of the decision variable on the trial to a criterion placed along the decision-variable axis. The subject makes one response if the value falls below the criterion and makes the other response if it falls above the criterion. In Bayesian statistical models, this decision variable would typically correspond to the log likelihood ratio—the log of the ratio of the probability of the modeled pattern of neural activity given category $b$ to the probability given category $a$. However, it is important to point out that basic SDT makes no assumption about how the decision variable is computed. Its purpose is only to provide a principled interpretation of the four possible outcomes of trials in the identification task. Indeed, basic SDT can provide a principled interpretation even if there is no location
within the nervous system where a single decision variable is represented.

Within the SDT framework, the probabilities of the four possible stimulus–response outcomes are determined by two parameters: the number \( d' \) of standard deviations separating the two distributions, and the decision criterion \( \gamma \). Thus, without loss of generality, the two normal distributions can be represented as normal distributions with standard deviations of 1.0 and means at \(-d'/2\) and \(d'/2\) (Figure 1). In SDT, the parameter \( d' \) represents the intrinsic discriminability of the two categories, independent of the particular decision criterion adopted by the subject.

A typical use of SDT is to estimate the subject’s discriminability and decision criterion from the proportions of trials falling into the four possible stimulus–response outcomes already mentioned. Because the two proportions for each category must sum to 1.0, estimates of discriminability and criterion are obtained from just two proportions—for example, \( p(B|b) \) and \( p(B|a) \). In the special case of a yes–no detection task these two proportions represent the proportion of hits and false alarms (Green & Swets, 1974). These two proportions correspond to the areas under the two normal distributions above the decision criterion.

The estimates of the discriminability and the criterion are obtained by solving the following pair of equations:

\[
p(B|b) = \Phi\left(\frac{d'}{2} - \gamma\right) \tag{1}
\]
\[
p(B|a) = \Phi\left(-\frac{d'}{2} - \gamma\right) \tag{2}
\]

where \( \Phi(\cdot) \) is the standard normal integral function. Two subjects could differ in their intrinsic ability to identify the categories (the values of \( d' \)), in their decision biases (the values of \( \gamma \)), or in both. SDT provides a principled framework for measuring these quantities, allowing better comparison of performance across subjects or between subjects and models.

The main goal here is to describe an extension of the standard SDT framework that allows measurement of two additional quantities (two decision-variable correlations) that provide increased power for comparing subjects and testing models. This is a straightforward extension related to existing signal-detection analyses, including the estimation of classification images (Ahumada & Lovell, 1971; Ahumada, 1996; for a review, see Murray, 2011), estimation of the ratio of external to internal noise (Spiegel & Green, 1981; Burgess & Colborne, 1988), and estimation of choice probabilities in neurophysiology experiments (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Haefner, Gerwinn, Macke, & Bethge, 2013; Pitkow, Liu, Angelaki, DeAngelis, & Pouget, 2015; Michelson, Pillow, & Seidemann, 2017; Seidemann & Geisler, in press).

Importantly, the extension proposed here does not require any special or new experimental design features; it can be applied to data from almost any two-alternative identification experiment in which the stimuli within each of the two categories are not physically identical on each trial. The only requirement is that the specific category and stimulus presented on each trial be known. There are many published studies where the experimenters saved the specific stimuli, or the information necessary to generate the stimuli (e.g., random-number seeds); however, there are also many studies where that is not the case. Another goal of this report is to emphasize the value of saving the stimuli, or the information for regenerating them, and the value of making the specific stimuli and responses publicly available.

The extended SDT framework is illustrated in Figure 2. We assume that a subject and an arbitrary model (or two subjects, or the same subject on two occasions) are performing the same task with the same stimuli. Thus, we can represent the subject’s normalized decision variable along the horizontal axis and the model’s (or the second subject’s) normalized decision variable along the vertical axis. The subject will have some arbitrary discriminability \( d'_{s} \) and some arbitrary decision criterion \( \gamma'_{s} \). Similarly, the model will have some arbitrary discriminability \( d'_{m} \) and some arbitrary decision criterion \( \gamma'_{m} \). As usual, Equations 1 and 2 can be used to estimate the discriminability and decision criterion of the subject and model from their respective response proportions \( p_s(B|b), p_s(B|a) \) and \( p_m(B|b), p_m(B|a) \).

An accurate model would have the same discriminability as the subject under the same conditions. However, even if a model has the same discriminability

![Figure 1. Standard signal-detection theory framework for interpreting responses in a two-alternative identification task. The two categories are represented by \( A \) and \( B \), and the two possible behavioral responses by \( A \) and \( B \). The discriminability of the two categories \( d' \) is the number of standard deviations separating the distributions of the decision-variable values for the two categories. If the decision variable on a trial exceeds the decision criterion \( \gamma \), the response is \( B \) (the stimulus was from category \( b \)); otherwise the response is \( A \).](http://arvojournals.org/)
as the subject, the model and subject may vary in how correlated their responses are on a trial-by-trial basis. In the SDT framework, this is represented by the correlation between the decision variables. The better the model (or the more similar the two subjects), the higher is the correlation between the values of their decision variables (i.e., the more accurate the trial-by-trial predictions). These decision-variable correlations (DVCs) need not be the same for the two categories of stimuli. Thus, by measuring DVCs in addition to discriminabilities and decision criteria, we should have more power to discriminate between models, estimate model parameters, and characterize subjects based on individual differences.

The joint distribution of the two decision variables is given by one bivariate normal distribution for category \(a\) and another bivariate normal distribution for category \(b\) (represented by the ellipses in Figure 2). All the standard deviations are 1.0, because the decision variables are normalized. The mean for category \(a\) is at \((-d_s'/2, -d_m'/2)\) and for category \(b\) is at \((d_s'/2, d_m'/2)\). The DVC for category \(a\) is \(\rho_a\) and for category \(b\) is \(\rho_b\). Formally, the two distributions are given by

\[
bvn_b(z_s, z_m) = \phi\left(\frac{z_s - d_s}{2}, \frac{z_m - d_m}{2}; \rho_b\right) \quad (3)
\]

\[
bvn_a(z_s, z_m) = \phi\left(\frac{z_s + d_s}{2}, \frac{z_m + d_m}{2}; \rho_a\right), \quad (4)
\]

where \(\phi(x, y; \rho)\) is the standard bivariate normal density function

\[
\phi(x, y; \rho) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1 - \rho^2)}\right]. \quad (5)
\]

Once the discriminability and criterion values have been estimated from Equations 1 and 2, the DVCs can be estimated from four proportions that are directly available from the subject and model responses to the stimuli in each category. For category \(a\) they are the proportion of times the subject and the model both responded \(A\), \(p(\text{AA}|a)\); the proportion of times the subject responded \(A\) and the model responded \(B\), \(p(\text{AB}|a)\); the proportion of times the subject responded \(B\) and the model responded \(A\), \(p(\text{BA}|a)\); and the proportion of times the subject and the model both responded \(B\), \(p(\text{BB}|a)\). Similarly, for category \(b\), the proportions are \(p(\text{AA}|b)\), \(p(\text{AB}|b)\), \(p(\text{BA}|b)\) and \(p(\text{BB}|b)\). These proportions correspond to the volume under the bivariate normal distribution within the four quadrants defined by the two decision criteria. As long as no more than one of the four proportions is zero, then it is straightforward to obtain the maximum-likelihood estimate of the DVC using Equations 3–5 (see Appendix). This estimate of the DVC is Pearson’s tetrachoric correlation coefficient (Pearson, 1900; Stuart & Ord, 1991) applied in the SDT framework.

