The Relationship Between Signal Detection Theory and Bayesian Decision Theory

Authors: Hu Xiao, Hu Xiao

Date: 2023-06-30T19:48:46+00:00

Abstract

Signal detection theory is widely employed to explain individuals' decision-making processes across various types of cognitive tasks. However, a significant limitation of classical signal detection theory is its difficulty in further elucidating the intrinsic psychological mechanisms that underlie the process by which individuals set their reporting criteria. This article, from the perspective of Bayesian decision theory, provides an in-depth examination of individuals' decision rules in signal detection tasks. Initially, based on Bayes'theorem, this article introduces the fundamental tenets of Bayesian decision theory. Subsequently, it explores how Bayesian decision theory accounts for the decision rules of an ideal observer, as well as the divergence between individuals' decision outcomes and those of the ideal observer in practical signal detection tasks. Following this, the article investigates the distinctions between classical signal detection theory and Bayesian decision theory within the unequal-variance signal detection model. Finally, this article briefly presents empirical research evidence supporting Bayesian decision theory.

Full Text

The Relationship Between Signal Detection Theory and Bayesian Decision Theory

HU Xiao
Faculty of Psychology, Beijing Normal University, Beijing 100875

Abstract: Signal detection theory (SDT) has been widely applied to explain decision-making processes in various cognitive tasks. However, a significant limitation of classical SDT is its difficulty in elucidating the underlying psychological mechanisms by which individuals set their response criteria. This article examines decision rules in signal detection tasks from the perspective of Bayesian decision theory (BDT). We first introduce the fundamental concepts of BDT

based on Bayes' theorem. Next, we discuss how BDT explains the decision rules of an ideal observer and characterizes deviations between actual participants and the ideal observer in empirical signal detection tasks. We then examine the differences between classical SDT and BDT in unequal-variance signal detection models. Finally, we briefly review empirical research supporting BDT.

Keywords: signal detection theory; Bayesian decision theory; prior probability; likelihood function; response criterion

1 Introduction

Signal detection theory (SDT) is one of the most widely applied computational models in experimental psychology. Since psychologist John A. Swets and his collaborators systematically introduced SDT into psychology (Green & Swets, 1966; Tanner & Swets, 1954), researchers have extensively used SDT models to explain the underlying mechanisms of perception, memory, reasoning, and other psychological processes (Banks, 1970; Mamassian, 2016; Wixted, 2020). A search for "signal detection theory"in the PsycArticles and PsycInfo databases yields over 4,000 publications, with 500 appearing between 2020 and 2022 alone, demonstrating that SDT not only holds an important place in the history of psychological research but remains highly active today.

In SDT-based experimental designs, researchers present participants with two types of stimuli: "signal"stimuli and "noise"stimuli. Participants must determine which stimuli are signals and which are noise (Wickens, 2001). For example, in an auditory task, participants might hear white noise alone or white noise with a specific tone added as a signal; they must judge on which trials the signal appears (Egan et al., 1959). Similarly, in recognition memory tasks, participants first learn and memorize a list of words, after which researchers present "old"words (studied items) as signals and "new"words (unstudied items) as noise; participants must judge whether each word is old or new (Mickes et al., 2007; Wixted, 2007). Typically, SDT assumes that signal and noise stimulus intensities follow normal distributions, with the signal distribution having a higher mean than the noise distribution. The difference between these means is called the discriminability index (d'), which reflects an individual' s ability to discriminate signals from noise—higher d' indicates greater discriminability (Wickens, 2001).

A central question in SDT concerns how individuals decide which stimuli are signals and which are noise. Classical SDT posits that individuals set a response criterion C directly on the stimulus intensity axis. If the current stimulus intensity exceeds C, the individual judges it as a signal; otherwise, it is judged as noise (Wixted, 2020). Figure 1 illustrates an SDT model example where the noise distribution has a mean of 0, the signal distribution has a mean of d' , and both distributions have a standard deviation of 1. When a stimulus' s intensity exceeds C, it is judged as a signal; otherwise, it is judged as noise. In signal detection tasks, responses can be classified into four types: hits (signal correctly

identified as signal), false alarms (noise incorrectly identified as signal), misses (signal incorrectly identified as noise), and correct rejections (noise correctly identified as noise) (Wickens, 2001). The probabilities of these four outcomes are represented by the areas under the normal distributions for signal or noise to the left or right of C (see Figure 1).

