

# Simple methods for evaluating and comparing binary experiments

Thomas A. Weber

Published online: 5 March 2009  
© Springer Science+Business Media, LLC. 2009

**Abstract** We consider a confidence parametrization of binary information sources in terms of appropriate likelihood ratios. This parametrization is used for Bayesian belief updates and for the equivalent comparison of binary experiments. In contrast to the standard parametrization of a binary information source in terms of its specificity and its sensitivity, one of the two confidence parameters is sufficient for a Bayesian belief update conditional on a signal realization. We introduce a confidence-augmented receiver operating characteristic for comparisons of binary experiments for a class of “balanced” decision problems, relative to which the confidence order offers a higher resolution than Blackwell’s informativeness order.

**Keywords** Classification problems · Decision making under uncertainty · Informativeness · Value of information

Where observation is concerned,  
Chance favors only the prepared mind.  
—LOUIS PASTEUR (1822–1895).<sup>1</sup>

---

<sup>1</sup> Lecture given on December 7, 1854, at the Université de Lille, France. See [Eves \(1988\)](#).

T. A. Weber (✉)  
Department of Management Science and Engineering, 442 Terman Engineering Center,  
Stanford University, Stanford, CA 94305-4026, USA  
e-mail: webert@stanford.edu

## 1 Introduction

A decision maker can sometimes obtain access to information in the form of a signal (also termed “experiment” or “information source”<sup>2</sup>) before choosing an optimal action under uncertainty. The efficient acquisition of information frequently requires finding the “best” among several such signals. The decision maker uses the information provided by his preferred signal to update the prior belief about the distribution of outcomes conditional on any given decision. In this article, we introduce a “confidence order,” based on likelihood ratios, for comparing different binary information sources. The confidence order, which is weaker than (i.e., implied by) Blackwell’s informativeness order, allows the equivalent comparison of information sources, as long as they “balance” the decision problem. The latter means that the expected gains from using the information source are similar across different states of the world. In addition to an order for different information sources, the confidence parameters provide a natural decomposition of a Bayes update with respect to signal realizations. The latter facilitates the Bayes belief update, which may allow decision makers to bypass the need for a complex mathematical model.

There are many practical problems in which an experiment has only two outcomes, one of which indicates the realization of an uncertain event  $X$ , and the other does the contrary. To fix ideas, let  $S$  and  $\bar{S}$  denote the possible outcomes of such a binary experiment. This experiment, relative to the event  $X$  and its complement  $\bar{X}$ , can be characterized by a *contingency table* as shown in Table 1, where TP, FP, FN, and TN represent the number of either obtained or expected “true positive,” “false positive,” “false negative,” and “true negative” joint realizations of the state of the world together with the experimental outcome.

Without loss of generality, one can normalize the four entries to sum up to 100%. The contingency table is then fully specified by two parameters in addition to the decision maker’s prior belief  $p = TP + FN$  about the probability of  $X$ . The two parameters can be chosen freely, as long as they, in conjunction with  $p$ , allow a one-to-one mapping to and from the entries of the contingency table. Much of the standard analysis of the performance of a binary experiment is conducted in terms of its *sensitivity* =  $TP/(TP+FN)$  and its *specificity* =  $TN/(TN+FP)$ , which describe the probability of  $S$  conditional on  $X$  and the probability of  $\bar{S}$  conditional on  $\bar{X}$ , respectively.<sup>3</sup> Different experiments can be compared in terms of their  $(1 - specificity, sensitivity)$ -tuples, corresponding to different points in the square  $[0, 1] \times [0, 1]$ . When a collection of

<sup>2</sup> In information theory, the classical model of communication is an information source generating a message that is sent as a signal over a channel (e.g., in the form of electrical impulses on an electric wire; cf. Shannon and Weaver 1948, p. 7). Thus, from the receiver’s point of view, the information conveyed by the information source arrives in the form of a signal. In other disciplines, such as economics, the terms “information source” and “signal” are therefore generally synonymous. In statistics, an experiment is an information source that produces messages in the form of “outcomes.” Here, we often use the term “signal” to describe an “experiment” or an “information source,” for the signal is what is ultimately observed by the receiver.

<sup>3</sup> Equivalently, as is usually done in statistics, one can analyze the performance of a binary signal in terms of its *type-I error* =  $FP/(FP + TN)$ , i.e., the probability of  $S$  conditional on  $\bar{X}$ , and its *type-II error* =  $FN/(FN + TP)$ , i.e., the probability of  $\bar{S}$  conditional on  $X$ .

**Table 1** Contingency table

	$X$	$\bar{X}$
$S$	TP	FP
$\bar{S}$	FN	TN

such points has been generated by varying a critical parameter within a given experimental setup, the family of experiments is generally referred to as a “receiver operating characteristic (ROC),” and a curve connecting the points is called an “ROC curve.” For example, when analyzing the performance of a breast-cancer detection technology such as mammography, a one-parameter family of experiments is naturally obtained by moving the diagnosis intensity from “malignant” over “suspicious” and “indeterminate” to “benign.” The ROC curve obtained for mammography with respect to these different thresholds can then be compared with the ROC curve of a different breast-cancer detection technology, such as magnetic resonance imaging (MRI), with respect to the same detection thresholds (Kriege et al. 2004; Leach et al. 2005). Such ROC-curve comparisons are useful in many other decision settings as well, ranging from fraud detection (Cavusoglu and Raghunathan 2004) and biomedicine (Lasko et al. 2005) to machine-learning applications such as email filtering (Hastie et al. 2001; Fawcett 2006).

Instead of adhering to the standard  $(1 - \text{specificity}, \text{sensitivity})$ -coordinates of the ROC space, as is common practice, we consider here an equivalent parametrization of binary experiments in terms of appropriate likelihood ratios  $\kappa$  and  $\lambda$ , which we term “confidence parameters.” For example, the confidence of a signal realization  $S$  as an indicator for the event  $X$  is simply the ratio between the conditional probability of  $S$  given  $X$  (i.e., its sensitivity) and the conditional probability of  $S$  given not  $X$  (i.e., one minus its specificity). The confidence parameters isolate positive and negative belief updates, which allow a graphical representation of a Bayes update as a function of the decision maker’s prior belief  $p$  about the probability of  $X$ . The appropriateness of the confidence parametrization for decision problems is highlighted by the fact that the decision maker’s optimal action depends only on the single confidence parameter that corresponds to the observed outcome of the experiment. Confidence parameters therefore naturally align decisions with signal realizations.

In addition to the evaluation of experiments, the confidence parametrization is also useful for the equivalent comparison of binary experiments in the ROC space. In the above example of breast-cancer testing, even though the ROC curve for MRI always lies above the ROC curve for mammography relative to the same detection thresholds (Leach et al. 2005), MRI cannot be considered a substitute for mammography for the following two reasons (Nemec et al. 2007): first, while for any given detection threshold MRI has a higher sensitivity, its specificity is generally lower than for mammography; second, the cost of an MRI is currently about ten times the cost of a mammogram. In Sect. 3, we show that an ordering based on an increase in the confidence parameters offers a higher resolution than a comparison with respect to Blackwell’s informativeness criterion, which requires dominance in both sensitivity and specificity. We demonstrate that the comparison in terms of confidence parameters

is valid for decision problems that are balanced in terms of the gains generated from making appropriate decisions relative to the information source, a condition which can be satisfied for signals with sufficiently small errors. The question of how to choose among information sources when their costs are different is discussed in Remark 6.

### 1.1 Related literature

[Bayes \(1764\)](#) fundamental contribution on how to update beliefs with observed evidence forms the basis for much of the modern theory of choice under uncertainty. Using a non-Bayesian, frequentist approach, [Hoel and Peterson \(1949\)](#) provide a solution to the standard classification problem. Their solution maximizes the probability of a correct classification and is based on earlier work by [Neyman and Pearson \(1933\)](#) and [Wald \(1939\)](#). [Birnbaum and Maxwell \(1960\)](#) examine the classification problem from a Bayesian viewpoint and, by tracing the “error probability pattern” for different experiments, introduce the notion of the now widely used ROC curves. The importance of likelihood ratios for the binary classification problem was noted by [Birnbaum \(1961\)](#), who examined the ordering of binary experiments using the tools developed by [Bohnenblust et al. \(1949\)](#) and [Blackwell \(1951, 1953\)](#). [Birnbaum \(1962\)](#) generalized these insights in a “likelihood principle,” which postulates that all evidence obtained by an experiment about an unknown event is contained in the corresponding likelihood function. This is in contrast to the frequentist view, in which once the decision maker stops collecting evidence, natural concerns arise about the probabilistic characteristics of the missing data (one could have obtained by continuing experimentation). Despite having been qualified by [Berger and Wolpert \(1984\)](#) and [Hill \(1987\)](#) among others, Birnbaum’s likelihood principle persists to this day. To obtain a visual tool for decision makers, we instead transform the ROC parameters of sensitivity and specificity for binary information sources to an equivalent set of likelihood ratios, which we term “confidence” parameters. Each likelihood ratio, together with the decision maker’s prior belief about the probability of an uncertain event, is completely sufficient for obtaining the Bayes update conditional on the corresponding signal realization. This allows the visualization of the evaluation of information sources in a Bayesian update diagram. Another reason for the appropriateness of likelihood ratios for updating beliefs is that Bayes’ rule becomes linear in the relevant realization-specific likelihood ratio when written in terms of odds or log-odds, as suggested by [Barnard \(1949\)](#) and [Lindley \(1964\)](#).

For the equivalent comparison of experiments, [Blackwell \(1951\)](#) introduces “informativeness,” which he shows to be equivalent to the notion of statistical sufficiency, as a meaningful partial order of information sources. If an experiment is strictly preferred in terms of its informativeness, then this experiment can be considered (at least weakly) more valuable for *any* decision problem, provided that a decision maker is allowed to randomize over different available actions if necessary. Unfortunately, the strong requirement of statistical sufficiency by Blackwell’s informativeness order renders most information sources indistinguishable. Therefore, to compare information sources relative to a particular decision problem, there is a need for more resolution. [Lindley \(1956\)](#) introduces one such criterion based on entropy measures by [Shannon](#)

(1948). DeGroot (1962) showed that generalized entropy methods can in fact be viewed as equivalent to Blackwell's informativeness criterion. More recently, Athey and Levin (2001) restrict attention to a class of "monotone decision problems" (i.e., problems in which higher posterior beliefs induce higher actions), for which they define a "monotone information order" based on a seminal article by Lehmann (1988). For a class of "balanced" decision problems, which have similar gains for the optimal decisions relative to the realization of the state of nature, we introduce a confidence order to compare information sources in terms of their value to the decision maker. We find that for any given decision problem the value of a binary experiment is nondecreasing in at least one of its two likelihood ratios.<sup>4</sup>

In contrast to much of the earlier literature, the focus of our approach is not to have an exhaustive ordering for all decision problems, but to provide a simple method for comparing binary experiments for a large class of decision problems. These techniques can be used by decision makers with only a minimal need for mathematical modeling and may therefore be of direct practical value. The power of reparametrization for obtaining monotone comparative statistics in optimization problems that might not exhibit monotone behavior at the outset is examined by Strulovici and Weber (2008). The decision tools in this article can be interpreted as reparametrizations of balanced decision problems, in part to ensure that the value of information is monotonic in the confidence parameters.

## 1.2 Outline

The article proceeds as follows. Section 2 introduces the parametrization of binary signals using confidence parameters. We then introduce iso-confidence lines in the ROC space. In Sect. 3, we present a general "standard decision problem" and compare an information order based on confidence parameters to the considerably stronger (and thus generally less useful) Blackwell order of signals in terms of their informativeness. The confidence order can be used for ordering experiments that balance the decision problem. In Sect. 4, we discuss our findings and their limitations. In the appendix, we generalize the framework in this article to decision problems and experiments with nonbinary outcome spaces. We also briefly consider the use of the confidence parametrization for the aggregation of different informative signals and for the elimination of confidence-dominated portions of any given ROC curve.