Maximum-likelihood estimation of the DVC from these four proportions is appropriate when only the categorical responses \((A\ or \ B)\) are available, as when estimating DVCs between two subjects (or within the same subject on different occasions). However, for many (but not all) models, a decision variable is explicitly computed for each stimulus presentation. In this case, more reliable maximum-likelihood estimates can be obtained by directly using the values of the model’s decision variable (see Appendix). However, we emphasize that DVCs can be measured even for a model that does not produce an explicit decision variable (e.g., some neural-network models).

The distributions of the decision variables for the subject and model (or another subject) are the marginal distributions of the bivariate distribution, and hence the DVCs do not necessarily depend on the discriminabilities and decision criteria estimated with basic
The ellipses in Figure 2 represent a moderate positive correlation. If the decision variables were uncorrelated, then the ellipses would become circles. If the decision variables were negatively correlated, the ellipses would be rotated 90°. In principle, all of these correlations are possible for the same discriminabilities and decision criteria. For example, if a model and subject (or two subjects) are using different cues to perform a task, then it is possible that a stimulus that is easy for the model will be hard for the subject, and vice versa.

An important implication of this fact is that a given DVC can be consistent with many different values of standard measures of the behavioral correlation between the sequences of responses of the model and subject, or between two subjects. For example, Figure 3 shows how four different behavioral-correlation measures (defined in the Appendix) vary with the subject’s decision criterion $\gamma_s$ and discriminability $d'_s$, while the decision-variable correlation $\rho_b$ for category $b$ is held fixed at 0.5 (yellow curves). The simplest correlation measure is the fraction of trials in which the subject and model are in agreement (blue curves). Perhaps the best-known and most common correlation measure for binary data is the phi coefficient (also introduced by Pearson; Cramer, 1946, p. 282), which is equivalent to Matthews’s (1975) correlation coefficient (orange curves). Another popular correlation measure is Cohen’s (1960) kappa coefficient (gray curves). Finally, a common measure in the neurophysiology literature is choice probability (Britten et al., 1996; green curves). In all cases, the behavioral correlations vary substantially while the decision-variable correlation is held fixed.

We note that all the curves in Figure 3A are flipped about the vertical axis at 0 if the measures are computed for category $a$ rather than category $b$.) In the extended SDT framework, the fundamental quantities are the discriminabilities, decision criteria, and DVCs. The overall accuracy levels and behavioral correlations depend in a rather complex way on all six of these fundamental quantities.

In other words, the extended SDT framework plays an analogous role in interpreting trial-by-trial correlations to the role that the standard SDT framework plays in interpreting accuracy. Standard SDT recognizes the fact that percent correct is not the best measure of accuracy because percent correct is dependent on the value of the decision criterion (every point on a given receiver operating characteristic curve corresponds to a different percent-correct value). Extended SDT recognizes the fact that previous trial-by-trial measures are not the best measures of trial-by-trial correlation because they are dependent on the value of decision criterion and the discriminability.
Simulations

To better understand the implications of the SDT framework in Figure 2, we simulated model and subject decisions in the well-known task of detecting a target in white pixel noise. On each simulated trial we presented (with equal probability) a background of Gaussian white noise, or a background of noise plus an added target (a small Gaussian), to the simulated subject and to the model. We took the model to be the template-matching observer, which is the ideal observer for this particular task (Peterson, Birdsall, & Fox, 1954; Green & Swets, 1974; Burgess, Wagner, Jennings, & Barlow, 1981). On each trial, this model observer takes the dot product of the image with a template having the shape of the target:

\[
R_m = f_m \cdot I = \sum_{x,y} f_m(x,y)I(x,y),
\]

where \(f_m(x,y)\) is the matched template and \(I(x,y)\) is the input image (Figure 4A). The template response \(R_m\) is the (unnormalized) decision variable, which will have a normal distribution with the same standard deviation for target-present and target-absent trials. The normalized decision variable \(Z_m\) (as in Figure 2) was obtained by subtracting half the mean of the target-present response and dividing by the standard deviation.

The simulated subjects were degraded versions of the ideal template-matching observer (Figure 4B). Two kinds of degradation were simulated. One kind was suboptimal pooling, which was obtained by removing some fraction of the pixels at random from the matched template (i.e., setting them to zero). The other kind of degradation was some level of simulated sensor noise—Gaussian white noise of standard deviation \(\sigma_0\) added to the input image. In other words, the simulated subjects could have inefficiency due to early noise, suboptimal pooling (sampling inefficiency), or both (e.g., Burgess et al., 1981; Burgess, 1985, 1994; Eckstein, Ahumada, & Watson, 1997; Lu & Dosher, 1999; Pelli & Farell, 1999). Later, we also consider the effect of decision noise.

The symbols in Figure 5 plot the maximum-likelihood estimates of the DVC as a function of the root-mean-square (RMS) contrast of the background image (based on 500 trials per condition). These estimates were obtained from the categorical responses of the ideal observer and simulated subject (see Equations A1–A12). The solid curves show the functions expected from the closed-form expression (Equations A33–A36) derived in the Appendix. The horizontal axis is the root-mean-square contrast of a Gaussian white-noise background. Each curve is for a different simulated subject. All the simulated subjects had a fixed neural (pixel) noise corresponding to a root-mean-square contrast of 0.05. The simulated subjects differed in what fraction of pixels in the ideal observer’s matched template they applied. The amplitude of the Gaussian target is set so that the discriminability of the ideal-observer model was fixed at 1.0. The solid curves show the functions expected from the closed-form expression (Equations A33–A36) derived in the Appendix.
with the fraction of remaining pixels in the subject’s template. This is also expected, because the subject’s decision variable depends only on the background and neural variability at a subset of pixel locations. If the neural noise is set to 0.0, then all the curves in Figure 5 become horizontal lines. The DVCs in these simulations are the same for the two stimulus categories (here, target present and absent), but as mentioned before, this need not be the case in general. The simulations in Figure 5 demonstrate that DVC should vary systematically with fundamental properties of the subject and model, confirming that the extended SDT framework should be useful for testing models (and characterizing individual differences).

