

A new nonparametric test for the race model inequality

Luigi Lombardi, Marco D'Alessandro,
and
Hans Colonius*

Department of Psychology and Cognitive Science
University of Trento

*Department of Psychology
University of Oldenburg

Send all correspondence to:

Luigi Lombardi

Department of Psychology and Cognitive Science

Corso Bettini 31

I-38068 Rovereto (TN), Italy

Email: luigi.lombardi@unitn.it

Abstract

The race model inequality (RMI), as first introduced by Miller (1982), entails an upper bound on the amount of statistical facilitation for reaction times (RTs) attainable by a race model within the redundant-signals paradigm. A violation of RMI may be considered as empirical evidence for a coactivation model rather than a race model. Here, we introduce a novel nonparametric procedure for evaluating the RMI for single participant analysis. The statistical procedure is based on a new probabilistic representation that highlights some neglected, but important distributional features of the RMI. In particular, we show how the reconstructed distribution function under maximal statistical facilitation for a race model is characterized by a specific truncated-type property. The results of two Monte Carlo simulation studies suggest that our procedure efficiently controls for type I error with reasonable power. Finally, unlike previous proposals for single participant analysis (e.g., Maris and Maris, 2003), our approach is also more consistent with the typical way to collect RT data in experimental works. R script functions for running the statistical analysis on single participant data are made freely available to the readers on a dedicated web server.

Keywords: race model inequality, redundant-signals paradigm, truncated Kolmogorov-Smirnov test.

1 Introduction

A fundamental property of the human perceptual-motor system is its ability to deliver fast responses to signals presented in the environment. In the **redundant signals task**, the observer is required to react as quickly as possible, e.g. by pressing a response button, to the occurrence of a target stimuli. Typically, participants exhibit shorter response times (RTs) when two or more targets from different sensory modalities (visual, auditory, tactile) are presented simultaneously. The **expected** RT difference between single- and multiple-target conditions is called *redundancy gain*. In the literature, two alternative explanations have been suggested (Miller, 1982). The *race model* assumes that information from both targets is processed in separate sensory channels and that RT is primarily determined by the first channel (“winner of the race”) to transmit information to some more central interface in the processing hierarchy. Assuming that sensory processing time varies randomly from trial to trial, expected RT in the redundant condition is always smaller or equal to the minimum of the expected RTs **in the single target conditions** (Raab, 1962). Alternatively, the *coactivation model* assumes that redundancy gain occurs because information from each target is summed (somewhere) in the brain and that a response is triggered as soon as the combined information exceeds some threshold (Miller, 1982, 1986; Schwarz, 1994; Diederich, 1995).

The race model inequality (RMI), first suggested in Miller (1982), states that the RT distribution function for the redundant signal condition is never larger than the sum of the RT distribution functions of the two single signal conditions. Any violation of the RMI may be understood as an empirical support for a coactivation integration mechanism. In general, evaluation of the RMI has mainly been treated using statistical procedures for group data (e.g., Miller, 1982; Ulrich, Miller, & Schröter, 2007; Gondan, 2010), while statistical approaches for testing the RMI in single participant’s data is less common (e.g., Miller, 1986; Maris & Maris, 2003; Vorberg, 2008). **A first proposal for analyzing data in single participants appeared in a seminal work by Miller (1986) in which the violation of the RMI was evaluated using a bootstrapping procedure (Efron, 1979). More recently, a nonparametric test based on a modified version of the ordinary Kolmogorov-Smirnov (KS) statistic was introduced by Maris & Maris (2003). It was based on a probabilistic sampling procedure associated with a mixture distribution with equal weights for the two single target conditions. In practice, this procedure required that in each single signal trial a fair coin was tossed to determine in which**

channel the stimulus was presented next. This mixture distribution paradigm, however, had several practical limitations and was rarely met in most experimental work.

In this paper, we take advantage of the original idea introduced by Maris & Maris (2003) but reformulate it to propose a new nonparametric statistical procedure for evaluating the race model inequality for single-participant analysis. We will introduce a novel probabilistic context that highlights some neglected, but relevant distributional properties of the RMI. In particular, the proposed procedure is more consistent with the typical way to collect RT data in experimental works.