## 2 Evaluation of binary experiments

Let the *state space*  $\mathcal{X} = \{X, \bar{X}\}$  be the set of possible realizations (or outcomes) of a random variable  $\tilde{x}$  of interest to the decision maker. The decision maker's prior belief about the probability that the outcome  $X$  occurs is  $P(X)$ . When  $P(X)$  is either 0 or 1, the decision maker is absolutely sure of his prior belief and no experiment is powerful

---

<sup>4</sup> Because of the well-known "monotone-likelihood-ratio" stochastic-dominance order for comparing probability distributions, we prefer using the term "confidence order" instead of the somewhat more suggestive term likelihood-ratio order.

enough to change his mind. We restrict attention to the interesting case where this outcome is neither certain nor completely impossible so that

$$0 < P(X) < 1. \tag{1}$$

We assume that there is an information source in the form of a random variable  $\tilde{S}$  with realizations in the binary sample space  $S = \{S, \bar{S}\}$ . To make things nontrivial, we consider here information sources for which the message  $S$  does occur sometimes but not always so that

$$0 < P(S) < 1 \tag{2}$$

is naturally satisfied. The decision maker may consider  $S$  as an indicator for  $X$ , characterized by the conditional probability

$$\pi = P(S|X) \in (0, 1) \tag{3}$$

of giving a “true positive” indication, and the probability

$$\varepsilon = P(S|\bar{X}) \in (0, 1) \tag{4}$$

of giving a “false positive” indication. In relations (3) and (4), we assume that the message  $S$  is not a perfect indicator for  $X$ . This covers all interesting applications.<sup>5</sup> In addition, by switching the labels of  $S$  and its complement  $\bar{S}$  if necessary, we can restrict attention to the case in which

$$0 < \varepsilon \leq \pi < 1. \tag{5}$$

Otherwise the decision maker could—instead of  $S$ —simply take  $\bar{S}$  as an indicator for  $X$ . Relation (5) does not limit generality in any way; it solely imposes a “proper interpretation” of messages by the decision maker.

*Remark 1* The probability  $\varepsilon$  of obtaining  $S$  even though  $X$  is not realized is commonly referred to as *type-I error*, while its complement,  $1 - \varepsilon$ , is also called the signal’s *specificity*. Similarly, the probability  $1 - \pi$  of obtaining a message  $\bar{S}$  even though the outcome  $X$  occurred is referred to as *type-II error*. Its complement,  $\pi$ , is also called the signal’s *sensitivity*. □

Having observed  $S$ , the decision maker updates his beliefs using Bayes’ rule,

$$P(X|S) = \frac{P(S|X)P(X)}{P(S|X)P(X) + P(S|\bar{X})(1 - P(X))}. \tag{6}$$

<sup>5</sup> The cases of perfect information (when  $(\varepsilon, \pi) = (0, 1)$ ) and of one-sidedly noiseless information structures (when either  $\varepsilon = 0$  or  $\pi = 1$ ) can be obtained by taking the limits for  $\kappa \rightarrow \infty$  and/or  $\lambda \rightarrow \infty$ , where  $\kappa$  and  $\lambda$  are given in (9) and (13), respectively.

The decision maker’s posterior belief  $P(X|S)$  in (6) can be rewritten equivalently in the form

$$P(X|S) = \frac{\kappa P(X)}{1 + (\kappa - 1)P(X)}, \tag{7}$$

as a function solely of the decision maker’s prior belief  $P(X)$  and the parameter

$$\kappa = \frac{P(S|X)}{P(S|\bar{X})}, \tag{8}$$

which represents the decision maker’s *confidence in S as an indicator for X*. From relations (3)–(5), we conclude that

$$\kappa = \frac{\pi}{\varepsilon} \geq 1. \tag{9}$$

Thus, if we abbreviate the decision maker’s prior belief about the likelihood of  $X$  by  $p = P(X)$ , his posterior belief can be obtained from (7) by evaluating the function

$$g(\kappa, p) = \frac{\kappa p}{1 + (\kappa - 1)p}. \tag{10}$$

Fixing  $\kappa$  and varying  $p$  from 0 to 1, we thus obtain as the graph of  $g(\kappa, \cdot)$  a curve of the updated beliefs about the likelihood of  $X$  conditional on  $S$  for any possible Bayesian decision maker. By changing  $\kappa \geq 1$ , we obtain a family of such curves for other messages from potentially different information sources that satisfy relations (3)–(5). If instead of  $S$  the message  $\bar{S}$  is realized, the decision maker’s posterior belief becomes

$$P(X|\bar{S}) = \frac{P(\bar{S}|X)P(X)}{P(\bar{S}|X)P(X) + P(\bar{S}|\bar{X})(1 - P(X))}. \tag{11}$$

If, as before,  $p = P(X)$  denotes the decision maker’s prior belief, then his posterior belief after a realization of  $\bar{S}$  is given by

$$g(1/\lambda, p) = \frac{\left(\frac{1}{\lambda}\right) p}{1 + \left(\frac{1}{\lambda} - 1\right) p}, \tag{12}$$

where the parameter

$$\lambda = \frac{P(\bar{S}|\bar{X})}{P(\bar{S}|X)} = \frac{1 - \varepsilon}{1 - \pi} \geq 1 \tag{13}$$

represents the *confidence in  $\bar{S}$  as an indicator for  $\bar{X}$* , analogous to the earlier definition of  $\kappa$  in (8).

**Lemma 1** *Let  $\kappa, \lambda \geq 1$ . (i) If  $(\kappa, \lambda) \neq (1, 1)$ , then  $(\kappa - 1)(\lambda - 1) > 0$ . (ii) If  $\kappa > 1$ , then  $P(X|S) > P(X)$ . (iii) If  $\lambda > 1$ , then  $P(X|\bar{S}) < P(X)$ .*

*Proof* (i) Consider first the case where  $\kappa > 1$ . By relation (9), we have  $\pi > \varepsilon$  so that  $1 - \varepsilon > 1 - \pi$  and by (13), therefore,  $\lambda > 1$ . The converse also holds. Similarly,

if  $\kappa = 1$ , then by (9) and (13) it must be  $\lambda = 1$ . The converse also holds. (ii) Let  $p \in (0, 1)$  be the decision maker’s prior belief about the likelihood of  $X$ . Conditional on having observed the message  $S$ , the decision maker’s updated posterior belief is  $P(X|S) = g(\kappa, p) = \kappa p / (1 + (\kappa - 1)p)$  by virtue of (10). Since

$$\frac{\partial g(\kappa, p)}{\partial \kappa} = \frac{(1 - p)p}{(1 + (\kappa - 1)p)^2} > 0$$

for all  $p \in (0, 1)$  and all  $\kappa > 1$ , the decision maker’s posterior belief  $g(\kappa, p)$  is strictly increasing in  $\kappa$ . Thus,  $g(\kappa, p) > g(1, p) = 1$  for all  $p \in (0, 1)$ . (iii) The proof for this part proceeds analogously to the one for part (ii), by first showing that  $g(1/\lambda, p)$  is strictly decreasing in  $\lambda > 1$ . □

Part (i) of Lemma 1 states that restricting attention to  $\kappa \geq 1$  is without any loss in generality, as  $\kappa \geq 1$  if and only if  $\lambda \geq 1$ . Therefore, one can always label the messages of a binary information source such that  $\kappa, \lambda \geq 1$ . The message  $S$  is then an indicator for  $X$ , whereas  $\bar{S}$  is an indicator for  $\bar{X}$ . Parts (ii) and (iii) of this lemma state that when the confidence parameters  $\kappa$  and  $\lambda$  exceed 1, the decision maker’s prior belief about the likelihood of  $X$  increases upon receiving the message  $S$ , while it decreases upon receiving the message  $\bar{S}$ . Thus, increasing one of the two confidence parameters strictly increases the “spread,”  $P(X|S) - P(X|\bar{S})$ , in the decision maker’s posterior beliefs.

The following result shows that the signal description by the probability tuple  $(\varepsilon, \pi)$  is equivalent to a description by the confidence parameter tuple  $(\kappa, \lambda)$ ; the two parametrizations are linked by a smooth one-to-one mapping.

**Lemma 2** (i) If  $0 < \varepsilon < \pi < 1$ , then  $(\kappa, \lambda) = (\pi/\varepsilon, (1 - \varepsilon)/(1 - \pi)) > (1, 1)$ .  
 (ii) If  $\kappa, \lambda > 1$ , then

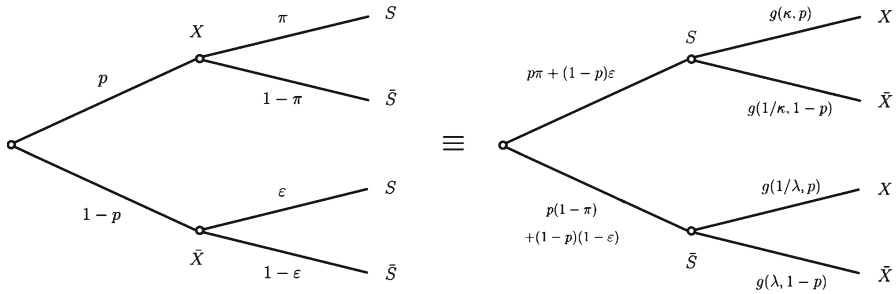
$$(\varepsilon, \pi) = \left( \frac{\lambda - 1}{\kappa\lambda - 1}, 1 - \frac{\kappa - 1}{\kappa\lambda - 1} \right) \tag{14}$$

and  $0 < \varepsilon < \pi < 1$ .

The parameter tuples  $(\varepsilon, \pi)$  and  $(\kappa, \lambda)$  are directly related by a one-to-one mapping on our domain of interest. The proof of Lemma 2, which contains a simple inversion of the relations (8) and (13), is elementary and has therefore been omitted. The key consequence for a decision maker is that he can move seamlessly back and forth between a description of an information source in terms of  $(\varepsilon, \pi)$  or in terms of  $(\kappa, \lambda)$ . We will do so in the remainder of the article and refer to  $(\varepsilon, \pi)$  as an *error parametrization* and  $(\kappa, \lambda)$  as a *confidence parametrization*. Figure 1 shows the binary event tree in the equivalent error and confidence parametrizations.

*Example 1* Let the state of the world  $X$  represent the presence of a certain disease in a patient. After a preliminary examination, a doctor’s prior belief about the presence of the disease is  $p = P(X) = 50\%$ . The doctor can now perform two different tests, each of which either produces a “symptom” ( $S$ ) or not ( $\bar{S}$ ). The first test has a specificity of 55% and a sensitivity of 70%, corresponding to the error characteristics  $(\varepsilon, \pi) = (45\%, 70\%)$ . From relation (14) in Lemma 2, we obtain the equivalent





**Fig. 1** Binary event tree in error parameters and confidence parameters

confidence parameters,  $(\kappa, \lambda) \approx (1.6, 1.8)$ . From parts (ii) and (iii) of Lemma 1, we see that the test outcomes, especially when the symptom is not observed, influence the posterior beliefs of a Bayesian decision maker. From relations (10) and (12), the decision maker’s posterior is  $P(X|S) \approx 62\%$  in the case of  $S$  and  $P(X|\bar{S}) \approx 36\%$  in the case of  $\bar{S}$ . The second test has a specificity of 80% and a sensitivity of 60%, corresponding to the error characteristics  $(\hat{\epsilon}, \hat{\pi}) = (20\%, 60\%)$  and a confidence  $(\hat{\kappa}, \hat{\lambda}) = (3, 2) > (\kappa, \lambda)$ . Based on this test the doctor can augment his posterior belief adjustments to 75% in case of  $S$  and 33% in case of  $\bar{S}$ . Thus, while the confidence characteristics of the two tests can be clearly ordered, this is not possible for their error characteristics; since the second test achieves a significant improvement in the type-II error, from 45% to 20%. The performance in terms of its type-I error, with an increase from 30% to 40%. We show below that while this makes the two tests incomparable in the sense of Blackwell, i.e., relative to the set of *all* decision problems, it is still possible to order the tests relative to a class of reasonably general classification problems. This is consistent with our intuition from comparing the two tests’ confidence characteristics.  $\square$