An obvious question is how the reliability of the estimate of DVC varies with the number of trials in the experiment, and with other aspects of the data. In any real situation, the confidence intervals could be estimated using bootstrap methods, but simulations are useful to get some idea of what to expect. Figure 6A plots the standard deviation of the estimated DVC as a function of the number of trials, for different levels of discriminability, when the true DVC was set to 0.5. These estimates were based on using only the categorical responses (see Appendix). The levels of discriminability were set by varying the amplitude of the target, with the neural and decision noise set to zero (Figure 4B). The DVC was set by removing pixels from the ideal matched template to obtain the simulated subject’s template. (The relevant equations in the Appendix are A33–A38.) The results in Figure 6A make intuitive sense: As the level of discriminability goes up, there is less overlap of the distributions, and hence it takes more trials to reliably estimate the error proportions $p(AB|b)$, $p(BA|b)$, $p(AB|a)$, and $p(BA|a)$. For $d'$ values of 2.0 or below, the estimates of DVC should be useful once the number of trials reaches about 100.

Figure 6B plots the standard deviation of the estimated DVC as a function of the number of trials, for different values of the true DVC, with the target amplitude set so that the discriminability ($d'$) was fixed at 1.0. Interestingly, the standard deviation of the estimated DVC is not strongly dependent on the true correlation, at least over the range of 0.1 to 0.5 (the approximate range we observe in the demonstration experiment described later). In terms of percentage, the confidence interval shrinks nearly linearly with the true DVC.

Demonstration experiment

In addition to the simulations, we carried out and analyzed the data from a small experiment with two human subjects (the authors).

Methods

Psychometric functions were measured for detection of a Gaussian-shaped target having a standard-deviation parameter value of 4 arcmin, for three levels of Gaussian white-noise background contrast (0.05, 0.1, and 0.2 RMS) with a virtual pixel width of 2 arcmin ($4 \times 4$ display pixels per virtual pixel) and mean luminance of 48 cd/m². The amplitude of the Gaussian target was defined to be the value at the peak of the target divided by twice the mean luminance (i.e., max amplitude = 1.0). The target and noise were presented for 250 ms in a blocked single-interval forced-choice
task, with feedback. The three psychometric functions were measured three times for a total of 150 trials per target amplitude (total of 750 trials per psychometric function). The two subjects saw exactly the same stimuli, so that we could estimate between-subjects DVCs. In addition, we replicated the entire experiment with exactly the same trials a couple of weeks later (a double-pass experiment), so we could estimate within-subject DVC.

Results

Figure 7A plots the measured psychometric functions for the two observers, for each level of background contrast. Figure 7B plots the estimated thresholds (the target amplitude when \( d^2 \approx 1 \)). Consistent with previous studies (e.g., Burgess et al., 1981), the squared threshold amplitude increases linearly with background contrast power (the square of the RMS contrast).

Figure 8A–8C plots the measured DVCs between the two subjects, as a function of target amplitude for each level of background contrast. Figure 8D–8F plots the measured within-subject (double-pass) DVCs. The DVC increases with background contrast; the average between-subjects correlations are 0.215, 0.258, and 0.425, and within-subject correlations are 0.398, 0.568, and 0.595. This result is qualitatively consistent with the hypothesis that the internal noise in the subjects is uncorrelated, and that as background contrast increases, the spatial structure of the external noise increasingly determines the values of the subjects’ decision variables. The fact that the within-subject correlations are higher than the between-subjects correlations suggest that the templates (i.e., set of features) or internal noise levels of the two subjects are somewhat different.

Interestingly, there is a trend for the DVC to increase with target amplitude, especially when background contrast is high, and a trend for DVC to be greater for target-present than target-absent trials (see insets and schematic ellipses in Figure 2). This is not consistent with the family of model observers in Figure 4B, even though the model observers correctly predict that threshold energy should increase linearly with background energy (Figure 7B). Specifically, this family of models predicts that neither the contrast of the target nor its presence or absence should have an effect on the DVC. In simple simulations, we have found that these trends are predicted if the two subjects (or the same subject on two occasions) have different levels of intrinsic position uncertainty (a factor not included in Figure 4B; but see later).

In the Appendix we show that for the class of models in Figure 4B, the double-pass experiment provides an estimate of the correlation of the subject’s template to that of the ideal matched template. Specifically, the cosine similarity between the subject’s template and the ideal matched template equals the square root of the ratio of the subject’s efficiency to the double-pass DVC:

\[
s = \sqrt{\frac{\eta}{\rho_{dp}}},
\]

where the efficiency is defined as \( \eta = d_{subject}^2/d_{ideal}^2 \). Furthermore, we show that if (in addition) the late decision noise is negligible compared to the sensor noise, then the double-pass DVC provides an estimate

![Figure 7. Psychometric functions and thresholds for background root-mean-square contrasts of 0.05, 0.1, and 0.2. (A) Psychometric functions for two subjects. (B) Average amplitude threshold squared as a function of background contrast power.](http://arvojournals.org/)
of the ratio of the sensor-noise standard deviation to that of the external stimulus noise:

$$\frac{\sigma_0}{\sigma} = \sqrt{\frac{1 - \rho_{dp}}{\rho_{dp}}}.$$  \hspace{1cm} (8)

Figure 9A–9C plots the estimated correlations of the subjects’ templates to the ideal template, and Figure 9D–9F plots the estimated relative standard deviation of the subjects’ sensor noise (under the strong assumption that late decision noise is negligible). These plots illustrate how DVCs can provide constraints useful for estimating the components of multistage models like the one in Figure 4B.

The solid circles in Figure 10A–10C are the DVCs between the template-matching ideal observer and the subjects, as a function of the target contrast. As expected, the DVCs are lower than those within the subjects. In the Appendix we show that the family of model human observers in Figure 4B predicts that the DVC between the human and ideal observers should equal the square root of the efficiency:

$$\rho = \sqrt{\eta}.$$  \hspace{1cm} (9)

The open symbols in Figure 10A–10C are the square-root of the efficiency. Remarkably, this prediction appears to hold quite well.

However, as in Figure 8 there is a trend for the DVC to increase with target amplitude. Thus, the pattern of results is not consistent with the class of models in Figure 4B, unless it is extended to allow the template, sensor noise, or decision noise to vary with target contrast.

The demonstration experiment and various formulas derived in the Appendix show that computing DVCs can provide considerable additional power for testing models of performance, beyond the power provided by computing only detectabilities (i.e., psychometric functions and thresholds) and decision biases.
Estimating model parameters

A strong model will not only predict the accuracy of the subjects in each condition but will also predict the trial-by-trial responses—the more accurate the prediction of the trial-by-trial responses in a condition, the larger the DVC. The best specific model in a parameterized family of models can be found by

Figure 9. Template cosine similarity and internal- to external-noise ratio. (A–C) Estimated cosine similarity of the subjects’ and ideal observer’s templates (Equation 7) as a function of target amplitude, for the three background contrasts. (D–F) Estimate of the ratio of internal sensor-noise level to external noise level (Equation 8) as a function of target amplitude, for the three background contrasts.