The manuscript is organized as follows. The next section specifies the RMI at both the population and sample level. Some relevant properties of the distribution functions associated with the RMI are discussed and a new statistical procedure based on a truncated version of the KS statistic is presented. In the third section, two simulation studies are reported to evaluate performance of the new statistical procedure. Finally, the fourth section presents discussion and conclusions.

2 Evaluating the race model inequality

In the redundant condition, let (S_1, S_2) be a pair of (jointly distributed) continuous random variables with S_1 the processing time for the first channel and S_2 for the second channel. Under the race model, define

$$Z = \min(S_1, S_2) \tag{1}$$

as random variable representing observable response time in the redundant condition, whereas S_1 and S_2 are considered unobservable variables (only their minimum is). According to the assumption of *context invariance* (Miller, 1982; Colonius, 1990; Luce, 1986), the marginal distributions of S_1 and S_2 are equal to the distributions of the RT random variables X and Y observed in the single target conditions, respectively¹.

Under context invariance, the distribution F_Z of Z equals:

$$F_Z(t) = F_X(t) + F_Y(t) - P(S_1 \leq t, S_2 \leq t) \tag{2}$$

¹Note that context invariance is a condition needed for testing the race model inequality but itself cannot be empirically tested (e.g. Luce, 1986).

where F_X and F_Y are the distribution functions of X and Y , respectively². The latter equation implies the following inequality:

$$F_Z(t) \leq \min\{1, F_X(t) + F_Y(t)\}, \quad \text{for all } t \quad (3)$$

It provides an upper bound for RT in the redundant condition under the race model and is called *race model inequality* (Miller, 1982). Writing

$$K(t) = \min\{1, F_X(t) + F_Y(t)\}, \quad \text{for all } t, \quad (4)$$

then $F_Z = K$ corresponds to a race model under *maximal attainable statistical facilitation*. More precisely, if S_1 and S_2 show maximum negative stochastic dependence with fixed marginals F_X and F_Y (Miller, 1982; Colonius, 1990, 2016), then a coupling exists (i.e., a bivariate distribution with maximal negative correlation) such that

$$\begin{aligned} P(S_1 \leq t_1, S_2 \leq t_2) &= C_{S_1 S_2}(t_1, t_2) \\ &= \min\{F_X(t_1), F_Y(t_2)\}, \quad \forall t_1, \forall t_2 \end{aligned}$$

with the mapping $C_{S_1 S_2}$ being a copula and where its diagonal section ($t_1 = t_2$) yields

$$F_Z(t) = K(t), \quad \text{for all } t.$$

Clearly, if $F_Z(t) > K(t)$ for some t , then the redundant signals effect is no longer compatible with predictions from the race model under context invariance. Distribution function $K(t)$ will be referred to as *reconstructed distribution function* below.

On the basis of the statistical facilitation principle and the context invariance assumption, we can formulate the following null hypothesis for testing the RMI:

$$H_0 : F_Z(t) \leq K(t), \quad \text{for all } t. \quad (5)$$

Since under the null hypothesis the distribution F_Z is most to the left if $F_Z = K$, we set the corresponding *point* null hypothesis as follows

$$H_0^* : F_Z(t) = K(t), \quad \text{for all } t. \quad (6)$$

²In this context, we assume that F_X and F_Y are absolutely continuous distribution functions with nonnegative arguments.