*Remark 2* It is instructive to represent the role of confidence parameters in a diagram, which we refer to as Bayesian update diagram. For this we first recall that a binary information source is fully characterized by the two independent parameters  $\pi$  and  $\epsilon$  in (3) and (4), albeit with the inconsequential restriction (5). Both of these parameters generically influence the decision maker’s posterior beliefs in (6) and (11) so that an easy graphical representation of the posterior as a function of  $(p, \epsilon, \pi)$  is not possible. The confidence parameters  $\kappa$  and  $\lambda$  in (8) and (13) have a separate influence, depending on which message is received from the information source. Conditional on having observed  $S$ , the decision maker’s posterior belief  $g(\kappa, p)$  in (10) depends only on  $p$  and  $\kappa$ , while conditional on having observed  $\bar{S}$ , the posterior  $g(1/\lambda, p)$  depends only on  $p$  and  $\lambda$ . Thus, using the confidence parameters  $\kappa$  and  $\lambda$ , it is possible to construct the Bayesian update diagram depicted in Fig. 2. The solid lines represent “iso-confidence lines” in the  $(p, q)$ -space, where  $q$  denotes the decision maker’s posterior belief of the likelihood of  $X$  after having observed a signal realization. The curves above the 45° line are iso- $\kappa$  curves (for a realization of  $S$ ), while the others are iso- $\lambda$  curves (for a realization of  $\bar{S}$ ). The Bayesian update diagram can be used to obtain Bayes updates graphically, without using Bayes’ rule explicitly (cf. Appendix A).  $\square$

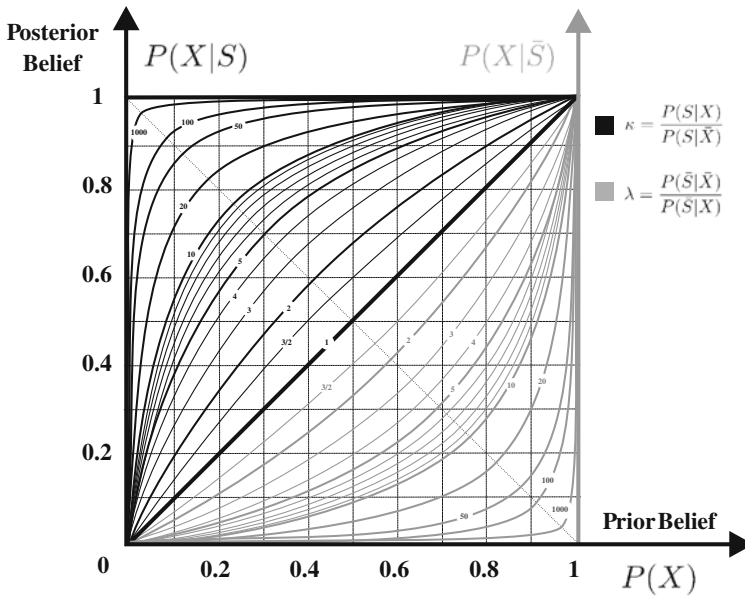


Fig. 2 Bayesian update diagram

### 3 Comparison of binary experiments

Before choosing an optimal action in a given decision problem, a decision maker may acquire information in the form of access to signal realizations that are correlated with the realization of a random outcome. When the use of one information source excludes the use of another (e.g., when the decision maker chooses between different *designs* of an experiment), to be able to make a choice between different sources or designs, it is necessary to have a preference order over all available experiments.<sup>6</sup> Blackwell (1951) introduced an ordering of signals in terms of their “informativeness,” which ranks different signals for *any* stochastic decision problem with a convex action set and a continuous payoff function. Blackwell’s order is equivalent to the concept of statistical sufficiency.<sup>7</sup> Because of the imposed strength of the Blackwell order, many, if not most, information sources cannot be compared under it. For example, while reading the local news of the *New York Times* might be useful for deciding where to dine in New York City, people living in San Francisco are likely to instead prefer the *San Francisco Chronicle* when pondering about where to eat out that night. Thus, in most practical settings Blackwell’s informativeness criterion is of no use to order information sources in terms of their value in a given choice situation. In this section, we first introduce a “standard decision problem,” relative to which we examine the ordering

<sup>6</sup> In Example 1, a doctor had the choice between two different diagnostic tests, and in Appendix B, we discuss how information from different tests may be *combined* to form posterior beliefs.

<sup>7</sup> A signal  $\bar{s}$  is (statistically) *sufficient* for another signal  $\bar{s}'$  if the conditional probability distribution of any data  $\bar{x}$ , given any realization of  $\bar{s}$ , is independent of the realization of  $\bar{s}'$ , i.e., when  $P(\bar{x}|\bar{s}, \bar{s}') = P(\bar{x}|\bar{s})$ .

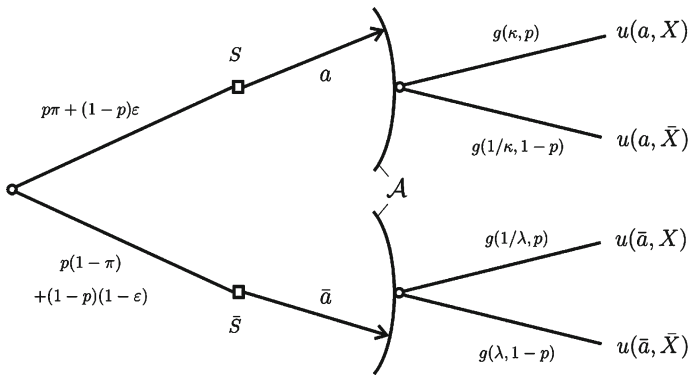


Fig. 3 The standard decision problem

of binary information sources. Historically, the ROC has been used for this purpose. The confidence order enables a (strict) comparison of many more experiments in the ROC space than the Blackwell order. It also allows one to eliminate certain portions from ROC curves, which are confidence-inconsistent (for details see Appendix C).

### 3.1 The standard decision problem

In order to compare different sources of information that the decision maker might gain access to, it is necessary to formulate a decision problem including preferences over action-outcome tuples (Fig. 3). We assume that *before* the random variable  $\tilde{x}$  realizes and *after* a message from the information source  $\tilde{s}$  is observed, the decision maker chooses an action  $a$  from a given nonempty, finite-dimensional, and compact *action set*  $\mathcal{A}$ . It is customary to assume that the action set  $\mathcal{A}$  is convex. Convexity of the action set is without any loss of generality (and is relaxed completely for the confidence order introduced in Sect. 3.3), as long as the decision maker can randomize over all actions in  $\mathcal{A}$ . Nonemptiness and compactness guarantee the existence of an optimal action for any reasonable (continuous) payoff function that the decision maker may have. The decision maker’s preferences over tuples of actions and outcomes  $(a, x) \in \mathcal{A} \times \mathcal{X}$  are assumed to be representable by a utility function  $u : \mathcal{A} \times \mathcal{X} \rightarrow \mathbb{R}$ .<sup>8</sup> As we show below, trying to compare different informative signals relative to the class of *all* such decision makers, i.e., without specifying prior beliefs and without making any further assumptions on the utility function  $u$ , necessarily leads to a very weak preference order which renders it impossible to distinguish most signals. It is therefore natural to expect that sharper criteria for the desirability or “informativeness” of signals must incorporate particulars of the decision problem: utility function, action set, and beliefs. Given the informative signal  $\tilde{s}$  with possible realizations  $S$  and  $\bar{S}$ , the *standard decision problem* is to find

<sup>8</sup> For details on which preferences allow a utility representation, see Fishburn (1970) or Kreps (1988).

$$a^* \in \arg \max_{a \in \mathcal{A}} E [u(a, \tilde{x}) | S] = \arg \max_{a \in \mathcal{A}} \{p\pi u(a, X) + (1 - p)\varepsilon u(a, \bar{X})\}, \tag{15}$$

and

$$\begin{aligned} \bar{a}^* \in \arg \max_{a \in \mathcal{A}} E [u(a, \tilde{x}) | \bar{S}] = \arg \max_{a \in \mathcal{A}} \{ & p(1 - \pi)u(a, X) \\ & + (1 - p)(1 - \varepsilon)u(a, \bar{X})\}, \end{aligned} \tag{16}$$

respectively. The decision maker’s optimal expected utility (also referred to as *decision value*) is

$$V^* = p (\pi u(a^*, X) + (1 - \pi)u(\bar{a}^*, X)) + (1 - p) (\varepsilon u(a^*, \bar{X}) + (1 - \varepsilon)u(\bar{a}^*, \bar{X})). \tag{17}$$

Without any information, the decision maker’s *default utility* is

$$V_0^* = \max_{a \in \mathcal{A}} E [u(a, \tilde{x})] = \max_{a \in \mathcal{A}} \{pu(a, X) + (1 - p)u(a, \bar{X})\}. \tag{18}$$

Any continuous function attains its maximum on a compact set so that solutions to the problems in (15), (16), and (18) exist.

*Remark 3* In the standard decision problem, we have implicitly assumed that the probability of an outcome remains unaffected by the decision maker’s action. It is easy to imagine decision problems where this assumption is unrealistic. For example, when deciding to wear a seat-belt while driving a car, the probability of getting seriously injured conditional on an accident (event  $X$ ) clearly depends on the chosen action. This dependence can easily be incorporated into our framework by choosing the payoff function  $u(a, x)$  appropriately. If  $\hat{P}(x|a)$  denotes the conditional probability of  $x \in \mathcal{X}$  given an action  $a$ , then (as long as  $P(x) > 0$ ) it is enough to replace  $u(a, x)$  by  $\hat{u}(a, x) = u(a, x) \frac{\hat{P}(x|a)}{P(x)}$ ; the formulation of the standard decision problem (15) and (16) is therefore without loss of generality.  $\square$

### 3.2 Blackwell’s informativeness order

Given an action  $a$  and an outcome realization  $x$ , the decision maker obtains the utility payoff  $u(a, x)$ . Consider now two informative signals  $\tilde{s}_1$  and  $\tilde{s}_2$  with realizations in the common sample space  $\mathcal{S}$ . The signal  $\tilde{s}_1$  is *at least as informative as the signal  $\tilde{s}_2$  in the sense of Blackwell* if all utilities attainable given  $\tilde{s}_2$  are also attainable given  $\tilde{s}_1$ .

*Remark 4* The precise definition proceeds as follows. Assume first that the action space  $\mathcal{A}$ , in addition to being nonempty and compact, is also convex. Let  $\alpha : \mathcal{S} \rightarrow \mathcal{A}$  be a policy that describes how the decision maker chooses an action  $\alpha(s) \in \mathcal{A}$ , contingent on a signal realization  $s \in \mathcal{S}$ . The *state-payoff vector* relative to the policy  $\alpha$  and the signal  $\tilde{s}$  is given by

$$v(\tilde{s}, \alpha) = (\pi u(\alpha(S), X) + (1 - \pi)u(\alpha(\bar{S}), X), \varepsilon u(\alpha(S), \bar{X}) + (1 - \varepsilon)u(\alpha(\bar{S}), \bar{X})).$$

The set of all attainable state-payoff vectors under the signal  $\tilde{s}$  with respect to all possible (measurable) policies is

$$\mathcal{V}(\tilde{s}) = \left\{ v \in \mathbb{R}^2 : v = v(\tilde{s}, \alpha) \text{ and } \alpha : \mathcal{S} \rightarrow \mathcal{A} \right\}.$$

The signal  $\tilde{s}_1$  is at least as informative as the signal  $\tilde{s}_2$  in the sense of Blackwell, if  $\mathcal{V}(\tilde{s}_2) \subseteq \mathcal{V}(\tilde{s}_1)$ . In the case where the action set  $\mathcal{A}$  is nonconvex, we allow the decision maker to randomize among elements of  $\mathcal{A}$ , which effectively amounts to convexifying the action set (and, more importantly, it also convexifies the set of all attainable state-payoff vectors). A randomized policy  $\tilde{\alpha}$  must then map  $\mathcal{S}$  into the set of all probability distributions over elements of  $\mathcal{A}$ , and a corresponding state-payoff vector can then be defined as the expected value of the state-payoff under the randomized policy. The definition of the informativeness is from then on analogous to the case where policies are deterministic.  $\square$

Blackwell (1951, 1953) showed that  $\tilde{s}_1$  is at least as informative as  $\tilde{s}_2$  for any decision maker facing a standard decision problem as introduced in Sect. 3.1 if and only if  $\tilde{s}_1$  is statistically sufficient for  $\tilde{s}_2$ .<sup>9</sup> We can therefore focus entirely on statistical sufficiency as determinant for the Blackwell ordering of informative signals. In our simple context, where we restrict attention to signals with binary messages, either  $S_i$  or  $\bar{S}_i$  for signal  $\tilde{s}_i$ , the signal  $\tilde{s}_1$  is statistically sufficient for  $\tilde{s}_2$  if and only if there exists a  $2 \times 2$  Markov matrix  $M$  such that<sup>10</sup>

$$\begin{bmatrix} \pi_2 & \varepsilon_2 \\ 1 - \pi_2 & 1 - \varepsilon_2 \end{bmatrix} = M \begin{bmatrix} \pi_1 & \varepsilon_1 \\ 1 - \pi_1 & 1 - \varepsilon_1 \end{bmatrix}, \quad (19)$$

where  $(\varepsilon_i, \pi_i) = (P(\tilde{s}_i \in S_i | \bar{X}), P(\tilde{s}_i \in S_i | X))$  for  $i \in \{1, 2\}$ . Using the concept of statistical sufficiency, it is possible to precisely characterize the Blackwell ordering in terms of the error characteristics of binary experiments.