Figure 10. Decision-variable correlation (DVC) between subjects and ideal observer as a function of target amplitude, for the three levels of background root-mean-square contrast. The solid symbols are the DVCs averaged across the two subjects. The error bars are bootstrapped 67% confidence intervals. The open symbols show the DVCs predicted from the subjects’ efficiencies, for the family of models in Figure 4B. The inset shows, for each background contrast, the DVC (solid bars) and square root of efficiency (open bars) averaged across target amplitude.
finding the parameters that maximize the predicted accuracy and DVC. The most efficient procedure for doing this will likely depend on the family of models and the specific task the subjects are performing.

As an example, consider the present detection-in-noise task and the family of models in Figure 4B, generalized to include the possibility of position uncertainty. This family of models assumes that the subject applies a template on each trial and compares the responses of that template to a decision criterion. Note that in this section we regard Figure 4B as the model observer (not the simulated subject), and thus its responses, criterion, template, and detectability will have the subscript \( m \).

DVCs can be exploited to estimate the subject’s template and level of position uncertainty. Using classification-image techniques (Ahumada, 1996; Murray, 2011), Eckstein, Beutter, Pham, Shimozaki, and Stone (2007) found that subjects’ templates for detecting Gaussian targets in noise have approximately the shape of a difference-of-Gaussians function. Thus, here we assumed that the model observer’s template \( f_m(x, y) \) is a difference-of-Gaussians function with three parameters, two standard deviations and a relative weight for center and surround: \( \sigma_c, \sigma_s, w \) (the center is scaled by \( w \) and the surround by \( 1 - w \)). The model observer in Figure 4B does not take into account the effects of position uncertainty, which is likely to be an important factor (Solomon, 2002; Tjan & Nandy, 2006). Thus, we generalized the model observer to allow inclusion of position uncertainty. Specifically, the template is now applied over a spatial region \( \Omega \) with a radius of \( r \) pixels, and the model observer’s unnormalized decision variable is the maximum of these template responses:

\[
R_m = \max_{x,y} f_m \cdot I. \tag{10}
\]

To estimate the subjects’ template and level of uncertainty, we set the model observer’s sensor and decision noise to zero (i.e., the response to a given stimulus is deterministic) and varied the difference-of-Gaussians parameters and the uncertainty parameter to maximize the average DVC of the model observer and subjects.

The inset in Figure 11 shows the estimated templates and uncertainty regions for the two subjects; the estimated model parameters are given in the figure caption. The solid black circles in the panels show DVCs between the model observer and subjects. For comparison, the open black circles show the DVCs between the ideal observer and subjects from Figure 10, and the solid gray circles show the double-pass DVCs from Figure 8D–8F. Remarkably, the model observer is able to predict the trial-by-trial responses of the subjects as accurately as, or more accurately than, presenting the same stimulus again to the subjects. This is possible

because the model observer does not have sensor or decision noise, whereas the subject has independent samples of sensor and decision noise on both trials with the same stimulus. In the Appendix (Equation A51), we show that the upper limit for DVC, given the model observers in Figure 4B, is the square root of the double-pass DVC (open gray circles in Figure 11).

This example demonstrates the potential usefulness of DVCs for estimating model parameters. It also demonstrates the important principle that, in general, models that optimally predict an organism’s trial-by-trial responses should not include a representation of neural noise within the organism. Thus, there is a fundamental difference between predicting trial-by-trial responses and predicting other aspects of performance (e.g., neural noise is needed to explain why the subjects’ double-pass DVCs are not 1.0). This principle also holds for models of neural response—in general, an optimal model of trial-by-trial neural responses will be deterministic and will predict the responses better than presenting the same stimuli twice.

### Discussion

Experiments measuring behavioral performance in humans and other animals often use a task where the stimulus on each trial is drawn from one of two categories, and the subject is required to identify the category. Such tasks are particularly common in the areas of perception, memory, and decision making. In all of these areas, the well-known framework of signal-detection theory (SDT) is commonly used to estimate both the subject’s intrinsic ability to identify the category \( (d') \) and the subject’s decision bias \( (\gamma) \). We described a straightforward extension of SDT that can provide additional power for testing models and quantifying individual differences. Responses to the same stimuli are recorded from a model and a subject, from two subjects, or from the same subject on two occasions. From these responses, six quantities are estimated (see Figure 2): two discriminability (identifiability) values, two decision-criterion values, and two decision-variable correlations (DVCs; one for each stimulus category). The discriminability and decision-criterion values are estimated in the usual way (e.g., from hits and false alarms). The DVCs can be estimated from the pairs of categorical responses to the stimuli within each category, or (for many models) from the subject’s categorical responses and the model’s continuous decision-variable values.

DVC is a principled measure of the relationship between the trial-by-trial responses made by a subject and a model, two subjects, or the same subject on multiple occasions. We found that other common
measures of correlation in trial-by-trial behavior do not properly take into account the effects of discriminability and decision bias on the estimation of correlation (see Figure 3). We then demonstrated and evaluated the extended framework in simulations and in an experiment involving detection of a Gaussian-shaped target in Gaussian white noise. Even this very simple experiment yielded a rich set of theoretical and experimental results. We found that it was possible to derive a new set of predicted relations for the well-known class of models in Figure 4. These relations allowed us to reject this class of models for detection in white noise. We then estimated parameters for an extended model and found that it is possible to predict trial-by-trial performance as well as or better than presenting the same stimuli twice to the subject. Finally, we showed that deterministic models are best for predicting trial-by-trial performance, while models with sources of internal noise are necessary to explain other aspects of performance.

**Measuring DVCs**

Estimation of DVCs requires that, in a given condition (e.g., a specific background contrast and target amplitude), the two categories contain multiple stimuli (exemplars). In many published studies, a given condition consists of exactly the same two stimuli presented multiple times (e.g., discrimination experiments with fixed stimuli). In these cases, all of the trial-to-trial variation in the subjects’ responses is due to unknown internal factors (e.g., neural noise). Because these experiments vary the stimuli so little, they provide less information for testing hypotheses about the underlying neural computations (and neural noise). In general, we argue it is much better (and more realistic compared to real-world tasks) to include some stimulus variation within conditions.

To compute DVCs between two subjects, it is necessary to present the same stimuli to both subjects. This is easy to do and is potentially useful for characterizing individual differences in trial-by-trial behavior. For example, if a number of subjects are tested on the same stimuli, then those with similar neural processing should have higher DVCs.