The advantage of SDT is its ability to separate discriminability index d' and response criterion C from accuracy data, allowing researchers to examine how task difficulty (d') and response bias (C) independently affect performance (Wickens, 2001). However, classical SDT (i.e., models using C directly to reflect decision criteria) has an important limitation: it cannot adequately explain the underlying psychological mechanisms of the decision process. While classical SDT simply assumes individuals compare current stimulus intensity directly to criterion C, it cannot explain why C is set at a particular location or why C varies across individuals or experimental conditions (Glanzer et al., 2009, 2019).

In fact, when SDT was first introduced, Swets and collaborators began using Bayesian decision theory (BDT) to explain decision processes in signal detection tasks (Green & Swets, 1966). Recently, the relationship between BDT and SDT has received increasing attention (Fleming & Daw, 2017; Glanzer et al., 2019; Maloney & Zhang, 2010). BDT proposes that individuals observe stimulus intensity and complete a Bayesian inference process to decide whether the stimulus is signal or noise (Fleming & Daw, 2017; Maloney & Zhang, 2010; Pouget et al., 2016). Compared to classical SDT, BDT provides a deeper theoretical explanation of decision-making in signal detection tasks (Glanzer et al., 2019; Lau, 2007).

This article examines individual decision rules within the SDT framework from a BDT perspective. We first introduce BDT' s basic concepts—how individuals make decisions through Bayesian inference in signal detection tasks. Next, we address the "ideal observer" problem in SDT, explaining how BDT accounts for deviations between real participants and ideal observers within equal-variance SDT models. We then discuss unequal-variance SDT models and examine differences between classical SDT and BDT in these models. Finally, we review empirical evidence supporting BDT.

2 Bayesian Decision Theory

Bayesian decision theory is founded on Bayes' theorem, which derives from the formula for joint probability (胡传鹏 et al., 2018). Consider two events, A and B. The probability of both occurring is called joint probability, denoted P(A, B). The joint probability can be expressed as:

$$P(A, B) = P(B|A)P(A) \tag{1}$$

This means the probability of both A and B occurring equals the probability of A occurring, P(A), multiplied by the conditional probability of B given A,

P(B|A). Here, P(A) is also called marginal probability—the probability of an event occurring regardless of other events—while P(B|A) is the conditional probability of B occurring given that A has occurred. The joint probability can also be written alternatively as:

$$P(A, B) = P(A|B)P(B) \tag{2}$$

Based on equations (1) and (2), we obtain:

$$P(A|B)P(B) = P(B|A)P(A) \tag{3}$$

Equation (3) is typically written as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{4}$$

Equation (4) is the general form of Bayes' theorem. Initially proposed to characterize mathematical relationships between marginal and conditional probabilities, psychologists soon recognized that Bayes' theorem could model human reasoning and decision-making (Kersten et al., 2004; Lau, 2007; Wickens, 2001). For instance, equation (4) can explain how individuals infer the probability of event A from observing event B. Here, P(A) is called prior probability—an individual' s prior belief about A' s probability before observing B. P(A|B) is posterior probability—reflecting the updated belief about A after observing B. In Bayesian inference, individuals update their beliefs about A' s probability from the prior P(A) to the posterior P(A|B) based on observing B.

Meanwhile, P(B|A) is called the likelihood function, reflecting how likely B is given that A occurs. Both prior probability and likelihood function play crucial roles: higher prior probability for A leads to higher posterior probability, and greater likelihood of B given A (i.e., larger likelihood function value) also leads to higher posterior probability for A after observing B.

According to BDT, in signal detection tasks individuals observe each stimulus' s intensity (denoted x) and complete a Bayesian inference process to determine the posterior probability that the current stimulus is signal or noise (Burgess, 1985; Fleming & Daw, 2017; Wickens, 2001). The following formula shows how individuals update the probability that the current stimulus is a signal (denoted S) based on intensity x:

$$P(S|x) = \frac{P(x|S)P(S)}{P(x)} \tag{5}$$

This means individuals have a prior belief P(S) about the probability of a signal before observation. After observing intensity x, they update this to posterior

probability P(S|x). Similarly, updating the probability that the stimulus is noise (denoted N) follows:

$$P(N|x) = \frac{P(x|N)P(N)}{P(x)} \tag{6}$$

BDT posits that individuals decide whether to judge the stimulus as signal or noise based on these posterior probabilities. If the posterior probability of signal exceeds that of noise, the stimulus is judged as signal; otherwise, it is judged as noise (Burgess, 1985; Fleming & Daw, 2017; Lau, 2007). Since the posterior probabilities of signal and noise sum to 1, individuals use P(S|x) = 0.5 as the decision criterion: judge as signal if P(S|x) > 0.5, and as noise if P(S|x) < 0.5.