**Proposition 1** *Let  $0 < \varepsilon < \pi < 1$  and  $0 < \hat{\varepsilon} < \hat{\pi} < 1$ . A signal with  $(P(\hat{S} | \bar{X}), P(\hat{S} | X)) = (\hat{\varepsilon}, \hat{\pi})$  is at least as informative as a signal with  $(P(S | \bar{X}), P(S | X)) = (\varepsilon, \pi)$  if and only if  $(-\hat{\varepsilon}, \hat{\pi}) \geq (-\varepsilon, \pi)$ .*

*Proof* By (19) the signal with characteristics  $(\hat{\varepsilon}, \hat{\pi})$  is at least as informative as the signal with characteristics  $(\varepsilon, \pi)$  if and only if a  $2 \times 2$  Markov matrix  $M$  can be found such that

$$\begin{bmatrix} \pi & \varepsilon \\ 1 - \pi & 1 - \varepsilon \end{bmatrix} = M \begin{bmatrix} \hat{\pi} & \hat{\varepsilon} \\ 1 - \hat{\pi} & 1 - \hat{\varepsilon} \end{bmatrix}.$$

<sup>9</sup> Kihlstrom (1984) provides an exposition of this result in a Bayesian setting similar to the one chosen here.

<sup>10</sup> A square matrix  $M$  is a Markov matrix (or “has the Markov property”) if all its entries are in  $[0, 1]$  and each of its columns sums up to 1.

It is easy to show that the latter can be accomplished if and only if the inequalities  $\hat{\varepsilon} \leq \varepsilon$  and  $\hat{\pi} \geq \pi$  are simultaneously satisfied. Only if the last two inequalities hold, the entries of

$$M = \begin{bmatrix} \hat{\pi} & \hat{\varepsilon} \\ 1 - \hat{\pi} & 1 - \hat{\varepsilon} \end{bmatrix}^{-1} \begin{bmatrix} \pi & \varepsilon \\ 1 - \pi & 1 - \varepsilon \end{bmatrix} = \frac{1}{\hat{\pi} - \hat{\varepsilon}} \begin{bmatrix} \pi - \hat{\varepsilon} & \varepsilon - \hat{\varepsilon} \\ \hat{\pi} - \pi & \hat{\pi} - \varepsilon \end{bmatrix}$$

are in  $[0, 1]$  and each of its columns sums up to 1 so that  $M$  has the Markov property. □

In Example 1, we already noted that the Blackwell criterion in Proposition 1 is generally too restrictive to be able to strictly order two signals (in the sense that one signal is at least as informative as the other but not vice versa). To compare the Blackwell order to our weaker confidence order that we introduce in the next subsection, it is useful to provide an alternative characterization based on what we term the ‘‘Blackwell-derivative.’’ Let  $V^*(\varepsilon, \pi)$  be the optimal expected utility given a signal of characteristics  $(\varepsilon, \pi)$ . From Proposition 1, we can conclude that the Blackwell order is equivalent to  $V^*$  being nondecreasing in  $(-\varepsilon, \pi)$ . Let us call

$$\Delta = p(u(a^*, X) - u(\bar{a}^*, X)) \tag{20}$$

the decision maker’s (expected)  $X$ -gain from adjusting his action from  $\bar{a}^*$  to  $a^*$  in the event  $X$ . Similarly,

$$\bar{\Delta} = (1 - p)(u(\bar{a}^*, \bar{X}) - u(a^*, \bar{X})) \tag{21}$$

is termed the decision maker’s  $\bar{X}$ -gain from adjusting his action from  $a^*$  to  $\bar{a}^*$  in the event  $\bar{X}$ . Since by (5)  $S$  is an indicator for  $X$  and  $\bar{S}$  is an indicator for  $\bar{X}$ , both of these gains must be nonnegative, i.e.,

$$\bar{\Delta}, \Delta \geq 0. \tag{22}$$

The following result effectively restates Proposition 1 and introduces the Blackwell-derivative.

**Proposition 2** *Let  $0 < \varepsilon < \pi < 1$ . Then the (directional) Blackwell-derivative is*

$$D_{(-\varepsilon, \pi)} V^* \cdot v = v_1 \bar{\Delta} + v_2 \Delta \geq 0, \tag{23}$$

along any direction  $v = (v_1, v_2) \in \mathbb{R}_+^2 \setminus \{0\}$ .

*Proof* Let  $v = (v_1, v_2) \in \mathbb{R}_+^2 \setminus \{0\}$ . Using relations (15)–(17) and (20)–(21), as well as the envelope theorem, we have that

$$\begin{aligned} D_{(\varepsilon, \pi)} V^* &= \frac{\partial}{\partial(\varepsilon, \pi)} \left[ p(\pi u(a^*, X) + (1 - \pi)u(\bar{a}^*, X)) + (1 - p)(\varepsilon u(a^*, \bar{X}) \right. \\ &\quad \left. + (1 - \varepsilon)u(\bar{a}^*, \bar{X})) \right] \\ &= \frac{\partial}{\partial(\varepsilon, \pi)} \left[ \pi p(u(a^*, X) - u(\bar{a}^*, X)) + \varepsilon(1 - p)(u(a^*, \bar{X}) - u(\bar{a}^*, \bar{X})) \right] \\ &= (-\bar{\Delta}, \Delta), \end{aligned}$$

which, by virtue of (22), implies that relation (23) holds.<sup>11</sup>  $\square$

Proposition 2 states that the directional (“Blackwell”)–derivative, corresponding to the slope of the optimal expected value  $V^*$  along any direction  $v$  between  $(-1, 0)$  and  $(0, 1)$ , must be nonnegative. In other words, the decision value cannot decrease if  $(-\varepsilon, \pi)$  increases, which corresponds precisely to the informativeness criterion in Proposition 1. We term (23) the *Blackwell criterion*, which needs to be satisfied for all standard decision problems. It therefore provides an important baseline comparison of information sources.

*Remark 5* A good analogy for ordering information sources relative to different classes of decision makers is stochastic dominance. While the first-order stochastic dominance order provides an excellent baseline comparison of probability distributions, it considers a reference class of decision makers (all those with utility functions increasing in the scalar outcome) that is too large to be able to order many probability distributions. The notions of second-, third-, and higher-order stochastic dominance impose successively weaker criteria on the comparison of probability distributions, effectively narrowing down the class of decision makers for which the comparison is valid. For example, second-order stochastic dominance implies a preference over distributions for all decision makers with utility functions that are increasing and concave in the scalar outcome.  $\square$

### 3.3 Confidence order

Comparing different informative signals for all possible decision makers with different utility functions  $u$  on  $\mathcal{A} \times \mathcal{X}$  and different prior beliefs necessarily leads to a very weak preference order, which makes it impossible to distinguish many, if not most, signals. It is therefore natural to expect that sharper criteria for the desirability or “informativeness” of signals must incorporate particulars of the decision problem, such as the decision maker’s utility payoffs and prior beliefs. Earlier work focusses on particular classes of decision problems, such as portfolio decisions (Athey 2000), or monotone decision problems (Athey and Levin 2001). Our ordering criterion based on the confidence parameters  $\kappa$  and  $\lambda$  is applicable for comparing binary experiments for our standard decision problem introduced in Sect. 3.1. We note that our initial assumption of the convexity of the action set  $\mathcal{A}$  is not necessary for the confidence order and can be dropped.

A *confidence order* for the comparison of binary signals can now be formulated as follows: a decision maker facing the standard decision problem in Sect. 3.1 is *at least as confident* in signal  $\tilde{s}_1$ , of confidence  $(\kappa_1, \lambda_1)$ , as in signal  $\tilde{s}_2$ , of confidence  $(\kappa_2, \lambda_2)$ , if

$$(\kappa_1, \lambda_1) \geq (\kappa_2, \lambda_2). \quad (24)$$

The “augmented” ROC space in Fig. 4, which contains the iso-confidence lines for  $\kappa$  and  $\lambda$ , shows that the confidence order is strictly weaker than the Blackwell order.

<sup>11</sup> The use of the envelope theorem for arbitrary choice sets (including our action set  $\mathcal{A}$ ) has been discussed by Milgrom and Segal (2002).

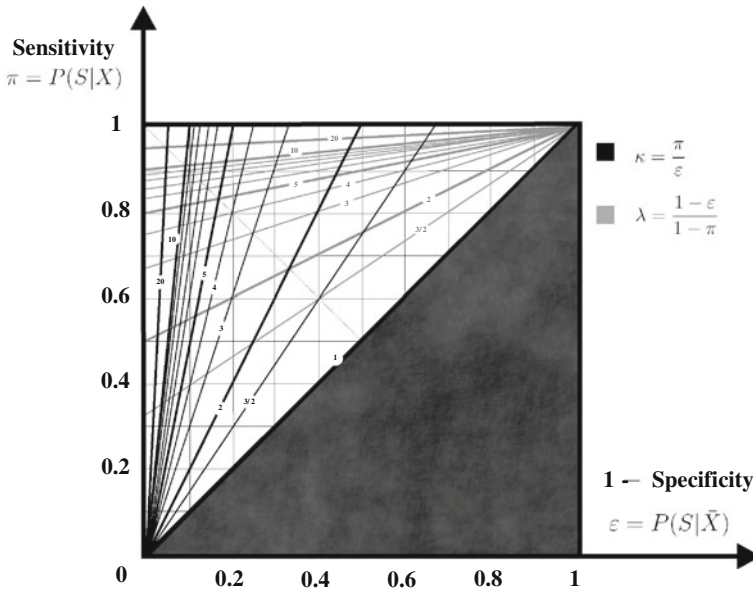


Fig. 4 ROC space with iso-confidence lines

The latter requires that

$$(-\varepsilon_1, \pi_1) \geq (-\varepsilon_2, \pi_2), \tag{25}$$

and, indeed, relation (25) implies (24) but not vice versa. The following result characterizes the confidence order based on  $(\kappa, \lambda)$ .