To compute DVCs on the same subject, it is necessary to present the same stimuli (at least) twice to the same subject (a double-pass experiment). As we saw, double-pass experiments can also be very useful within the context of a specific class of models. In addition, they can provide a useful benchmark. For many models, the maximum possible DVC between the model and subject is the square root of the DVC between subject and herself. However, a double-pass experiment is a special design that, it could be argued, is inefficient because it doubles the length of the experiment.

In contrast, computing DVCs between a subject and a model requires nothing more than keeping track of the specific stimuli and responses on each trial of the experiment. For stimuli that are randomly generated or sampled this can be as simple as describing the random-number generator, saving the random-number seed, and saving a short sequence of random numbers (this makes it possible to verify later that the random-number generator is behaving as expected and hence that the regenerated stimuli are correct). Even if an experimenter is not immediately interested in trial-by-trial behavior,
saving this information and making it available would make it possible for others to compute DVCs or other measures of the trial-by-trial performance.

Computing DVCs between subjects and models is useful for testing models. In principle, all six quantities in the extended SDT framework can be estimated for every stimulus condition in an experiment. For example, in the small experiment described here there were 15 conditions (5 target amplitudes \( \times \) 3 background contrasts). The 90 estimated quantities (\( d_0 \), criteria, and DVCs) provide a strong set of constraints on models of detection in white noise. Of course, in some experimental designs there may not be sufficient trials to reliably estimate all six quantities for each condition. Even so, estimates can be averaged across conditions and can still be useful for testing hypotheses. For example, in the demonstration experiment, computing the average DVC (across the five target amplitudes) for each of the three background contrasts would have been sufficient to provide a strong test of the model in Figure 4B—the average correlations approximately equal the square root of the average efficiencies.

The extended SDT framework is most directly applied in experiments with blocked conditions, where a number of responses are collected for each pair of fixed stimulus conditions. For experiments where the categories are continuously varying (e.g., adaptive psychophysical procedures) it should still be possible to apply the framework by binning the stimuli into approximately fixed categories.

Robustness to violations of the equal-variance assumption

The extended SDT framework makes the standard SDT assumption that the decision variables for the two categories, in a given condition, are normally distributed with equal variance (Figures 1 and 2). This assumption is likely to hold only approximately. We carried out calculations to get some idea about how robust the estimated quantities are to the violations of this assumption. Specifically, we varied the ratio of the variances of the two normal distributions. We set \( \sigma_a = \sqrt{2/(1 + k)} \) and \( \sigma_b = \sqrt{2k/(1 + k)} \), where \( k \) is the ratio of the two variances, with the sum of the variances equal to 2.0 (i.e., the average variance is 1.0). We set \( u_a = -0.5 \), \( u_b = +0.5 \), and the criterion at the cross point between the distributions. We assumed this is true for both the subject and model (or both subjects). For a true DVC of 0.5 (which was the same for both categories) we first computed the proportion of responses in the four quadrants for each category (see Figure 2) and then estimated \( (d_0^a, d_0^m, \gamma_0^a, \gamma_0^m, \rho_a, \rho_b) \) assuming equal variances. Figure 12 shows how the estimated discriminability, criterion, and DVC varied as a function of \( k \) (e.g., \( k = 0.5 \) to 2). As the ratio of the variances deviates from 1.0 there are modest systematic deviations of the estimated discriminability and criterion from the actual values. Interestingly, there is no systematic error in the estimate of the DVC. Overall, the extended SDT framework appears to be fairly robust to violations of the equal-variance assumption.

Relationship between DVC, choice probability, and choice correlation

Choice probability (CP) and choice correlation (CC) are popular measures of the trial-by-trial relationship between measured neural responses and behavioral responses (Britten et al., 1996; Haefner et al., 2013; Pitkow et al., 2015). CP and CC are typically computed...
only in neurophysiological studies, but they could potentially be computed when evaluating the trial-by-trial predictions of a behavioral model. Specifically, one could regard the value of a model’s decision variable on a trial as a hypothetical neural response. CP is defined to be the probability of correctly discriminating the subject’s behavioral responses, given the neural responses, using the standard SDT framework. In the Appendix we derive the formal relationship between DVC and CP. CC is an approximately linear function of CP, and is equivalent to computing a DVC computed under certain assumptions. Under these assumptions there is a one-to-one relationship between the measures, but like the other measures of trial-by-trial correlations, CP and CC are dependent on the subject’s discriminability and decision criterion (see Figure 3). Thus, DVC is a more general and flexible measure for evaluating models and should also be better for evaluating the relationship between neural activity and behavior (see also Michelson et al., 2017; Seidemann & Geisler, in press).

Keywords: signal detection theory, visual cognition, spatial vision, psychophysics, masking

Acknowledgments

The work was supported by NIH grants EY11747 and EY024662. We thank Johannes Burge and Robbe Goris for helpful suggestions.

Commercial relationships: none.
Corresponding author: Wilson S. Geisler.
Email: w.geisler@utexas.edu.
Address: Center for Perceptual Systems and Department of Psychology, University of Texas at Austin, Austin, TX, USA.

References


Appendix

Maximum-likelihood estimation of decision-variable correlation (DVC): Categorical responses

Here we consider maximum-likelihood estimation of the DVC when all the available responses are categorical. This is appropriate when computing DVCs between two subjects, the same subject on different occasions, or a model and a subject when the model does not generate a continuous decision variable on each trial (e.g., some kinds of neural-network models).

In the extended SDT framework, the proportion of trials in the four quadrants defined by the two decision criteria correspond to the volume under the bivariate normal distribution in those quadrants. Given Equations 3 and 5, the four integrals for category \( b \) are

\[
f(B, B|b) = \int_{-\infty}^{\gamma_m} \int_{-\infty}^{\gamma_n} b_{vn}(z_s, z_m) \, dz_s \, dz_m \tag{A1}\]

\[
f(A, A|b) = \int_{-\infty}^{\gamma_m} \int_{-\infty}^{\gamma_n} b_{vn}(z_s, z_m) \, dz_s \, dz_m \tag{A2}\]

\[
f(A, B|b) = \int_{-\infty}^{\gamma_m} \int_{-\infty}^{\gamma_n} b_{vn}(z_s, z_m) \, dz_s \, dz_m \tag{A3}\]

\[
f(B, A|b) = \int_{-\infty}^{\gamma_m} \int_{-\infty}^{\gamma_n} b_{vn}(z_s, z_m) \, dz_s \, dz_m \tag{A4}\]

and for category \( a \) they are

\[
f(B, B|a) = \int_{-\infty}^{\gamma_m} \int_{-\infty}^{\gamma_n} b_{vn}(z_s, z_m) \, dz_s \, dz_m \tag{A5}\]

\[
f(A, A|a) = \int_{-\infty}^{\gamma_m} \int_{-\infty}^{\gamma_n} b_{vn}(z_s, z_m) \, dz_s \, dz_m \tag{A6}\]

\[
f(A, B|a) = \int_{-\infty}^{\gamma_m} \int_{-\infty}^{\gamma_n} b_{vn}(z_s, z_m) \, dz_s \, dz_m \tag{A7}\]

\[
f(B, A|a) = \int_{-\infty}^{\gamma_m} \int_{-\infty}^{\gamma_n} b_{vn}(z_s, z_m) \, dz_s \, dz_m \tag{A8}\]