In equations (5) and (6), the marginal probability of observing intensity x, P(x), is identical. Specifically, P(x) is calculated as a weighted average of the likelihood functions—P(x|S) and P(x|N)—according to the prior probabilities of signal and noise, P(S) and P(N):

$$P(x) = P(x|S)P(S) + P(x|N)P(N) \tag{7}$$

Dividing equation (5) by equation (6) eliminates P(x):

$$\frac{P(S|x)}{P(N|x)} = \frac{P(x|S)P(S)}{P(x|N)P(N)} \tag{8}$$

Equation (8) reveals that Bayesian decision-making can be viewed as a process based on the ratio of posterior probabilities: if the ratio P(S|x)/P(N|x) exceeds 1, the stimulus is judged as signal; otherwise, as noise. According to Bayes'theorem, this posterior ratio is jointly determined by the likelihood ratio P(x|S)/P(x|N) and the prior probability ratio P(S)/P(N).

Since stimulus intensity x follows probability distributions (typically assumed normal) under signal and noise conditions, the likelihood function reflects the relative probability of observing intensity x given signal or noise—equivalent to the probability density of x in the signal or noise distribution (i.e., the ordinate value at x in the normal distribution). When the probability density of x in the signal distribution, P(x|S), is larger, or when it is smaller in the noise distribution, P(x|N), individuals are more likely to judge the stimulus as signal, and vice versa. Importantly, individuals may not know the true distributions of signal and noise. For instance, they may not know the actual means and standard deviations of these distributions (Lau, 2007). To complete Bayesian inference, individuals must subjectively estimate the distribution shapes and use these estimates to assess the likelihood of the current intensity x (Lau, 2007).

In some cases, individuals may have no prior preference between signal and noise. For example, before a signal detection experiment, participants may have no bias

about whether stimuli on a computer screen are signals or noise, simply considering them equally likely. Here we can assume equal prior probabilities: P(S) = P(N) = 0.5. In this case, the posterior probability ratio equals the likelihood ratio, so Bayesian decisions depend entirely on likelihood values. However, prior probabilities are not always 0.5 (Wickens, 2001). Some individuals may have subjective preferences, believing signals or noise are more likely. In such cases, Bayesian decisions based on posterior probability ratios are influenced by both likelihood ratios and prior probability ratios: when individuals believe signals are more probable a priori, they are more likely to judge stimuli as signals, and vice versa.

According to BDT, both likelihood function values and prior probabilities used in Bayesian inference come from subjective estimates. When do these subjective estimates yield optimal decisions? How do changes in these estimates affect decision processes? To address these questions, we next examine the "ideal observer" perspective in signal detection tasks to explore when decisions are optimal and how to explain deviations between real participants and ideal observers.

3 The Ideal Observer

Classical SDT posits that individuals set a response criterion C on the stimulus intensity axis and compare the current stimulus intensity x to C to decide whether the stimulus is signal or noise. Responses are classified as hits, false alarms, misses, and correct rejections, with hits and correct rejections considered correct responses, while false alarms and misses are errors. An interesting question arises: where should C be placed to maximize accuracy? In SDT, an individual who sets the optimal criterion to maximize accuracy is called an ideal observer (Wickens, 2001).

To calculate where an ideal observer would place C, we must first determine how C affects accuracy. Consider the SDT model shown in Figure 1, where signal and noise distributions have equal variance (both equal to 1)—this is called an equal-variance SDT model (Wickens, 2001). In this model, when the stimulus is a signal, accuracy (hit rate, HR) equals the area under the signal distribution to the right of C:

$$HR = 1 - \Phi(C|\mu = d', \sigma = 1) \tag{9}$$

Here, $\Phi(C|\mu, \sigma)$ represents the cumulative distribution function of a normal distribution with mean μ and standard deviation σ, equivalent to the area under the curve to the left of C. Similarly, when the stimulus is noise, accuracy (correct rejection rate, CRR) equals the area under the noise distribution to the left of C:

$$CRR = \Phi(C|\mu = 0, \sigma = 1) \tag{10}$$

Assuming signal stimuli occur with probability P(St) and noise stimuli with probability P(Nt) in the experiment (where t indicates these are the true probabilities, to distinguish them from subjective priors used in Bayesian decision-making), overall accuracy Pcorrect is:

$$P_{correct} = P(St)\cdot HR + P(Nt)\cdot CRR = P(St)\cdot[1-\Phi(C|\mu = d', \sigma = 1)] + P(Nt)\cdot\Phi(C|\mu = 0, \sigma = 1) \tag{11}$$