2.1 Population level representation

In what follows we will illustrate some relevant properties regarding the reconstructed distribution function K . Note that definitions, propositions, and corollaries are enumerated sequentially as they appear. We first introduce the so called *average distribution* for the single channel conditions:

$$M(t) = \frac{1}{2}(F_X(t) + F_Y(t)). \quad (7)$$

The average distribution M is a mixture distribution where the two original distribution functions F_X and F_Y have equal weights. **In our context, it is important to note that this mixture distribution constituted the main probabilistic token described by Maris and Maris (2003) in their original proposal. Here we will show how to modify this representation to adapt it to our new framework.**

Definition 1. Let $p \in (0, 1)$ and $\xi_p = M^{-1}(p) = \inf\{t : M(t) \geq p\}$ the p -quantile for the average distribution M . The p -truncated average distribution is represented as follows:

$$M_p(t) = \begin{cases} M(t)/M(\xi_p), & 0 < t \leq \xi_p \\ 1, & t > \xi_p \\ 0, & t = 0 \end{cases} \quad (8)$$

The p -truncated average distribution represents a measure of the fraction of the average distribution M up to the p -th quantile. The following proposition details the close relationship between an important instance of the truncated average distribution M_p and the reconstructed distribution K .

Proposition 2. The reconstructed distribution function K for the redundant condition is equivalent to the 0.5-truncated average distribution:

$$K(t) = M_{0.5}(t), \quad \text{for all } t. \quad (9)$$

Proof. First, for all $t \leq 0$ both distribution functions always take value 0 as F_X and F_Y have

nonnegative arguments. Next, suppose that $0 < t \leq \xi_{0.5}$, then

$$\begin{aligned} M_{0.5}(t) &= M(t)/M(\xi_{0.5}) \\ &= M(t)/0.5 \\ &= 2M(t) \\ &= K(t), \end{aligned}$$

as for all positive t that never exceed the 0.5-quantile of M , the rescaled distribution function $2M(t)$ cannot exceed 1. Finally, for all values of $t > \xi_{0.5}$ we have that $M(t) > 0.5$ and consequently $2M(t) > 1$ and so $K(t) = 1$. The same goes with $M_{0.5}(t)$ as by construction this distribution function is truncated at $\xi_{0.5}$. \square

Corollary 3. Under the maximal statistical facilitation assumption of a race model we have that:

$$F_Z = M_{0.5}.$$

Proof. Trivial. It immediately follows from Proposition 1 and the definition of the point null hypothesis (6). \square

In sum, the most relevant result of this section regards the characterization of the reconstructed RST distribution function K as a truncated-type distribution with truncation point localized at the population quantile $\xi_{0.5}$ of M . Moreover, under the point null hypothesis (6) also the RT distribution for the original redundant condition F_Z must be characterized by exactly the same property. **To our knowledge, this is the first time that this peculiar property about K has been explicitly characterized in the RMI literature. Later on in this paper (e.g., in the test procedure section), it will become clear how the truncation property of K directly plays a key role in the definition of the statistical test for the evaluation of the RMI in the redundant signals paradigm.**

2.2 Sample level representation

In this section we will introduce our main idea about the sampling procedure for the random variables involved in the RMI and the estimation of the corresponding distributions. Here, we will also highlight

the more relevant differences (at the sample level) between our approach and the original proposal by Maris and Maris (2003).

Let us assume that the redundant signal (RST) condition is associated with m distinct trials whereas the two single signal (SST) conditions are associated with n_1 and n_2 trials, respectively. Moreover, let $\mathcal{Z} = \{Z_1, Z_2, \dots, Z_m\}$, $\mathcal{X} = \{X_1, X_2, \dots, X_{n_1}\}$, and $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_{n_2}\}$ be the corresponding sets of random variables. In our context, we assume that each of the three sets is composed of independent identically distributed (i.i.d.) random variables with cumulative distribution functions (cdf) F_Z , F_X and F_Y , respectively. Taken together, the two SST conditions yield a combined set $\mathcal{W} = \mathcal{X} \cup \mathcal{Y} = \{X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}\}$ of $n_1 + n_2$ elements. In our representation, we also assume that all pairs $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ are independent. The latter *independent coupling condition* (Colonus, 2016) may easily be motivated by considering that the observations in the two SST conditions are usually recorded in separate experimental trials³. In sum, the combined set \mathcal{W} is understood as a collection of $n_1 + n_2$ independent and *non-identically* distributed random variables (e.g., Van Zuuijlen, 1978).