The log likelihood of the data given the DVC for category \( b \) is

\[
\ln L(\rho_b) = p(B, B|b) \ln f(B, B|b) + p(A, A|b) \ln f(A, A|b) + p(A, B|b) \ln f(A, B|b) + p(B, A|b) \ln f(B, A|b), \tag{A9}\]

and for category \( a \) it is
\[
\ln L(\rho_a) = p(B, B[a]) \ln f(B, B[a]) \\
+ p(A, B[a]) \ln f(A, A[a]) \\
+ p(A, B[a]) \ln f(A, B[a]) \\
+ p(B, A[a]) \ln f(B, A[a]),
\]

where \([p(B, B[b]), p(A, B[a]), p(A, B[b]), p(B, A[b])]\) are the proportions of observations in the four quadrants for category \(b\) and \([p(B, B[a]), p(A, A[a]), p(A, B[a]), p(B, A[a])]\) are the proportions of observations in the four quadrants for category \(a\). The maximum-likelihood estimates of the two DVCs are obtained by finding the maxima of the log likelihoods given by Equations A9 and A10:

\[
\hat{\rho}_b = \arg\max_{\rho_b} \ln L(\rho_b) \quad (A11)
\]

\[
\hat{\rho}_a = \arg\max_{\rho_a} \ln L(\rho_a) \quad (A12)
\]

For those who use MATLAB we note that

\[
\int \int bvn_b(z_x, z_m) dz_x dz_m = \text{mvncdf}(x_l, x_h, u_b, \Sigma_b) \quad (A13)
\]

with \(x_l = [x_l, y_l], x_h = [x_h, y_h], u_b = [d_l/2, d_m/2], \) and

\[
\Sigma_b = \begin{bmatrix} 1 & \rho_b \\ \rho_b & 1 \end{bmatrix}.
\]

Similarly,

\[
\int \int bvn_a(z_x, z_m) dz_x dz_m = \text{mvncdf}(x_l, x_h, u_a, \Sigma_a) \quad (A14)
\]

with \(x_l = [x_l, y_l], x_h = [x_h, y_h], u_a = [-d_l/2, -d_m/2], \) and

\[
\Sigma_a = \begin{bmatrix} 1 & \rho_a \\ \rho_a & 1 \end{bmatrix}.
\]

Inspection of Figure 2 shows that if at least three of the four quadrants contain a nonzero proportion of trials, then some constraints are placed on the possible DVC. However, if there are just two nonzero proportions, then the DVC is essentially unconstrained. This happens when one or more of the decision criteria does not sufficiently overlap the joint distribution. One case where this can happen is if the subject’s accuracy is too high. Another case is if the subject is highly biased. For example, if the subject makes no false alarms, then the DVC cannot be measured for the target-present category (although it may still be measurable for the target-absent category). In general, to estimate at least one of the DVCs, both the model and the subject (or the two subjects) must make some errors.

Note that the proportions in the four quadrants defined by the decision criteria must sum to 1.0 for each category. Thus, there are a total of six independent measured quantities, which exactly matches the number of parameters to estimate (2 \(d^s\) values + 2 decision criteria + 2 DVCs).

### Maximum-likelihood estimation of DVC: Model decision-variable responses

Here we consider maximum-likelihood estimation of the DVC when the continuous value of the decision variable of the model observer is available (e.g., when the decision variable is the log likelihood ratio of the two categories given the stimulus). This is the preferred method to estimate DVCs when the model decision variable is available, because it is generally a more reliable estimate (notice smaller error bars in Figures 10 and 12 than in Figure 8).

Let \(Q_i\) be the categorical (binary) response of the subject on trial \(i\), and let \(Z_i\) be the value of the model observer’s normalized decision variable on trial \(i\) (i.e., its value in the space of Figure 2). For category \(a\), the log likelihood of all the subject’s responses given the value of the model’s decision variable is given by

\[
\ln L(\rho_a) = \sum_{i=1}^{n} \ln p_a(Q_i|Z_i),
\]

where

\[
p_a(Q_i|Z_i) = \begin{cases} p_a(A|Z_i) & \text{if } Q_i = A \\ 1 - p_a(A|Z_i) & \text{if } Q_i = B \end{cases}
\]

with

\[
p_a(A|Z_i) = \Phi \left( \frac{\gamma_s + \frac{d}{2} - \rho_a Z_i - \rho_a \frac{d}{2}}{\sqrt{1 - \rho_a^2}} \right).
\]

Note that \(\Phi(x)\) is the standard normal integral function (in MATLAB, \(\Phi(x) = \text{normcdf}(x)\)). The maximum-likelihood estimate of the DVC is found by maximizing Equation A15 over all possible values of the DVC:

\[
\hat{\rho}_a = \arg\max_{\rho_a} \ln L(\rho_a). \quad (A18)
\]

We now derive Equation A17. From the definition of conditional probability (and Figure 2), we have

\[
p_a(A|y) = \frac{1}{\phi(y)} \int_{-\infty}^{x_c} \phi(x, y) dx, \quad (A19)
\]

where \(\phi(x, y)\) is the standard bivariate normal density function (Equation 5) and \(\phi(y)\) is the standard univariate normal density function, where

\[
x = x_c + d/2, \quad y = Z_i + d/2, \quad \text{and} \quad x_c = \gamma_s + d/2.
\]

From Equations 5 and A19 we have
Completing the square in the exponent and simplifying gives

\[
p_a(A|y) = \Phi \left( \frac{x_c - \rho_{aw}}{\sqrt{1 - \rho^2}} \right).
\]  \hspace{1cm} (A21)

Substituting for \( y \) and \( x_c \) gives Equation A17.

Equations analogous to A15 to A18 hold for estimating the DVC for category \( b \). The main difference is that

\[
p_b(A|Z_i) = \Phi \left( \frac{y_s - \frac{d_x}{2} - \rho_p Z_i + \rho_p \frac{d_x}{2}}{\sqrt{1 - \rho_p^2}} \right).
\]  \hspace{1cm} (A22)

Finally, note that when the values of the model decision variable are available, then DVCs can be estimated as long as the subject makes some errors (no errors need be made by the model).