To find the C that maximizes Pcorrect, we first derive its first derivative:

$$\frac{dP_{correct}}{dC} = P(Nt)\cdot\phi(C|\mu = 0, \sigma = 1) - P(St)\phi(C|\mu = d', \sigma = 1) \tag{12}$$

Here, $\phi(C|\mu, \sigma)$ represents the probability density at intensity C in a normal distribution with mean μ and standard deviation σ. The ideal observer' s criterion, Cideal, should be set where this derivative equals zero:

$$P(Nt)\cdot\phi(C_{ideal}|\mu = 0, \sigma = 1) - P(St)\phi(C_{ideal}|\mu = d', \sigma = 1) = 0 \tag{13}$$

When individuals judge all stimuli with intensity above Cideal as signals and all below as noise, accuracy is maximized (Wickens, 2001). Using the normal probability density function, we can solve for Cideal' s location:

$$\frac{\phi(C_{ideal}|\mu = d', \sigma = 1)}{\phi(C_{ideal}|\mu = 0, \sigma = 1)} = \frac{P(Nt)}{P(St)} \tag{14}$$

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{(C_{ideal}-d')^2}{2}} = \frac{P(Nt)}{P(St)}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{C_{ideal}^2}{2}} \tag{15}$$

$$C_{ideal} = \frac{d'}{2} - \frac{\ln[P(Nt)/P(St)]}{d'} \tag{16}$$

In the equal-variance model, when signals and noise occur with equal probability (P(St) = P(Nt) = 0.5), Cideal equals d' /2—the ideal observer places the criterion exactly midway between the signal and noise distribution means, at their intersection (Wickens, 2001).

We can also understand the ideal observer' s decision rule from a BDT perspective. Equation (13) can be rewritten as:

$$\frac{\phi(C_{ideal}|\mu = d', \sigma = 1)P(St)}{\phi(C_{ideal}|\mu = 0, \sigma = 1)P(Nt)} = 1 \tag{17}$$

This closely resembles equation (8) describing Bayesian decision-making. In equation (17), the ratio of true probabilities P(St)/P(Nt) resembles the prior

probability ratio in Bayesian decisions, and the ratio of probability densities at Cideal resembles the likelihood ratio. BDT indicates that when individuals know the objective prior probabilities and true distribution shapes (i.e., true likelihood values), Bayesian decision-making ensures ideal observer performance—maximizing accuracy (Burgess, 1985). Combining equations (8) and (17), when stimulus intensity equals Cideal, the posterior probability ratio (product of likelihood ratio and prior ratio) equals 1. When intensity exceeds Cideal, the posterior ratio exceeds 1, leading to a "signal" judgment; when below Cideal, it is less than 1, leading to a "noise" judgment. This Bayesian decision rule matches the ideal observer's rule in classical SDT.

However, empirical analyses show that while real participants' criteria approximate the ideal observer's (Knill, 1998; Legge et al., 2002; Stretch & Wixted, 1998), they rarely place the criterion exactly at Cideal from equation (16), indicating systematic deviations (Lau, 2007; Wickens, 2001). BDT explains these deviations through two sources: First, participants' subjective estimates of prior probabilities may differ from objective probabilities. Figure 2A illustrates a case where objective probabilities are equal, but participants subjectively believe signals are more likely. Based on their subjective priors, they set a criterion left of the distributions' midpoint to maximize accuracy, causing deviation from the ideal observer. Second, participants' subjective estimates of distribution shapes may differ from objective distributions, causing biased likelihood estimates. In Figure 2B, participants correctly estimate the noise distribution mean but overestimate the signal distribution mean, leading them to set a criterion at the midpoint of their subjective estimates, which deviates from the true ideal criterion (Lau, 2007; Wickens, 2001). Thus, BDT provides a deeper explanation than classical SDT for why criteria are set at specific locations, linking them to subjective estimates of probability and distribution shape.

These conclusions derive from the equal-variance model in Figure 1. In this model, although BDT offers deeper theoretical explanation than classical SDT, they are mathematically equivalent. Even if participants' subjective estimates of priors and likelihoods differ from true values, as long as they believe signal and noise variances are equal, both theories yield mathematically equivalent decision processes: judge as signal when x exceeds criterion C, and as noise when x is below C (Glanzer et al., 2009, 2019; Wickens, 2001). However, in real tasks, signal and noise variances are not always equal—these are called unequal-variance SDT models (Mickes et al., 2007). Here, the relationship between classical SDT and BDT becomes more complex.