**Proposition 3** *Let  $\kappa, \lambda > 1$ . Then the components  $\partial V^*/\partial \kappa$  and  $\partial V^*/\partial \lambda$  of the confidence-derivative  $D_{(\kappa, \lambda)} V^*$  satisfy*

$$(\bar{\Delta} - \kappa \Delta) \frac{\partial V^*}{\partial \lambda} \leq 0 \leq \left( \bar{\Delta} - \left( \frac{1}{\lambda} \right) \Delta \right) \frac{\partial V^*}{\partial \kappa}. \tag{26}$$

*Proof* Assume that  $\kappa, \lambda > 1$ . Using the envelope theorem as in the proof of Proposition 2 and the chain rule for differentiation, the confidence-derivative can be computed as follows:

$$D_{(\kappa, \lambda)} V^* = \frac{\partial V^*}{\partial(\varepsilon, \pi)} \frac{\partial(\varepsilon, \pi)}{\partial(\kappa, \lambda)} = (-\bar{\Delta}, \Delta) \begin{bmatrix} -(\lambda - 1)\lambda & \kappa - 1 \\ -(\lambda - 1) & (\kappa - 1)\kappa \end{bmatrix} \frac{1}{(\kappa\lambda - 1)^2}. \tag{27}$$

Hence,  $\partial V^*/\partial \kappa \geq 0$  if and only if  $\bar{\Delta} - (1/\lambda)\Delta \geq 0$ , and  $\partial V^*/\partial \lambda \geq 0$  if and only if  $\bar{\Delta} - \kappa\Delta \leq 0$ . The last two relations are equivalent to the confidence criterion (26), which concludes our proof.  $\square$

We term the relation (26) characterizing the confidence order the *confidence criterion*. The first inequality in (26) states that the decision value  $V^*$  is nondecreasing in the confidence  $\lambda$  if and only if the  $\bar{X}$ -gain  $\bar{\Delta}$  is at least as large as  $\kappa$  times the  $X$ -gain. Thus,



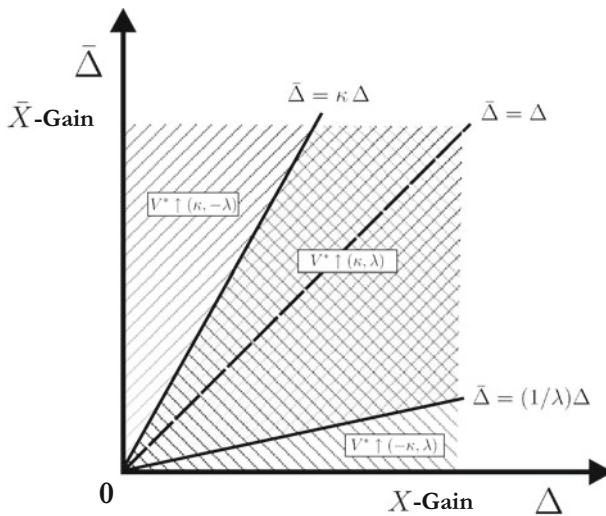


Fig. 5 Monotonicity of the information value in  $\kappa$  and  $\lambda$

for the decision value  $V^*$  to increase in the confidence  $\lambda$  of  $\bar{S}$  as an indicator for  $\bar{X}$ , the gain  $\bar{\Delta}$  from adjusting from  $a^*$  to  $\bar{a}^*$  for a realization of  $\bar{X}$  has to exceed  $\kappa \Delta$ . Similarly, the second inequality in (26) states that the decision value  $V^*$  is nondecreasing in the confidence  $\kappa$  if and only if the  $X$ -gain  $\Delta$  is at least as large as  $\lambda$  times the  $\bar{X}$ -gain. In other words, for the decision value  $V^*$  to increase in the confidence  $\kappa$  of  $S$  as an indicator for  $X$ , the gain  $\Delta$  from adjusting from  $\bar{a}^*$  to  $a^*$  for a realization of  $X$  has to exceed  $\lambda \bar{\Delta}$ . When the gains  $\Delta$  and  $\bar{\Delta}$  are strictly positive, we obtain a somewhat more intuitive statement of the confidence criterion in the form

$$D_{(\kappa, \lambda)} V^* \geq 0 \Leftrightarrow \left( \frac{\bar{\Delta}}{\Delta} \leq \kappa \text{ and } \frac{\Delta}{\bar{\Delta}} \leq \lambda \right). \tag{28}$$

The decision problem is called *balanced* relative to  $(\kappa, \lambda)$  if the right-hand side of (28) is satisfied. It is clear that *any decision problem with nonzero gains is balanced if only the experiment is good enough*, i.e., if  $\kappa, \lambda$  are sufficiently large. Thus, by relation (28) it is

$$\kappa, \lambda \gg 1 \Rightarrow V^* \text{ is nondecreasing in } (\kappa, \lambda). \tag{29}$$

Figure 5 illustrates the confidence criterion (26) for the monotonicity of  $V^*$  in  $\kappa$  and  $\lambda$ . It particular, it shows that  $V^*$  is always increasing in at least one of the confidence parameters. It also illustrates that *when the gains  $\Delta$  and  $\bar{\Delta}$  are similar, then the decision value  $V^*$  is increasing in  $(\kappa, \lambda)$* , i.e.,

$$\Delta \approx \bar{\Delta} \Rightarrow V^* \text{ is nondecreasing in } (\kappa, \lambda). \tag{30}$$

### 3.4 Comparison of information orders

The confidence order in Sect. 3.3 is weaker than the Blackwell order in Sect. 3.2. While the Blackwell criterion (23) in Proposition 2 guarantees the monotonicity of the optimal decision value  $V^*$  in the error parameters  $(-\varepsilon, \pi)$  for all standard decision problems, the confidence criterion (26) allows for more flexible monotonicity properties of  $V^*$  in the confidence parameters  $(\kappa, \lambda)$ . Nonetheless, as the examples in this section show, the confidence order can provide a much sharper criterion for the value of information and is therefore of use in many practical situations.

*Example 2* A decision maker can take one of two possible actions,  $a = 0$  or  $a = 1$ . The cost of taking action  $a$  is  $ca$ , where  $c \in [0, 1]$  is a given constant. The gross utility for choosing  $a = 1$  corresponds to the realization of a random outcome  $\tilde{x}$  in  $\mathcal{X} = \{0, 1\}$ . The decision maker’s utility for action  $a$  at outcome  $x$  is therefore

$$u(a, x) = a(x - c).$$

The decision maker is interested in the event  $X = \{1\} \subset \mathcal{X}$ . Given a prior belief of  $p = P(X)$ , his expected net payoff is

$$E[u(a, \tilde{x})|p, c] = a(p - c).$$

One can think of  $c$  as a cost threshold, above which taking the action  $a = 1$  becomes beneficial. Such a decision problem is archetypal. For example, many capital investment decisions can be based on thresholds (Keisler 2004). The decision maker’s optimal default policy  $a_0^*$  is to choose 1 whenever  $p > c$ , to choose 0 whenever  $p < c$ , and to choose either of the two actions if  $p = c$ . This leads to the expected default payoff of  $V_0^* = [p - c]_+$ . Using an informative signal of confidence  $(\kappa, \lambda) \geq 1$ , the optimal expected payoff becomes

$$V^* = P(S)[g(\kappa, p) - c]_+ + (1 - P(S))[g(1/\lambda, p) - c]_+,$$

where

$$P(S) = \frac{(1 + (\kappa - 1)p)(\lambda - 1)}{\kappa\lambda - 1}.$$

An optimal policy conditional on  $S$  is to choose the action  $a^* = 1$  if and only if the decision maker’s posterior  $g(\kappa, 1)$  exceeds  $c$ . If, on the contrary, the decision maker observes  $\bar{S}$ , it is optimal to choose  $\bar{a}^* = 1$  if  $g(1/\lambda, p) > c$ . Consider now the value of the informative signal to the decision maker. The  $X$ -gain and  $\bar{X}$ -gain in this example can be computed using (20) and (21) to  $\Delta = p(1 - c)$  and  $\bar{\Delta} = (1 - p)c$ , respectively. By the confidence criterion (26) in Proposition 3, we therefore have that  $D_{(\kappa, \lambda)}^* \geq 0$  if and only if

$$(1 - p)c - \kappa p(1 - c) \leq 0 \leq (1 - p)c - (1/\lambda) p(1 - c). \tag{31}$$

It is possible to obtain the same inequalities using the following, slightly different reasoning. If neither  $a^*$  nor  $\bar{a}^*$  is different from  $a_0^*$ , then the signal cannot have any value for the decision maker. Hence, the informative signal can be worth a positive amount if and only if it has the power to actually change decisions, in which case

$$(g(\kappa, p) - c)(g(1/\lambda, p) - c) \leq 0$$

must necessarily hold. The last condition is equivalent to

$$c \in [P(X|\bar{S}), P(X|S)] = [g(1/\lambda, p), g(\kappa, p)].$$

In other words, the decision threshold  $c$  and the decision maker's prior belief  $p$  about a favorable outcome have to be such that

$$1/\lambda \leq \frac{c}{1-c} \frac{1-p}{p} \leq \kappa. \quad (32)$$

Relation (32), which is equivalent to relation (31), provides tight lower bounds in  $\kappa$  and  $\lambda$  for the signal to have a positive value. For example, if  $p = 0.5$  and  $c = 0.2$ , then the confidence criterion (31) requires that  $\lambda > 0$  for the information value to be strictly increasing in  $\kappa$ ,  $\lambda > 1$  (note that  $\kappa \geq 1/4$  is automatically satisfied). By the equivalent relation (32), the condition  $\lambda > 4$  (together with  $\kappa > 1$ ) is also necessary and sufficient for the value of information to be positive. This is illustrated in Fig. 6.<sup>12</sup>  $\square$

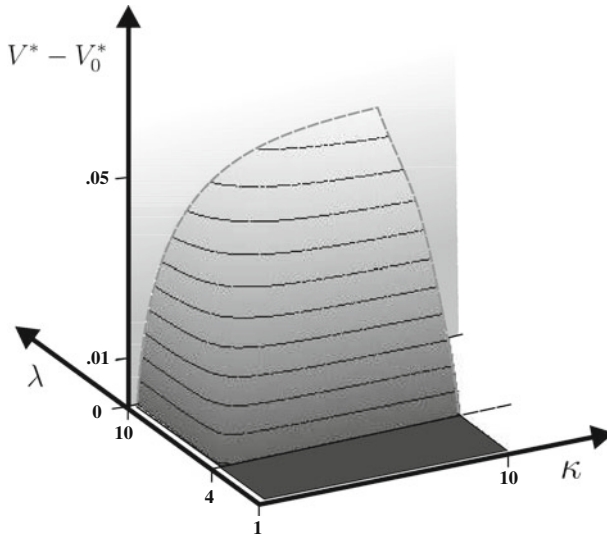
*Remark 6* The natural question arises of how one can compare two signals,  $\tilde{s}_1$  and  $\tilde{s}_2$ , when their costs (denoted by  $C_1$  and  $C_2$ , respectively) are different. While the maximum tolerable difference in the signal-acquisition cost can be determined only as the difference between the respective (gross) decision values they induce,<sup>13</sup> we note that, as long as the decision maker's utility is quasilinear in money, the balancedness condition remains unaffected by the costliness of signals. Further, a priori statements about which of the two signals is preferable are impossible unless one has bounds on the rate of change of the decision value in the confidence parameters of a balanced problem. For example, if it is known that  $\frac{\partial V^*}{\partial \kappa} \in [\underline{\gamma}_\kappa, \bar{\gamma}_\kappa]$  and  $\frac{\partial V^*}{\partial \lambda} \in [\underline{\gamma}_\lambda, \bar{\gamma}_\lambda]$  with known constants  $\underline{\gamma}_\kappa, \underline{\gamma}_\lambda, \bar{\gamma}_\kappa, \bar{\gamma}_\lambda$ , then

$$\underline{\gamma}_\kappa (\kappa_2 - \kappa_1) + \underline{\gamma}_\lambda (\lambda_2 - \lambda_1) \leq V_2^* - V_1^* \leq \bar{\gamma}_\kappa (\kappa_2 - \kappa_1) + \bar{\gamma}_\lambda (\lambda_2 - \lambda_1)$$

for  $(\kappa_2, \lambda_2) \geq (\kappa_1, \lambda_1)$ . This means that as long as the cost difference  $\Delta C = C_2 - C_1$  between signal 2 and signal 1 is smaller than  $\underline{\gamma}_\kappa (\kappa_2 - \kappa_1) + \underline{\gamma}_\lambda (\lambda_2 - \lambda_1)$  signal 2

<sup>12</sup> The value of information in Fig. 6 exhibits the characteristic Radner and Stiglitz (1984) nonconcavity. This nonconcavity appears *irrespective* of how the performance of the information source is parameterized.

<sup>13</sup> The cash value of information for single-person decision problems has been thoroughly examined by LaValle (1968). In strategic situations, the value of additional information might well be negative, even when it is being sold by one decision maker to another (see Weber and Croson 2003 for an example).