**DVC between optimal and suboptimal observers**

In the simulations, the stimulus on target-present trials is

\[ I(x, y) = aT(x, y) + N(x, y) + \bar{I}, \]  \hspace{1cm} (A23)

and on target-absent trials it is

\[ I(x, y) = N(x, y) + \bar{I}, \]  \hspace{1cm} (A24)

where \( N(x, y) \) is white Gaussian noise with a variance of \( \sigma^2 \) and \( \bar{I} \) is the mean luminance. The target shape \( T(x, y) \) is defined to have a maximum of 1.0. The amplitude of the target is \( a \). We simulated and analyzed two kinds of observers: the ideal template-matching observer and suboptimal model subjects (see Figure 4).

The ideal observer’s matched template \( f_m(x, y) \) is defined to be the target shape normalized by its energy,

\[ f_m(x, y) = \frac{T(x, y)}{\sqrt{T \cdot T}}, \]  \hspace{1cm} (A25)

where \( \cdot \) represents the dot-product operator defined in Equation 6. Thus, the unnormalized decision-variable value on target-present trials is

\[ R_m = aT \cdot f_m + f_m \cdot N + f_m \cdot \bar{I}, \]  \hspace{1cm} (A26)

and on target-absent trials it is

\[ R_m = f_m \cdot N + f_m \cdot \bar{I}. \]  \hspace{1cm} (A27)

The model subject’s template \( f_s(x, y) \) is an arbitrary template, and we assume that the simulated subject has some additive independent neural noise \( N_0(x, y) \) with variance \( \sigma^2_0 \) and some late decision noise \( N_d \) with variance \( \sigma^2_d \) (see Figure 4). Thus, the simulated subject’s unnormalized decision-variable value on target-present trials is given by

\[ R_s = aT \cdot f_s + f_s \cdot N + f_s \cdot \bar{I} + f_s \cdot N_0 + N_d, \]  \hspace{1cm} (A28)

and on target-absent trials it is given by

\[ R_s = f_s \cdot N + f_s \cdot \bar{I} + f_s \cdot N_0 + N_d. \]  \hspace{1cm} (A29)

The correlation between the decision variables is

\[ \rho = \frac{\text{COV}(R_m, R_s)}{\text{SD}(R_m)SD(R_s)}. \]  \hspace{1cm} (A30)

The covariance and standard deviations are unaffected by additive constants (the target and mean background luminance), and thus the DVC (in the present case) is the same for both categories of trials (target present and target absent):

\[ \rho = \rho_a = \rho_b = \frac{\text{COV}(f_m \cdot N, f_s \cdot N + f_s \cdot N_0 + N_d)}{\text{SD}(f_m \cdot N)SD(f_s \cdot N + f_s \cdot N_0 + N_d)}. \]  \hspace{1cm} (A31)

From the definition of covariance and standard deviation, we have

\[ \rho = \frac{\sigma^2 f_s \cdot f_m}{\sqrt{\sigma^2 f_m \cdot f_m} \sqrt{\sigma^2_0 f_s + \sigma^2 f_s \cdot f_s + \sigma^2_d}}. \]  \hspace{1cm} (A32)

Thus,

\[ \rho = \frac{s}{\sqrt{r_0 + r_d}}, \]  \hspace{1cm} (A33)

where \( s \) is the cosine similarity of the subject and model templates,

\[ s = \frac{f_s \cdot f_m}{\sqrt{f_s \cdot f_s} \sqrt{f_m \cdot f_m}}; \]  \hspace{1cm} (A34)

\( r_o \) is the ratio of the internal-noise variance to the external-noise variance,

\[ r_o = \frac{\sigma^2_0}{\sigma^2}; \]  \hspace{1cm} (A35)

and \( r_d \) is the ratio of the decision-noise variance to the external-noise variance scaled by 1 over the subject template energy,

\[ r_d = \frac{\sigma^2_d}{\sigma^2 f_s \cdot f_s}. \]  \hspace{1cm} (A36)

We also note that
\[ d_m = \frac{aT \cdot f_m}{\sigma \sqrt{f_m \cdot f_m}} \quad (A37) \]

and

\[ d_s = \frac{aT \cdot f_s}{\sqrt{\sigma^2 f_s \cdot f_s + \sigma_0^2 f_s \cdot f_s + \sigma_d^2 (T \cdot f_m)^2}}. \quad (A38) \]

Therefore, the efficiency of the simulated subject is

\[ \eta = \frac{\sigma^2 f_m \cdot f_m}{\sigma^2 f_s \cdot f_s + \sigma_0^2 f_s \cdot f_s + \sigma_d^2 (T \cdot f_m)^2}. \quad (A39) \]

Dividing top and bottom by \( (T \cdot T)^2 \) yields

\[ \eta = \frac{\sigma^2 f_m \cdot f_m}{\sigma^2 f_s \cdot f_s + \sigma_0^2 f_s \cdot f_s + \sigma_d^2 (f_m \cdot f_m)^2}. \quad (A40) \]

and so

\[ \eta = \frac{\sigma^2}{\sigma^2 + \sigma_0^2 + \sigma_d^2 (f_m \cdot f_m)} \left( \frac{f_m \cdot f_m}{f_s \cdot f_s} \right)^2. \quad (A41) \]

\[ \eta = \frac{s^2}{1 + r_0 + r_d}. \quad (A42) \]

Thus, this simulated subject satisfies the strong constraints that the DVCs be the same for both stimulus categories, and the DVC equals the square root of the efficiency (compare Equations A33 and A42):

\[ \rho = \sqrt{\eta}. \quad (A43) \]

In other words, for this class of subject models the DVC with the ideal observer is completely determined by the discriminabilities. Note that this relation holds for every specific background and target contrast. Thus, it holds for the more general class of models where the neural noise, subject template, and decision noise vary as a function of background and target contrast.

**DVC between the same suboptimal observers**

A useful technique for estimating the relative effect of internal and external noise on an observer’s performance is the double-pass technique, where the subject is presented the same stimuli twice during the experiment (Swets, Shipley, McKey, & Green, 1959; Spiegel & Green, 1981; Burgess & Colborne, 1988). The data from double-pass experiments can also be analyzed by estimating the DVC of the subject with herself in the two passes. For the class of model human observers considered here, the first- and second-pass responses are given by

\[ R_s(1) = f_s \cdot N + f_s \cdot N_0(1) + N_D(1) \quad (A44) \]

\[ R_s(2) = f_s \cdot N + f_s \cdot N_0(2) + N_D(2). \quad (A45) \]

In other words, the sample of external noise \( N \) is the same, but the samples of neural noise and decision noise are different. The DVC in the double-pass experiment is

\[ \rho_{dp} = \frac{\text{COV}(R_s(1), R_s(2))}{\text{SD}(R_s(1)) \cdot \text{SD}(R_s(2))} = \frac{f_s \cdot f_s \sigma_N^2 + f_s \cdot f_s \sigma_d^2}{f_s \cdot f_s \sigma_N^2 + f_s \cdot f_s \sigma_d^2 + \sigma_d^2} \quad (A46) \]

\[ \rho_{dp} = \frac{1}{1 + r_0 + r_d}. \quad (A47) \]

Thus, for this class of subject model, the double-pass decision correlation provides an estimate of \( r_d + r_d \):

\[ r_0 + r_d = \frac{1 - \rho_{dp}}{\rho_{dp}}. \quad (A48) \]

If one assumes that there is no late decision noise, then the double-pass decision correlation provides an estimate of \( r_0 \), the ratio of the internal-noise to external-noise variance (see also Burgess & Colborne, 1988).