4 Unequal-Variance Signal Detection Models

To simplify models, researchers often assume equal variances for signal and noise distributions. However, this assumption does not always hold empirically (Green & Swets, 1966; Wixted, 2020). In recognition memory tasks, for instance, the signal distribution (old words) has a significantly larger standard deviation than the noise distribution (new words), with a typical ratio of about 1.25:1 (Mickes

et al., 2007; Rotello, 2017). Wixted (2007) proposed that old words have higher intensity in recognition memory because they were learned during the study phase. Only if all old words were learned equally would the signal variance match the noise variance. Since learning varies across items, this variability causes the signal distribution' s variance to exceed the noise distribution' s. Thus, equal-variance models, while theoretically simple, may oversimplify signal detection processes, whereas unequal-variance models may better capture real task dynamics.

Not all signal detection tasks require unequal-variance models. In two-alternative forced-choice (2AFC) tasks, where a signal (e.g., old word) and noise (e.g., new word) are presented simultaneously and participants must select the signal, equal-variance models can quantify performance even when variances differ (Wickens, 2001). As this is not directly relevant to our topic, we will not elaborate further; interested readers may consult existing literature on 2AFC tasks (Macmillan & Creelman, 2004; Wickens, 2001).

We now examine decision-making in unequal-variance models from a BDT perspective. Figure 3 shows an unequal-variance model where the noise distribution has mean 0 and standard deviation $\sigma_N = 1$, while the signal distribution has mean d' and standard deviation $\sigma_S \neq 1$ (here $\sigma_S > 1$, as in recognition memory). To simplify, we assume equal true probabilities for signal and noise (0.5 each) and that participants know these priors and likelihoods perfectly. With equal priors, decisions depend entirely on the likelihood ratio: according to equation (8), judge as signal when the likelihood ratio exceeds 1, and as noise otherwise. However, in unequal-variance models, signal and noise distributions intersect at two points (C1 and C2). When x is less than C1 or greater than C2, the probability density is higher in the signal distribution, leading to a "signal" judgment; when x is between C1 and C2, the density is higher in the noise distribution, leading to a "noise" judgment. Thus, a specific likelihood ratio corresponds to two different criteria on the intensity axis.

Glanzer et al. (2009) derived the mathematical relationship between likelihood ratio and criterion C in unequal-variance models. They first computed the likelihood ratio at intensity C, then took its natural logarithm (denoted λ):

$$\lambda = \ln\left[\frac{\phi(C|\mu = d', \sigma = \sigma_S)}{\phi(C|\mu = 0, \sigma = 1)}\right] \tag{18}$$

Since the logarithm is monotonically increasing, λ's maxima/minima correspond to those of the likelihood ratio. Further analysis shows (Glanzer et al., 2009) that λ relates quadratically to C (see Figure 4). When $\sigma_S > 1$, λ has a minimum (λ^*); when $\sigma_S < 1$, it has a maximum (λ^*). Only when $\lambda = \lambda^*$ does a unique criterion C* exist on the intensity axis; otherwise, each λ value corresponds to two criteria. When $\lambda = 0$ (likelihood ratio = 1), the two criteria C1 and C2 divide the intensity axis into three regions. If $\sigma_S > 1$, the region between C1

and C2 is judged as noise and the outer regions as signal; if $\sigma_S < 1$, the pattern reverses.

Classical SDT, in contrast, assumes that even in unequal-variance models, individuals simply set a single criterion C, judging as signal when x exceeds C and as noise otherwise (Mickes et al., 2007; Wickens, 2001). Thus, classical SDT and BDT make different predictions for unequal-variance models, and their decision outcomes are no longer equivalent (unlike in equal-variance models).

These differences manifest in receiver operating characteristic (ROC) curves, which plot hit rate (HR) against false alarm rate (FAR) while holding objective distributions constant (Wickens, 2001). In classical SDT, varying C changes both HR and FAR. In BDT (assuming perfect knowledge of priors and likelihoods), varying signal/noise probabilities changes prior probabilities, affecting the likelihood ratio where posterior ratio equals 1 (i.e., the critical λ value), which in turn changes criteria locations and thus HR and FAR. Plotting all possible HR-FAR pairs yields the ROC curve.

In equal-variance models, both theories (being mathematically equivalent) produce ROC curves above the diagonal when d' > 0, with larger area under the curve for higher d' (Figure 5A). However, in unequal-variance models, classical SDT can produce ROC curves that fall below the diagonal even when d' > 0. For example, when $\sigma_S > 1$, HR may be lower than FAR when both approach 1 (Figure 5B); when $\sigma_S < 1$, this occurs when both approach 0 (Figure 5C). In contrast, BDT produces ROC curves that always remain above the diagonal (Figure 5D), a conclusion that can be proven mathematically (Wickens, 2001).