Definition 4. Given the combined set \mathcal{W} , the *average cumulative distribution function* of \mathcal{W} is represented as follows:

$$A(t) = \frac{1}{n_1 + n_2} (n_1 F_X(t) + n_2 F_Y(t)). \quad (10)$$

Unlike the average distribution M , the structure of A clearly depends on the relative frequencies of the random variables in \mathcal{W} for the two classes of distributions F_X and F_Y . Here we will highlight a relevant property of A when equal sample sizes are considered.

Proposition 5. For all sample sizes n , if $n_1 = n_2 = n$, then the average distribution A reduces to the population average distribution M .

³Unlike the observed pairs (X, Y) , the unobservable channel processing times (S_1, S_2) may be correlated

Proof. Trivially, for any $n = 1, 2, \dots$,

$$\begin{aligned} A(t) &= \frac{1}{2n}(nF_X(t) + nF_Y(t)) \\ &= \frac{1}{2}(F_X(t) + F_Y(t)) \\ &= M(t), \end{aligned}$$

for all t . \square

Note that, unlike the average mixture distribution, here A literally represents the average of distributions associated with a *fixed* number $2n$ of independent and (*non*)*identically* distributed random variables. The latter makes the sampling procedures for the two representations quite different. In particular, for the average mixture distribution M , the sampling always requires that each time an SST trial is administered in an experiment, a fair coin is tossed to determine to which channel the next stimulus is presented (e.g., Maris & Maris, 2003). By contrast, for the averaged distribution A (with equal sample sizes) we have a fixed number n of trials for each of the two SST conditions. Even if the two sampling procedures asymptotically converge to the same empirical distribution (i.e., when $n \rightarrow +\infty$), they may substantially deviate for small to medium sample sizes. In this work, we will take advantage of the equal sample size assumption to illustrate some useful estimation procedures for the reconstructed distribution function K . In particular, in the estimation of K we will replace the estimator of $M_{0.5}(t)$ in Eq. (9) with the corresponding estimator of $A_{0.5}(t)$ based on the observed experimental data for the combined set \mathcal{W} .

Now, let \mathbf{z} , \mathbf{x} , and \mathbf{y} be random samples associated with \mathcal{Z} , \mathcal{X} , and \mathcal{Y} , respectively. Hereafter, we assume that \mathbf{x} and \mathbf{y} have equal sample size n while \mathbf{z} has size m with no ties in the full data set⁴. Let $\widehat{F}_{Z,m}$, $\widehat{F}_{X,n}$, and $\widehat{F}_{Y,n}$ be the corresponding **empirical distribution functions** (ecdfs) associated with the three samples. It is well known that ecdfs are consistent and unbiased estimators of the associated **distribution functions under representative sampling** (e.g., Pratt & Gibbons, 1981). Moreover, because all transformations described in Eqs. (4,7-9) are continuous mappings, it follows that \widehat{K} , \widehat{A} , and $\widehat{A}_{0.5}$ are also consistent estimators of K , A , and $A_{0.5}$, respectively (Serfling, 1980). In particular, the

⁴The issues of unequal sample sizes and eventual ties in the samples will be discussed in more details in the final section of the manuscript.

reconstructed distribution K can be estimated using one of the following three equivalent procedures:

$$\widehat{K}(t) = \min\{1, \widehat{F}_{X,n}(t) + \widehat{F}_{Y,n}(t)\} \quad (11)$$

$$= \min\{1, 2\widehat{A}(t)\} \quad (12)$$

$$= \widehat{A}_{0.5}(t) \quad (13)$$

where $\widehat{A}^{-1}(0.5) = \widehat{\xi}_{0.5} = \inf\{t : \widehat{A}(t) \geq 0.5\}$ is the empirical 0.5-quantile computed on \widehat{A} . However, under the equal sample size assumption, the estimator for K can also be directly computed from the combined sample vector $\mathbf{w} = (\mathbf{x}, \mathbf{y})$. In particular, the ecdf of \mathbf{w} represents a direct estimator of the averaged distribution A . Similarly, the empirical 0.5-quantile $\widehat{\xi}_{0.5}$ for \widehat{A} is also equivalent to the n -th ordered value $w_{(n)}$ in \mathbf{w} . Therefore, a direct estimation of K can be obtained by computing the ecdf on the subsample of \mathbf{w} composed of its first n ordered observations: $w_{(1)}, w_{(2)}, \dots, w_{(n)}$. Table 1 reports a hypothetical data set which may serve as a guiding example to illustrate the different estimation procedures considered here. Several properties of the (empirical) averaged distribution functions and their relations with order statistics have been extensively investigated in the statistical literature (e.g., Van Zuuijlen, 1976; Van Zuuijlen, 1978).