Comparison of Equations A42 and A47 shows that

\[ \eta = \rho_{dp} s^2. \quad (A49) \]

Thus, the data from the double-pass experiment also provide an estimate of the cosine similarity of the subject and model template,

\[ s = \sqrt{\frac{\eta}{\rho_{dp}}}. \quad (A50) \]

However, the double-pass experiment does not provide an estimate of the relative contributions of early neural noise and late decision noise.

If the subject is using the optimal template (or the same template as a model observer; i.e., \( f_m = f_s \)), and has only neural and decision noise, then Equations A33 and A47 imply that

\[ \rho_{dp} = \rho^2. \quad (A51) \]

**Behavioral-correlation measures**

Here we define the three behavioral-correlation measures plotted in Figure 3.

The fraction of trials where responses are in agreement is given by

\[ a = p_BB + p_AA. \quad (A52) \]

The phi coefficient (Matthew’s correlation coefficient) is given by
Cohen’s kappa coefficient is given by
\[ \kappa = \frac{p_0 - p_e}{1 - p_e}, \]
where \( p_0 = \frac{p_{BB} + p_{AA}}{p_{BB} + p_{BA} + p_{AA} + p_{AB}} \) and \( p_e = \frac{(p_{BB} + p_{BA})(p_{BB} + p_{AB}) + (p_{AB} + p_{AA})(p_{BA} + p_{AA})}{p_{BB} + p_{BA} + p_{AB} + p_{AA}} \).

**Relationship between DVC and choice probability (CP)**

CP is the probability of correctly discriminating the subject’s behavioral responses, given the value of a model’s decision variable, using the standard SDT framework. In neurophysiology experiments this decision variable is computed from the neural responses on a trial, and in the present case it is computed from the stimulus on a trial. In the notation of Figure 2, CP can be computed for each stimulus category. For category \( a \) the two relevant distributions are \( p_a(z_m|A) \) and \( p_a(z_m|B) \). For category \( b \) they are \( p_b(z_m|A) \) and \( p_b(z_m|B) \). CP is then usually defined as the area under the receiver operating characteristic (ROC) for the pair of distributions. Equivalently, one can compute the choice probability from the hits and false alarms for the optimally placed criterion. Consider trials for category \( a \). From Bayes’s rule we have

\[
p_a(z_m|A) = \frac{p_a(A|z_m)p_a(z_m)}{p_a(A)} \quad \text{(A55)}
\]

\[
p_a(z_m|B) = \frac{p_a(B|z_m)p_a(z_m)}{p_a(B)} \quad \text{(A56)}
\]

Using Equations A17 and 2, and the fact that \( \Phi(-x) = 1 - \Phi(x) \), we have

\[
p_a(z_m|A) = \frac{1}{\Phi\left(\frac{d_A}{2} + \gamma_s\right)} \Phi\left(\frac{\gamma_s + \frac{d_A}{2} - \rho_a z_m - \rho_a \frac{d_m}{2}}{\sqrt{1 - \rho_a^2}}\right) \phi\left(z_m + \frac{d_m}{2}\right) \quad \text{(A57)}
\]

\[
p_a(z_m|B) = \frac{1}{\Phi\left(-\frac{d_A}{2} - \gamma_s\right)} \Phi\left(-\frac{\gamma_s - \frac{d_A}{2} + \rho_a z_m + \rho_a \frac{d_m}{2}}{\sqrt{1 - \rho_a^2}}\right) \phi\left(z_m + \frac{d_m}{2}\right) \quad \text{(A58)}
\]

The proportions of hits and false alarms are given by

\[
\phi = \frac{p_{BB}p_{AA} - p_{BA}p_{AB}}{\sqrt{(p_{BB} + p_{BA})(p_{AB} + p_{AA})(p_{AB} + p_{BB})(p_{AA} + p_{BA})}}. \quad \text{(A53)}
\]

Finally, the CP in the current single-interval task is given by

\[
\text{CP}_a = \Phi\left[\Phi^{-1}(p_{A|a}) - \Phi^{-1}(p_{B|a})\right] \quad \text{(A61)}
\]

The area under the ROC in a single-interval task actually gives the percent correct that would have
occurred in a two-interval task. This slightly inflated accuracy is obtained by dividing by $\sqrt{2}$ rather than 2 in Equation A60.

Analogous equations hold for category $b$:

$$p_b(z_m|A) = \frac{1}{\Phi(-\frac{d_s}{2} + \gamma_s)} \Phi\left(\frac{\gamma_s - \frac{d_s}{2} - \rho_b z_m + \rho_b \frac{d_s}{2}}{\sqrt{1 - \rho_b^2}}\right) \phi\left(z_m - \frac{d_m}{2}\right) \tag{A62}$$

$$p_b(z_m|B) = \frac{1}{\Phi(\frac{d_s}{2} - \gamma_s)} \Phi\left(-\gamma_s + \frac{d_s}{2} + \rho_b z_m - \rho_b \frac{d_s}{2}\right) \phi\left(z_m - \frac{d_m}{2}\right) \tag{A63}$$

$$p_{B|b} = \int_{\gamma_s}^\infty p_b(z_m|B)dz_m \tag{A64}$$

$$p_{A|b} = \int_{\gamma_s}^\infty p_b(z_m|A)dz_m \tag{A65}$$

$$\text{CP}_b = \Phi\left[\Phi^{-1}(p_{B|b}) - \Phi^{-1}(p_{A|b})\right] \tag{A66}$$

The values of CP depend on the subject’s discriminability and decision criterion, and on the DVC, but not on the model’s discriminability; thus, these equations can be simplified by setting $d_m = 0$. If there is a desire to express the DVC as an equivalent CP, then one method is to estimate the DVC as described here, substitute the value into these equations, and then set the subject’s criterion to align with the peak of the subject’s decision-variable distribution, which has the same effect as setting $\gamma_s = 0$ and $d_s = 0$. Figure A1 shows the relationship between CP and DVC when $\gamma_s = 0$ and $d_s = 0$, based on hits and false alarms, and based on the area under the ROC.