These differences are more apparent in zROC curves (Macmillan & Creelman, 2004). zROC curves transform HR and FAR to z-scores:

$$z_{HR} = \Phi^{-1}(HR|\mu = 0, \sigma = 1) \tag{19}$$

$$z_{FAR} = \Phi^{-1}(FAR|\mu = 0, \sigma = 1) \tag{20}$$

where $\Phi^{-1}(p|\mu = 0, \sigma = 1)$ is the inverse cumulative distribution function of the standard normal distribution, converting probabilities (0 to 1) to z-scores ($-\infty$ to $+\infty$). Plotting zHR against zFAR yields the zROC curve (Figure 6). When the zROC curve lies above the line y = x, hit rate exceeds false alarm rate. In classical SDT, zHR and zFAR have a linear relationship with slope equal to σ_N/σ_S (Wickens, 2001). With σ_N typically set to 1, when $\sigma_S > 1$ the slope is less than 1, causing the zROC line to intersect y = x on the right side and produce HR < FAR at high rates. When $\sigma_S < 1$, the slope exceeds 1, intersecting on the left side and producing HR < FAR at low rates (Figure 6A). In BDT, zHR and zFAR have a curvilinear relationship that always remains above y = x (Figure 6B) (Macmillan & Creelman, 2004).

Since classical SDT and BDT predict different ROC and zROC patterns in unequal-variance models, one could theoretically examine empirical ROC/zROC curves and extreme HR/FAR data to determine which theory better explains real decision-making. For example, in recognition memory tasks, one could increase old word frequency and decrease new word frequency to induce a liberal criterion (HR and FAR both near 1). If HR < FAR, this would support classical SDT; if HR always exceeds FAR, this would support BDT.

However, such designs are difficult to implement. In classical SDT' s unequal-variance model, HR < FAR occurs only when both approach extreme values (0 or 1), and the difference is small (Figures 5B and 5C). Sampling error severely affects observed HR-FAR differences, making it difficult to accurately estimate true differences (Glanzer et al., 2019; Macmillan & Creelman, 2004). Are there alternative experimental designs to compare these theories?

5 Empirical Evidence Supporting Bayesian Decision Theory

Researchers have proposed that two-condition experiments can compare classical SDT and BDT (Glanzer et al., 2009, 2019; Semmler et al., 2018; Stretch & Wixted, 1998). In these within-subject designs, the same participants complete two signal detection tasks differing only in difficulty, with identical requirements. For example, participants might complete two recognition memory tasks where the only difference is study time per item (longer time = easier task; Glanzer et al., 2009). Participants judge each stimulus as signal or noise and provide confidence ratings on a Likert scale. Glanzer et al. (2019) emphasized that participants must know which task condition each trial belongs to. Therefore, the two difficulty conditions should be presented in separate blocks; if using a mixed-list design, trials must be clearly marked (e.g., by color) to indicate their condition.

Likelihood ratio theory explains decision-making in two-condition experiments (Glanzer et al., 2009, 2019; Semmler et al., 2018). Based on BDT, it comprises two assumptions: (1) Likelihood ratio invariance—the critical likelihood ratio threshold remains constant across difficulty conditions. (2) True likelihood ratio—participants know the true distribution shapes and use actual likelihood ratios for decisions. Note that the true likelihood ratio assumption is stronger than general BDT, which allows subjective likelihood ratios to differ from true values (Lau, 2007; Wickens, 2001). Likelihood ratio theory posits that both assumptions hold simultaneously.

From this theory, Glanzer et al. (2009, 2019) predicted three phenomena: mirror effect, variance effect, and zROC length effect. Likelihood ratio theory readily explains these, while classical SDT struggles to account for all three. Empirical studies across cognitive domains (perception, memory, reasoning, mental rotation) consistently demonstrate these effects (Glanzer et al., 2009, 2019; Hilford et al., 2015, 2019; Semmler et al., 2018), supporting BDT. Due to space constraints, we focus on the zROC length effect.

The zROC length effect was first reported by Stretch and Wixted (1998), who reanalyzed data from Ratcliff et al. (1994). In that study, participants completed two recognition memory tasks (easy vs. difficult). In the easy condition, participants studied each old word for 3 s; in the difficult condition, only 1 s. During the recognition test, participants used a 6-point scale: 1 = "sure new," 2 = "probably new," 3 = "maybe new," 4 = "maybe old," 5 = "probably old," 6 = "sure old." Thus, participants made old/new judgments and reported confidence. In classical SDT, confidence rating tasks generalize binary judgments: an n-point scale corresponds to (n-1) criteria on the intensity axis (Wickens, 2001). As Figure 7A shows, when intensity falls below C1, participants respond "1" ; between C1 and C2, "2" ; and so on. Stretch and Wixted found that in the easy condition (higher d'), all criteria shifted inward, shortening distances between them (Figure 7A). They replicated this finding in a new experiment.