**Fig. 6** Value of information as a function of  $(\kappa, \lambda)$  in Example 2 for  $\frac{c}{1-c} \frac{1-p}{p} = \frac{1}{4}$

is preferable, whereas signal 1 is a better deal if the cost difference  $\Delta C$  exceeds  $\bar{\gamma}_\kappa(\kappa_2 - \kappa_1) + \bar{\gamma}_\lambda(\lambda_2 - \lambda_1)$ . □

### 4 Discussion

The examples in this article illustrate the usefulness of considering binary information sources in terms of their confidence parameters  $(\kappa, \lambda)$  instead of the standard error parameters  $(\varepsilon, \pi)$ . A key reason for this is that the confidence parameters naturally link the performance of the information source to the different messages it can send, thus rendering the Bayes update after a given signal realization dependent solely on a single parameter and the decision maker’s prior belief. This establishes a direct correspondence between the magnitude of a Bayes update and the confidence parameters, which also allows for a simple graphical Bayes update (cf. Appendix A). Relative to an event  $X$ , the parametrization idea naturally generalizes to *any* information source with an arbitrary number of messages, i.e., to any partition  $\mathcal{P}_S = \{S_i\}_{i \in \mathcal{I}}$  of the sample space  $\mathcal{S}$ ,<sup>14</sup> where  $\mathcal{I}$  is a countable index set. It is enough to define a (generalized) confidence  $\kappa_i(X) \in [0, \infty)$  for any possible realization  $S_i$  by setting  $\kappa_i(X) = P(S_i|X)/P(S_i|\bar{X})$  for all  $i \in \mathcal{I}$ . Hence, the confidence parametrization presented in Sect. 2 generalizes to *any* information source relative to *any* event  $X$ .<sup>15</sup> Nonetheless, by doing this, one

<sup>14</sup> The partition  $\mathcal{P}_S$  of  $\mathcal{S}$  has at most countably many elements  $S_i$ , which are pairwise disjoint and whose union is equal to  $\mathcal{S}$  (up to a set of Lebesgue-measure zero).

<sup>15</sup> In the general case, the Bayesian update diagram can also be used to find the posterior belief  $P(X|S_i) = g(\kappa_i(X), p)$  (corresponding to relation (34)), by *either* using the upper portion of the diagram for  $\kappa = \kappa_i(X)$  when  $\kappa_i(X) \geq 1$  or else using its lower portion for  $\lambda = 1/\kappa_i(X) > 1$ .

invariably associates a given message with a given event, and it is therefore simpler to think of the information source as binary in the first place.

The equivalent comparison based on confidence parameters can be generalized to nonbinary information sources as follows: if—relative to an event  $X$ —the confidence parameters  $\hat{\kappa}_i(X)$  of one source weakly exceed the confidence parameters  $\kappa_i(X)$  of another source, for all  $i \in \mathcal{I}$ , a decision maker is said to be at least as confident in the first source as in the second as an indicator for  $X$ . With this definition, all our results in Sect. 3 carry over in the more general framework, which is detailed for any finite outcome and sample space in Appendix D. The Blackwell order still implies the confidence order, which proves to be a useful weaker order than Blackwell's informativeness criterion. Such a generalization is of mostly a theoretical rather than a practical interest. The diagrams in Figs. 2 and 4 may be applicable by real-world decision makers, without requiring much knowledge of probability theory.<sup>16</sup> The concept of  $X$ - and  $\bar{X}$ -gains helps in deciding (via the confidence criterion (26) in Proposition 3) whether binary information sources for a given decision problem can be effectively compared using the confidence order. One of the key heuristic insights in Sect. 3 is that if these adjustment gains are similar, then the confidence order leads to a sharp distinction of information sources (cf. Fig. 5).

**Acknowledgements** The author would like to thank an anonymous reviewer, Sam Chiu, Ron Howard, Elisabeth Paté-Cornell, Ross Shachter, and Jeffrey Strnad for helpful suggestions and discussions.

## Appendix

### A. Graphical methods

The following examples illustrate the use of the Bayesian update diagram (Fig. 2) and the confidence-augmented ROC space (Fig. 4) for the evaluation and comparison of binary information sources.

*Example 3* A decision maker's prior belief about the probability that an important event  $X$  occurs is  $p = P(X) = 50\%$ . To update his beliefs, the decision maker can conduct a binary experiment with realizations  $S$  or  $\bar{S}$ . This experiment has as error characteristics  $(\varepsilon, \pi) = (0.25, 0.5)$ . From Lemma 2, we obtain the equivalent confidence parameters,  $(\kappa, \lambda) = (\pi/\varepsilon, (1 - \varepsilon)/(1 - \pi)) = (2, 3/2)$ . Figure 7 shows how to use the Bayesian update diagram to find the decision maker's posterior beliefs,  $P(X|S) \approx 67\%$  in the case of  $S$  and  $P(X|\bar{S}) \approx 40\%$  in the case of  $\bar{S}$ . The diagram also shows that using an experiment with higher confidence  $(\hat{\kappa}, \hat{\lambda}) = (3, 10) > (\kappa, \lambda)$ , the decision maker can increase the spread in his posterior belief adjustments, to 75% in case of  $S$  and 9% in case of  $\bar{S}$ .  $\square$

<sup>16</sup> In contrast to the Bayes update, the graphical comparison of information sources via iso-confidence lines in the ROC space as presented in Appendix A does not generalize to general information sources in a straightforward manner. The reason is that the confidence-augmented ROC space, which directly corresponds to the parametrization of the signal characteristics in terms of  $P(S_i|X)$  for all  $i \in \mathcal{I}$ , has as many dimensions as there are elements in the index set  $\mathcal{I}$ .

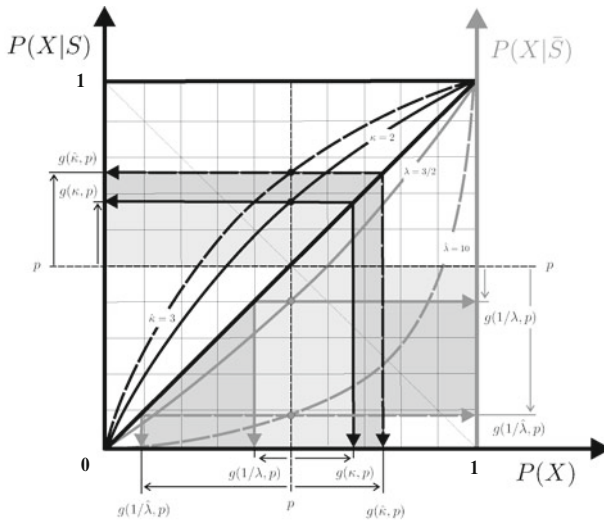
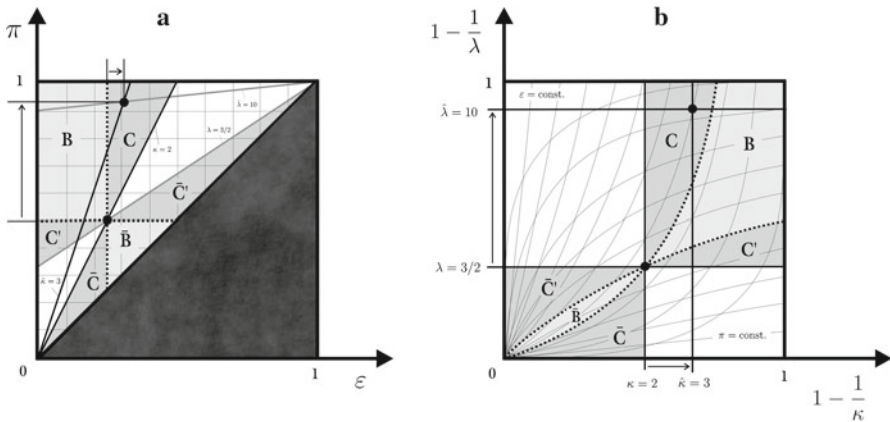


Fig. 7 Bayesian updates in Example 3

While the last example illustrates the straightforward use of the Bayesian update diagram in Fig. 2 to find posterior beliefs, given prior beliefs and confidence characteristics of a binary signal, there are other ways in which the diagram can be used. In certain situations, the posterior beliefs might be given, the signal characteristics known, and one seeks the relevant prior beliefs. Or, it may be the case that one wishes to find the appropriate confidence characteristics of a test that has the power to change a given prior belief into a given posterior belief, consistent with Lemma 1.

*Example 4* A prosecution lawyer is interested in obtaining forensic evidence that could prove the guilt of an accused “beyond a reasonable doubt.” While the precise probability standard for the required burden of proof may be disputable, for any given probability standard, say, 90% to fix ideas, the lawyer can focus on the  $\kappa$  needed to convince the most conservative juror, whose prior belief the lawyer estimates at  $p \in (0, 1)$ .<sup>17</sup> For instance, if  $p = 10\%$ , then from the Bayesian update diagram she can conclude that the proof must be the realization of a signal with confidence parameter  $\kappa$  somewhere between 50 and 100. Alternately, for a given piece of evidence whose probabilistic characteristics can be judged fairly, the lawyer can use the diagram to estimate the lowest prior that was necessary to lift beliefs above the required standard of proof. It is clear that the defense team can use the diagram in exactly the opposite way, introducing evidence with a  $\lambda$  as large possible to cast doubt on the defendant’s guilt. As long as different, successively introduced pieces of evidence can be considered statistically independent, it is possible to cycle several times through the diagram, each time using a newly obtained posterior belief as the prior belief when considering the next piece of evidence. □

<sup>17</sup> Our discussion is notwithstanding the question whether a juror is actually capable to perform correct Bayesian inferences; see [Kaye and Koehler \(1991\)](#) for a discussion of this question.



**Fig. 8** Comparing information sources **a** in ROC Space, **b** in confidence space (cf. Example 3)

The application in Example 2 represents a class of recurring binary classification problems, in which one of two possible classification actions is costly and its payoff depends on the realization of a random variable. The confidence order precisely orders signals of different values, whereas the Blackwell criterion is unable to distinguish certain signals. The latter point is made more precise by reconsidering Example 3.

*Example 3* (continued) A decision maker needs to decide between two different experiments, with respective confidence parameters  $(\kappa, \lambda) = (2, 3/2)$  and  $(\hat{\kappa}, \hat{\lambda}) = (3, 10)$ , in order to update beliefs about an important event. Our earlier discussion already showed that the Blackwell order cannot distinguish between the two experiments. In Fig. 8a and b, region  $B$  describes all the tests in the ROC space (cf. also Fig. 4) that are Blackwell-preferred to  $(\kappa, \lambda)$ . Regions  $C$  and  $C'$ , on the other hand, contain all signals which are not in  $B$  but still dominate  $(\kappa, \lambda)$  in the confidence order. The regions  $\bar{C}$ ,  $\bar{C}'$ , and  $\bar{B}$  are defined analogously, only that they contain the experiments that are dominated by the point  $(\kappa, \lambda)$  in the confidence-sense or Blackwell-sense, respectively. From Fig. 8, it becomes evident that the confidence order makes significantly more information sources comparable than the Blackwell order. If the cost of the decision maker’s action is given by  $c$  and the state of the world is described by a random payoff of either 0 or 1, then we are back to the same decision problem as in Example 2. Since  $V^*$  is nondecreasing in the confidence parameters, we can conclude that the experiment  $(\hat{\kappa}, \hat{\lambda})$  must be strictly preferred to the experiment  $(\kappa, \lambda)$ . □

Many real-world decisions need to be taken quickly and in the presence of substantial uncertainty. It is often impossible to carefully construct a probabilistic model and use sound decision-analytic tools to select optimal actions. Real-world decision makers often lack the training and intuition necessary to integrate different pieces of information and to then correctly update their probability assessments. For example, credit approval decisions have to be made online and are based on a very limited amount of information; a juror in court needs to update her subjective beliefs about a defendant’s guilt in a room filled with other jurors; a doctor needs to decide about how to best come-up with a reliable diagnosis while a patient is waiting; and, if she is

ever to finish her pile, an admissions officer has to quickly integrate pieces of evidence such as transcript and reference letters when deciding about whether or not to extend an offer to an applicant. A common feature of these choice situations is that a human decision maker is confronted with a fairly simple classification decision (e.g., yes/no) or with a simple information acquisition problem (e.g., which piece of evidence or test to select). The techniques for visualizing the Bayes' belief updates following the realization of informative signals and for graphically comparing different binary information sources and experiments can be used ad hoc and without first generating a formal model of the decision problem. The strength of the Bayesian update diagram (Fig. 2) and the confidence-augmented ROC space (Fig. 4) is that these visual aids can be used to make fairly precise normative judgements without a lot of formal analysis.

## B. Information aggregation in confidence parameters

At the end of Example 4, the question arose how it may be possible to aggregate different pieces of evidence. Such problems of *information aggregation* are ubiquitous in real-life decision problems and are encountered whenever decision makers have access to multiple information sources. Clemen and Winkler (1993) take a general perspective for the aggregation of information based on influence diagrams, where the different information sources might be weighted differently (e.g., depending on the varying importance of the different expert opinions). For the purposes of discussing the confidence order, here we make the demanding assumption that the decision maker knows the joint probabilistic behavior of his different information sources.<sup>18</sup> The confidence parametrization of information sources in terms of  $(\kappa, \lambda)$  is particularly well suited for obtaining a description of an aggregate information source in terms of the characteristics of its constituent information sources, at least when those sources are statistically independent conditional on the event of interest. The following lemma provides a general solution for the aggregation of two information sources. More than two experiments can be aggregated by applying the result several times in a row.