Glanzer et al. (2009, 2019) reformulated this using zROC curves. They computed (n-1) HR-FAR pairs by varying criteria across the n-point scale. For a 6-point scale, treating responses 1 as "noise" and 2-6 as "signal" yields one HR-FAR pair; treating 1-2 as "noise" and 3-6 as "signal" yields another; and so on. Plotting zROC curves from these pairs revealed that the easy condition' s zROC curve was significantly shorter (Figure 7B), meaning HR and FAR varied less across criteria—criteria were closer together. This is the zROC length effect.

Likelihood ratio theory naturally derives this effect from the invariance assumption. According to BDT (equation (8)), confidence judgments correspond to setting (n-1) thresholds on posterior probability ratios (denoted $\beta_1, \beta_2, ..., \beta_{n-1}$). If the posterior ratio falls below β_1, participants respond "1" ; between β_1 and β_2, "2" ; etc. (Green & Swets, 1966). Glanzer et al. (2009, 2019) assumed that when tasks differ only in difficulty, these thresholds β remain constant across conditions. Since posterior ratio equals prior ratio times likelihood ratio, and subjective prior estimates are approximately constant (Fleming & Daw, 2017; Wickens, 2001), participants effectively set constant thresholds on likelihood ratios across conditions. This is the likelihood ratio invariance assumption.

Deriving the zROC length effect from this assumption in equal-variance models: with constant likelihood ratio thresholds, the log-likelihood ratio (λ) thresholds are also constant across conditions. Using the normal density function (see equations (14) and (15)), the relationship between λ and criterion C is (Glanzer et al., 2009):

$$\lambda = \ln\left[\frac{\phi(C|\mu = d', \sigma = 1)}{\phi(C|\mu = 0, \sigma = 1)}\right] = d'C - \frac{{d'}^2}{2} \tag{21}$$

When λ is constant across conditions, decreasing task difficulty (increasing d') causes criteria C to contract around d' /2, shortening distances between them—producing the zROC length effect. Glanzer et al. (2009) showed this holds even in unequal-variance models, provided λ remains constant. While classical SDT could ad hoc explain criteria contraction as direct adjustment on the intensity

axis, it cannot specify the underlying mechanism. Thus, the zROC length effect supports BDT. Glanzer et al. (2009) similarly derived mirror and variance effects, which empirical studies confirm.

However, likelihood ratio theory' s true likelihood ratio assumption—that participants know true distribution shapes—has been questioned. Some doubt whether real participants possess such knowledge (Balakrishnan & Ratcliff, 1996; Criss & McClelland, 2006). Semmler et al. (2018) responded that knowledge accumulates through lifelong learning; even if participants cannot precisely estimate means and standard deviations, they learn how distribution separation varies with task difficulty (Turner et al., 2011; Wixted & Gaitan, 2002). This remains largely theoretical and requires future investigation.

6 Conclusion

This article examined decision-making in signal detection tasks through Bayesian inference. Compared to classical SDT, BDT provides deeper explanations of the mental mechanisms underlying signal-noise discrimination. However, BDT involves many parameters (subjective estimates of priors and likelihoods) that cannot all be estimated from hit and false alarm rates alone. For instance, criterion shifts could reflect changes in either prior or likelihood estimates, which are difficult to disentangle empirically (Fleming & Daw, 2017; Lau, 2007). While likelihood ratio theory simplifies BDT by assuming subjective likelihood ratios approximate true values (Glanzer et al., 2009, 2019; Semmler et al., 2018), this assumption is contested. Therefore, classical SDT (and its criterion C) remains a useful tool for data analysis.

The BDT discussed here does not fully capture decision mechanisms. For example, BDT assumes decisions depend solely on posterior probability ratios, but real decisions are also influenced by rewards and penalties. A driver approaching an intersection must quickly judge traffic light color (green = signal, red = noise). A miss (judging green as red) has minor consequences (waiting longer), but a false alarm (judging red as green) is catastrophic (causing accidents). Thus, drivers may adopt a strict criterion, 倾向于判断为红灯, independent of perceptual information. Green and Swets (1966) noted that when outcomes have different utilities, the goal may shift from maximizing accuracy (ideal observer) to maximizing reward or minimizing punishment, making decisions depend on both posterior probabilities and outcome consequences.