**Lemma 3** *Let  $\tilde{s}_1$  and  $\tilde{s}_2$  be two information sources on the common sample space  $\mathcal{S}$ . (i) If the events  $S_1, S_2 \subset \mathcal{S}$ , with  $0 < P(S_1|X), P(S_2|X) < 1$ , are statistically independent conditional on  $X$ , then the decision maker's possible posterior beliefs are*

$$\begin{cases} P(X|S_1 \cap S_2) = g(\kappa_1 \kappa_2, p), \\ P(X|S_1 \cap \bar{S}_2) = g(\kappa_1 / \lambda_2, p), \\ P(X|\bar{S}_1 \cap S_2) = g(\kappa_2 / \lambda_1, p), \\ P(X|\bar{S}_1 \cap \bar{S}_2) = g(1 / (\lambda_1 \lambda_2), p), \end{cases} \quad (33)$$

*depending on the joint realizations of the two signals. (ii) If the events  $S_1$  and  $S_2$  are statistically dependent conditional on  $X$ , then the decision maker's posterior beliefs*

<sup>18</sup> Much of the literature on the aggregation of information sources deals precisely with the question of how to generate a reasonable probability distribution for the joint experiments when combining the observations from the available information sources.



can be obtained as

$$P(X|S) = g\left(\frac{P(S|X)}{P(S|\bar{X})}, p\right), \quad (34)$$

where  $S \in \{S_1 \cap S_2, S_1 \cap \bar{S}_2, \bar{S}_1 \cap S_2, \bar{S}_1 \cap \bar{S}_2\}$  is the relevant joint event.

*Proof* (i) Consider first the case where the decision maker observes the message  $S_1 \cap S_2$ . Because of the assumed conditional independence of the two signals, it is

$$P(S_1 \cap S_2|X) = P(S_1|X)P(S_2|X) = \kappa_1\kappa_2,$$

where  $\kappa_1$  and  $\kappa_2$  are defined as in (8). Thus, by setting  $\kappa = \kappa_1\kappa_2$  and  $S = S_1 \cap S_2$  we obtain the first relation in part (i) of Lemma 3. The other relations hold by analogous arguments. (ii) The assertion follows directly from Bayes' rule (6) and the definition of  $g$  in (10),

$$P(X|S) = \frac{\frac{P(S|X)}{P(S|\bar{X})} p}{1 + \left(\frac{P(S|X)}{P(S|\bar{X})} - 1\right) p} = g\left(\frac{P(S|X)}{P(S|\bar{X})}, p\right),$$

for any  $S \in \{S_1 \cap S_2, S_1 \cap \bar{S}_2, \bar{S}_1 \cap S_2, \bar{S}_1 \cap \bar{S}_2\}$ .  $\square$

From Lemma 3, we conclude that conditionally independent information sources can be aggregated by multiplying their event-specific confidence characteristics. Morris (1977) has obtained an analogous multiplicative result for combining the judgements of independent experts. Part (ii), which is a generalization of part (i), states that whether information sources are statistically independent or not, it is always possible to form posterior beliefs using relations (8) and (9) by taking  $S$  as the appropriate joint event generated by the realizations of the two constituent signals. In that case, the confidence ratio  $P(S|X)/P(S|\bar{X})$  can be any nonnegative number and does not have to be  $> 1$ .

*Example 4* (continued) The prosecution lawyer introduced at the beginning of Example 4 is now interested in determining what "confidence" his case needs in order to convince the most conservative juror on the panel. With an estimated prior of  $p$  for the most conservative juror and a required burden of proof of at least 90%, the prosecution lawyer can try to assemble different pieces of evidence with an aggregate confidence of at least  $\kappa \geq (0.9/(1 - 0.9))(1 - p)/p$ . This lower bound is obtained from solving (7) for  $\kappa$ , given a left-hand side of 0.9. For example, when  $p = 10\%$ , the minimum aggregate confidence needed is 81. This number (which can also be obtained approximately using the Bayesian update diagram in Fig. 2) would still need to be multiplied by the aggregate confidence  $\lambda$  of the defense team's case for proving the defendant's innocence, in order to yield a conviction. Thus, if the lawyer estimates the defense team's  $\lambda$  at 10, it would be a good policy for the prosecution team to strive for an aggregate  $\kappa$  of at least 810. The aggregation can be performed pairwise using Lemma 3 repeatedly. Naturally, these results are indicative only of the strength of the evidence needed, as jurors may not be perfect Bayesian decision makers or may be naive in interpreting the evidence presented in court (Froeb and Kobayashi 1996).  $\square$

The last example illustrates the fact that the confidence parametrization of binary signals and the Bayesian update diagram in Fig. 2 can be used to solve the “inverse problem” of finding the minimum confidence needed to ensure that a decision maker (e.g., a juror) update a prior belief to a prescribed threshold level. We group such applications in the realm of *robust decision making*, since they focus on performance guarantees for worst-case scenarios. The example also illustrates the *odds interpretation* of the Bayes update in relation (7) (cf. Barnard 1949; Lindley 1964). Indeed, if we denote by  $h = p/(1 - p)$  the decision maker’s “prior odds,” and by  $\hat{h}(S) = P(X|S)/(1 - P(X|S))$  the decision maker’s “posterior odds” conditional on receiving the message  $S$ , then (7) can be written equivalently in the form

$$\hat{h}(S) = \kappa h, \tag{35}$$

which relates prior odds and posterior odds in a simple proportional relationship. Similarly, if  $\hat{h}(\bar{S})$  represents the posterior odds conditional on receiving the message  $\bar{S}$ , then

$$\hat{h}(\bar{S}) = (1/\lambda)h. \tag{36}$$

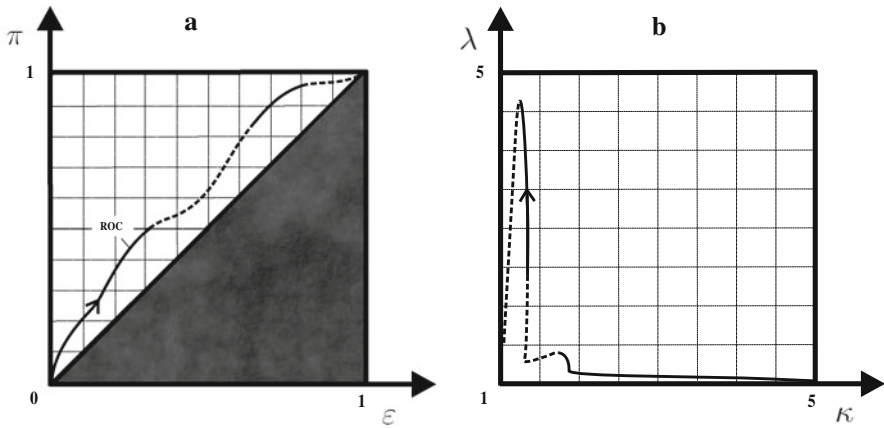
The odds interpretation of Bayes’ updating formula in (35) and (36) is useful for determining critical confidence parameters as appropriate ratios of prior and posterior odds. In Example 4, the prior odds are  $h = 0.1/(1 - 0.1) = 1/9$  and the posterior odds are  $\hat{h}(S) = 0.9/(1 - 0.9) = 9$ , which by virtue of (35) implies the lower bound,  $\kappa \geq \hat{h}(S)/h = 81$ .

*Remark 7* Aggregating different information sources generally leads to a new information source with more than two messages. While there are useful applications of this within our framework (cf. Example 4), we restrict attention in the main text to binary information sources to be able to formulate graphical methods for evaluating and comparing information sources. Appendix D introduces a generalization for nonbinary decision problems and experiments. □

### C. Confidence-consistent ROC curves

When considering an experimental method that is represented as an ROC curve (as in Fig. 9a) in terms of its confidence properties, one may be able to eliminate some portions of the curve from consideration if they are confidence-dominated by other points of the curve. The Blackwell order allows such an elimination only if the ROC curve is not monotonically increasing. The confidence order, by contrast, also allows the elimination of certain nonconvexities that naturally appear when ROC curves are generated by interpolating experimental data points. To identify dominated points, one first considers an equivalent representation  $\mathcal{R} = \{(\kappa(d), \lambda(d))\}_{d \in \mathcal{D}}$  of the ROC curve in confidence parameters (cf. Fig. 9b) and eliminates from  $\mathcal{R}$  any strictly dominated points, so that the remaining truncated graph  $\hat{\mathcal{R}}$  is given by

$$\hat{\mathcal{R}} = \{(\kappa, \lambda) \in \mathcal{R} : (\kappa, \lambda) \leq (\hat{\kappa}, \hat{\lambda}) \in \hat{\mathcal{R}} \Rightarrow (\kappa, \lambda) = (\hat{\kappa}, \hat{\lambda})\}.$$



**Fig. 9** Elimination of confidence-dominated designs from an ROC curve (a) by restricting attention to confidence-undominated points of the equivalent representation (b) in  $(\kappa, \lambda)$ -coordinates

Transforming the truncated graph  $\hat{\mathcal{R}}$  back into ROC space, one obtains a (generally discontinuous) receiver operating characteristic that does not contain any experiments that are confidence-dominated by other experiments on the curve.<sup>19</sup>

D. Generalization to nonbinary state and sample spaces

We now provide a generalization of the binary framework to decision problems with any finite state space  $\mathcal{X} = \{X_0, \dots, X_n\}$  and experiments with any finite sample space  $\mathcal{S} = \{S_0, \dots, S_m\}$ , where  $m, n$  are given positive integers.<sup>20</sup> For  $m = n = 1$ , we recover the binary situation discussed in the main text. To keep things nontrivial, we assume that the decision maker’s belief  $p_i = P(X_i)$  about the likelihood of the outcome  $X_i$  is neither 0 nor 1 so that

$$p_i \in (0, 1)$$

for all  $i$ . The information source (experiment) is characterized by nontrivial conditional probabilities

$$\pi_{ij} = P(S_j|X_i) \in (0, 1)$$

for all  $i, j$ . Analogous to the binary situation, we refer to the likelihood ratio

$$\kappa_{ij} = \frac{P(S_j|X_i)}{P(S_j|\bar{X}_i)}$$

<sup>19</sup> Flach and Wu (2003) offer a different rationale for eliminating nonconcavities (referred to as “non-convexities” in their framework) by constructing the convex hull of the empirical ROC curve, under the assumption of linear tradeoffs between different objectives.

<sup>20</sup> A generalization to more general (measurable) outcome and sample spaces is straightforward but is beyond the scope of this article.

as the *confidence of the message*  $S_j$  as an indicator for the event  $X_i$ . A value of  $\kappa_{ij} < 1$  means that  $S_j$  tends to disconfirm the realization of  $X_i$  and should therefore be viewed as an indicator for  $\bar{X}_i$ , the complement of  $X_i$ . On the other hand, if  $\kappa_{ij} > 1$ , then the decision maker's posterior belief conditional on the realization of  $S_j$  exceeds his prior, i.e.,

$$\kappa_{ij} > 1 \Rightarrow P(X_i|S_j) > p_i.$$

The latter implication (and its converse) holds, because, by virtue of the same arguments as in Sect. 2, the decision maker's posterior is

$$P(X_i|S_j) = g(\kappa_{ij}, p_i),$$

where the function  $g(\cdot)$  is given by (10). As a result, the Bayesian update diagram (Fig. 2) can be used, depending on the magnitude of  $\kappa_{ij}$ , by either replacing  $\kappa$  by  $\kappa_{ij} \geq 1$  or replacing  $\lambda$  by  $1/\kappa_{ij} > 1$ . We consider the *standard decision problem* of finding the optimal action

$$a_j^* \in \arg \max_{a \in \mathcal{A}} \left\{ \sum_{i=0}^n p_i \pi_{ij} u(a, X_i) \right\}$$

for all signal realizations  $j \in \{0, \dots, m\}$ . All other primitives of the decision problem are defined as in Sect. 3.1. Analogous to the binary situation, the decision maker's action contingent on the message  $S_j$  depends only on  $\kappa_j = (\kappa_{0j}, \dots, \kappa_{mj})$ , not on any other characteristic of the information source. Since for any event  $X_i$  the conditional probabilities  $P(S_j|X_i)$  sum up to 1, i.e.,  $\sum_{j=0}^m \pi_{ij} = 1$ , we can choose a *reference message*, say  $S_0$ , and express the conditional probabilities  $\pi_{i0}$  by the conditional probabilities for the other messages, i.e.,

$$\pi_{i0} = 1 - \sum_{j=1}^n \pi_{ij},$$

for all  $i$ , and then we obtain for the expected value of the decision problem

$$V^* = \sum_{i=0}^n p_i \left[ u(a_0^*, X_i) + \sum_{j=1}^n \pi_{ij} \left( u(a_j^*, X_i) - u(a_0^*, X_i) \right) \right].$$

In other words, the decision value is obtained as the sum of the expected value of the decision conditional on the reference message, plus the (expected)  $X_i$ -gains  $\Delta_{ij}$  of  $a_j^*$  over  $a_0^*$ ,

$$\Delta_{ij} = p_i \left( u(a_j^*, X_i) - u(a_0^*, X_i) \right),$$

weighted by the conditional probabilities  $\pi_{ij}$ . In other words,

$$V^* = \bar{u}_0^* + \sum_{i,j=0}^{n,m} \pi_{ij} \Delta_{ij},$$

where  $\bar{u}_0^* = \sum_{i=0}^n p_i u(a_0^*, X_i)$ . Note that in the binary case it was possible to switch the reference message with the state of the world considered (taking as reference message always the signal realization that represents the strongest indicator for a given state), which led to the *nonnegative* gains  $\Delta$  and  $\bar{\Delta}$  in (20) and (21). This convenient definition for the binary case proves to be cumbersome in the general situation. Hence, the  $X_i$ -gains  $\Delta_{ij}$  can have either sign, depending on how  $a_j^*$  compares to  $a_0^*$  in terms of the decision maker’s payoffs given the outcome  $X_i$ .

To find how the decision value  $V^*$  changes in the confidence parameters  $\kappa_{ij}$ , we first note that all these parameters cannot be chosen independently. Indeed, if we let  $\pi_j = (\pi_{0j}, \dots, \pi_{nj})$  and  $p = (p_0, \dots, p_n)$ , then

$$\kappa_{ij} = \frac{(1 - p_i)\pi_{ij}}{(p \cdot \pi_j) - p_i\pi_{ij}},$$

which implies that  $p_i\pi_{ij} = \kappa_{ij}(p \cdot \pi_j)/(1 - p_i + p_i\kappa_{ij})$ , whence (by summing over  $i$ ) the consistency condition

$$\sum_{i=0}^n \frac{p_i\kappa_{ij}}{(1 - p_i) + p_i\kappa_{ij}} = 1$$

for all  $j$ . We therefore term one outcome, without loss of generality  $X_0$ , a *reference state* and restrict attention to changes of  $\kappa_{ij}$  relative to states and messages other than this reference state and other than the reference message, i.e., for  $i, j \geq 1$ . Given a matrix of confidence parameters  $K = [\kappa_{ij}]_{i,j=0}^{n,m}$  (with elements satisfying the above consistency condition), the corresponding matrix of conditional probabilities  $\Pi = [\pi_{ij}]_{i,j=0}^{n,m}$  is given by

$$\Pi = Q(K, p),$$

where  $Q(K, p) = [Q_{ij}(K, p)]_{i,j=0}^{n,m} \in \{[y_{ij}]_{i,j=0}^{n,m} : \kappa_{ij} = (1 - p_i)y_{ij}/(p \cdot y_j - p_i y_{ij}), \sum_{k=0}^m y_{ik} = 1\}$ . The following result (which is a simple consequence of the chain rule of differentiation) summarizes how the decision value  $V^*$  changes as a function of the confidence parameters of the experiment.

**Lemma 4** *The change of the decision value  $V^*$  in the confidence parameter  $\kappa_{ij}$  is given by*

$$\frac{\partial V^*}{\partial \kappa_{ij}} = \sum_{i=0}^n \Delta_{ij} q_{ij},$$

where  $q_{ij} = \partial Q_{ij}(K, p)/\partial \kappa_{ij}$ .

In our confidence order a signal with confidence-parameter matrix  $\hat{K}$  (weakly) dominates a signal with confidence-parameter matrix  $K$ , denoted  $\hat{K} \succeq K$ , if and only if

$$(\kappa_{ij} - 1)(\hat{\kappa}_{ij} - \kappa_{ij}) \geq 0$$

for all  $i, j$ . In other words,  $K$  is (weakly) dominated by  $\hat{K}$  if for any confidence parameter  $\kappa_{ij}$  in  $K$  that exceeds 1, the corresponding confidence parameter  $\hat{\kappa}_{ij}$  exceeds  $\kappa_{ij}$  (at least weakly), while the converse is true (i.e.,  $\hat{\kappa}_{ij} \leq \kappa_{ij}$ ) if  $\kappa_{ij} < 1$ . Therefore, a decision maker can be more confident in a signal if messages become stronger indicators for (or against) any given outcome  $X_i$  in  $\mathcal{X}$ . Hence, in the context of general finite outcome and message spaces we say that a decision problem is “balanced” if

$$\hat{K} \succeq K \Rightarrow \sum_{i=0}^n \Delta_{ij} q_{ij} \geq 0.$$

*Remark 8* We now show that the above generalized confidence-order condition reduces to the balancedness conditions in the binary case, where  $m = n = 1$ . In that case,  $\kappa_{11} = 1/\kappa_{01} = \kappa$ , and  $\kappa_{00} = 1/\kappa_{10} = \lambda$ , where—without any loss in generality—we have that  $\kappa, \lambda \geq 1$ . Furthermore, with  $X_0 = \bar{X}$  and  $X_1 = X$  it is  $\pi_{11} = 1 - \pi_{10} = \pi$ ,  $\pi_{01} = 1 - \pi_{00} = \varepsilon$ , and  $\Delta_{11} = \Delta$ ,  $\Delta_{10} = \Delta_{00} = 0$ , and  $\Delta_{10} = -\bar{\Delta}$ . The function  $Q(K, p)$  corresponds in the binary case to the mapping in Lemma 2,

$$Q(K, p) = \frac{1}{\kappa\lambda - 1} \begin{bmatrix} \lambda(\kappa - 1) & \lambda - 1 \\ \kappa - 1 & \kappa(\lambda - 1) \end{bmatrix},$$

and is independent of the decision maker’s prior belief  $p$ . Differentiating  $Q$  with respect to  $(\kappa, \lambda)$  yields the expression on the right-hand side of (27) so that we immediately recover the binary balancedness condition (or confidence criterion) in the form (28) from the more general approach. □

### References

Athey, S. (2000). *Investment and information value for a risk averse firm*. Working paper 00-30, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA.

Athey, S., & Levin, J. (2001). *The value of information in monotone decision problems*. Working paper 98-24, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA.

Barnard, G. A. (1949). Statistical inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 11(2), 115–149.

Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society London*, 53, 370–418. (Reprinted in: Barnard, G. A. (1958). *Biometrika*, 45(3/4), 293–315).

Berger, J.O., & Wolpert, R. (1984). *The likelihood principle*. Lecture notes, monograph series, Vol. 61, Hayward, CA: Institute of Mathematical Statistics.

Birnbaum, A. (1961). On the foundations of statistical inference: Binary experiments. *Annals of Mathematical Statistics*, 32(2), 414–435.

Birnbaum, A. (1962). On the foundations of statistical inference. *Journal of the American Statistical Association*, 57(298), 269–306.

- Birnbaum, A., & Maxwell, E. (1960). Classification procedures based on Bayes's formula. *Applied Statistics*, 9(3), 152–169.
- Blackwell, D. (1951). Comparison of experiments. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability* (pp. 93–102). Berkeley: University of California Press.
- Blackwell, D. (1953). Equivalent comparison of experiments. *Annals of Mathematical Statistics*, 24(2), 265–272.
- Bohnenblust, H., Shapley, L., & Sherman, S. (1949). *Reconnaissance in game theory*. Research memorandum RM-208, RAND Corporation.
- Cavusoglu, H., & Raghunathan, S. (2004). Configuration of detection software: A comparison of decision and game theory approaches. *Decision Analysis*, 1(9), 131–148.
- Clemen, R. T., & Winkler, R. L. (1993). Aggregating point estimates: A flexible modeling approach. *Management Science*, 39(4), 501–515.
- DeGroot, M. (1962). Uncertainty, information, and sequential experiments. *Annals of Mathematical Statistics*, 33(2), 404–419.
- Eves, H. W. (1988). *Return to mathematical circles*. Boston, MA: Prindle, Weber, and Schmidt.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
- Fishburn, P. C. (1970). *Utility theory for decision making*. New York, NY: Wiley.
- Flach, P. A., & Wu, S. (2003). Repairing concavities in ROC curves. In *Proceedings of the 2003 UK workshop on computational intelligence* (pp. 38–44). Bristol, UK: University of Bristol.
- Froeb, L. M., & Kobayashi, B. H. (1996). Naive, biased, yet Bayesian: Can juries interpret selectively produced evidence? *Journal of Law, Economics, and Organization*, 12(1), 257–276.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York, NY: Springer.
- Hill, B. M. (1987). The validity of the likelihood principle. *American Statistician*, 41(2), 95–100.
- Hoel, P. G., & Peterson, R. P. (1949). A solution to the problem of optimum classification. *Annals of Mathematical Statistics*, 20(3), 433–438.
- Kaye, D. H., & Koehler, J. J. (1991). Can jurors understand probabilistic evidence? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 154(1), 75–81.
- Keisler, J. (2004). Value of information in portfolio decision analysis. *Decision Analysis*, 1(3), 177–189.
- Kihlstrom, R. E. (1984). A Bayesian exposition of Blackwell's theorem on the comparison of experiments. In M. Boyer & R. E. Kihlstrom (Eds.), *Bayesian models in economic theory* (pp. 13–31). North-Holland, Amsterdam, NL: Elsevier Science.
- Kreps, D. (1988). *Notes on the theory of choice*. Boulder, CO: Westview Press.
- Kriege, M., Brekelmans, C. T., Boetes, C., et al. (2004). Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *New England Journal of Medicine*, 351(5), 427–437.
- Lasko, T. A., Bhagwat, J. G., Zou, K. H., & Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, 38(5), 404–415.
- LaValle, I. H. (1968). On cash equivalents and information evaluation in decisions under uncertainty. *Journal of the American Statistical Association*, 63(321), 252–290.
- Leach, M. O., Boggis, C. R. M., Dixon, A. F. et al. (2005). Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: A prospective multicentre cohort study (MARIBS). *Lancet*, 365(9473), 1769–1778.
- Lehmann, E. L. (1988). Comparing location experiments. *Annals of Statistics*, 16(2), 521–533.
- Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, 27(4), 986–1005.
- Lindley, D. V. (1964). The Bayesian analysis of contingency tables. *Annals of Mathematical Statistics*, 35(4), 1622–1643.
- Milgrom, P., & Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2), 583–601.
- Morris, P. A. (1977). Combining expert judgments: A Bayesian approach. *Management Science*, 23(7), 679–693.
- Nemec, C. F., Listinsky, J., & Rim, A. (2007). How should we screen for breast cancer? Mammography, ultrasonography, MRI. *Cleveland Clinic Journal of Medicine*, 74(12), 897–904.
- Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London: Series A*, 231, 289–337.
- Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford, UK: Oxford University Press.

- Radner, R., & Stiglitz, J. E. (1984). A nonconcavity in the value of information. In M. Boyer & R. E. Kihlstrom (Eds.), *Bayesian models in economic theory* (pp. 33–52). North-Holland, Amsterdam, NL: Elsevier Science.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656.
- Shannon, C. E., & Weaver, W. (1948). *The mathematical theory of communication*. Chicago, IL: University of Illinois Press.
- Strulovici, B. H., & Weber, T. A. (2008). Monotone comparative statics: Geometric approach. *Journal of Optimization Theory and Applications*, 137(3), 641–673.
- Wald, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *Annals of Mathematical Statistics*, 10(4), 299–326.
- Weber, T. A., & Croson, D. C. (2003). Selling less information for more: Garbling with benefits. *Economics Letters*, 83(2), 165–171.