Additionally, BDT views confidence judgments as setting thresholds on posterior probability ratios (Glanzer et al., 2009, 2019; Green & Swets, 1966). However, recent work suggests posterior probabilities from stimulus intensity do not fully capture confidence. The information used for stimulus decisions may differ from that used for confidence judgments: some stimulus information may be lost in confidence reports, while additional reasoning may enrich them (Hu et al., 2021; Jang et al., 2012; Maniscalco & Lau, 2012; Shekhar & Rahnev, 2021). Fleming and Daw (2017) extended BDT to incorporate these differences, but others argue

confidence mechanisms may not fully conform to BDT even with this extension (Adler & Ma, 2018; Li & Ma, 2020). Whether confidence judgments follow Bayesian principles remains debated and requires further research.

胡传鹏, 孔祥祯, Wagenmakers, E.-J., Ly, A., 彭凯平. (2018). 贝叶斯因子及其在 JASP 中的实现. 心理科学进展, 26(6), 951–965.

Adler, W. T., & Ma, W. J. (2018). Comparing Bayesian and non-Bayesian accounts of human confidence reports. *PLoS Computational Biology, 14*(11), e1006572.

Balakrishnan, J. D., & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance, 22*(3), 615–633.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74*(2), 81–99.

Burgess, A. (1985). Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America A, 2*(9), 1498–1507.

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language, 55*(4), 447–460.

Egan, J., Schulman, A. I., & Greenberg, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *The Journal of the Acoustical Society of America, 31*(6), 768–773.

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. *Psychological Review, 124*(1), 91–114.

Glanzer, M., Hilford, A., Kim, K., & Maloney, L. T. (2019). Generality of likelihood ratio decisions. *Cognition, 191*, 103931.

Glanzer, M., Hilford, A., & Maloney, L. T. (2009). Likelihood ratio decisions in memory: Three implied regularities. *Psychonomic Bulletin & Review, 16*(3), 431–455.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Hilford, A., Glanzer, M., Kim, K., & Maloney, L. T. (2019). One mirror effect: The regularities of recognition memory. *Memory & Cognition, 47*(2), 266–278.

Hilford, A., Maloney, L. T., Glanzer, M., & Kim, K. (2015). Three regularities of recognition memory: the role of bias. *Psychonomic Bulletin & Review, 22*(6), 1646–1664.

Hu, X., Zheng, J., Su, N., Fan, T., Yang, C., Yin, Y., Fleming, S. M., & Luo, L. (2021). A Bayesian inference model for metamemory. *Psychological Review, 128*(5), 824–855.

Jang, Y., Wallsten, T. S., & Huber, D. E. (2012). A stochastic detection and retrieval model for the study of metacognition. *Psychological Review, 119*(1), 186–200.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology, 55*(1), 271–304.

Knill, D. C. (1998). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Research, 38*(11), 1683–1711.

Lau, H. C. (2007). A higher order Bayesian decision theory of consciousness. *Progress in Brain Research, 168*, 35–48.

Legge, G. E., Hooven, T. A., Klitz, T. S., Stephen Mansfield, J., & Tjan, B. S. (2002). Mr. Chips 2002: new insights from an ideal-observer model of reading. *Vision Research, 42*(18), 2219–2234.

Li, H.-H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nature Communications, 11*(1), 2004.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user' s guide*. London: Psychology press.

Maloney, L. T., & Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research, 50*(23), 2362–2374.

Mamassian, P. (2016). Visual confidence. *Annual Review of Vision Science, 2*(1), 459–481.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422–430.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review, 14*(5), 858–865.

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience, 19*(3), 366–374.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 763–785.

Rotello, C. M. (2017). Signal Detection Theories of Recognition Memory. In J. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference* (Second

Edition) (pp. 201–225). Academic Press.

Semmler, C., Dunn, J., Mickes, L., & Wixted, J. T. (2018). The role of estimator variables in eyewitness identification. *Journal of Experimental Psychology: Applied, 24*(3), 400–415.

Shekhar, M., & Rahnev, D. (2021). The nature of metacognitive inefficiency in perceptual decision making. *Psychological Review, 128*(1), 45–70.

Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1397–1410.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*(6), 401–409.

Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. *Psychological Review, 118*(4), 583–613.

Wickens, T. D. (2001). *Elementary signal detection theory.* New York: Oxford university press.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*(1), 152–176.

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(2), 201–233.

Wixted, J. T., & Gaitan, S. C. (2002). Cognitive theories as reinforcement history surrogates: The case of likelihood ratio models of human recognition memory. *Animal Learning & Behavior, 30*(4), 289–305.

Note: Figure translations are in progress. See original paper for figures.