Note that, at the sample level, the main difference between our proposal and the mixture representation by Maris and Maris (2003) pertains the way the reconstructed distribution K is estimated. In particular, in the mixture representation, \widehat{K} is estimated as follows

$$\widehat{K}(t) = \min\{1, 2\widehat{M}(t)\}$$

with \widehat{M} being the ecdf of the mixture sample drawn from the population mixture distribution M (see Eq. (7)) which makes the two proposals clearly different at the sample level as detailed earlier in the manuscript.

We conclude this section by highlighting some important relationships regarding the maximal statistical facilitation property for a race model, the combined set \mathscr{W} , and the antithetic variate method for continuous random variables. We recall that the method of *antithetic variates* (e.g., Thompson, 2000, Colonius & Diederich, 2006) can be used to generate a pair of maximally negative dependent random variables from the two SST conditions in line with the maximal statistical facilitation assumption for the RMI (Miller, 1982; Colonius, 1990).

Definition 6. Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ and $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ denote the order statistics for the sets \mathcal{X} and \mathcal{Y} , respectively. The antithetic variate transformations are defined as follows:

$$R_i = \min\{X_{(i)}, Y_{(n-i+1)}\}, \quad (14)$$

$$T_i = \max\{X_{(i)}, Y_{(n-i+1)}\}, \quad i = 1, \dots, n. \quad (15)$$

Note that, the distribution of the antithetic variates R_i ($i = 1, \dots, n$) corresponds to the reconstructed distribution for the redundant condition under maximal statistical facilitation. That is to say: $F_R = K$ as defined in (4).

Proposition 7. Let $W_{(1)}, W_{(2)}, \dots, W_{(2n)}$, (n even) denote the order statistics for the set \mathcal{W} , then

$$W_{(i)} = R_{(i)} \quad \text{and} \quad W_{(n+i)} = T_{(i)}, \quad i = 1, \dots, n.$$

Proof. The problem stated here is a particular instance of *merging networks*, which are networks that can merge two sorted input sequences into one sorted output sequence (called MERGER[n] in Cormen, Leiserson, Rivest, & Stein 2001). Specifically, $R_{(i)}$ and $T_{(i)}$ ($i = 1, \dots, n$) are the two sorted input sequences and $W_{(i)}$ with ($i = 1, \dots, 2n$) corresponds to the sorted output sequence. For details, see Cormen et al. (2001, pp. 716–718). \square

Notice, however, that the close relationship between the combined set \mathcal{W} and the antithetic variate method can be further generalized to all positive integer n (not necessarily even) as highlighted in the following remark.

Remark 8. The identity $F_R = A_{0.5}$ follows by Proposition 2 and 5. Moreover, at the sample level, $\hat{A}^{-1}(0.5) = \hat{\xi}_{0.5} = w_{(n)}$ and consequently $r_{(i)} = w_{(i)}$ for all $i = 1, 2, \dots, n$. A similar construction can be verified for the remaining half sample $w_{(n+i)}$ ($i = 1, 2, \dots, n$) of \mathbf{w} with respect to the corresponding antithetic variate quantities $t_{(i)}$. Note that this line of reasoning simply requires that n be a natural number (either even or odd).

To summarize, the most relevant result of this section regards the characterization of the reconstructed RST cumulative distribution function K as a truncated distribution under the point null

hypothesis (6). Moreover, Proposition 7 shows that computing an antithetic variate transformation simply coincides with directly selecting the first n order statistics from the combined set \mathscr{W} . Overall, this means that in the estimation of the cdf K the adoption of (i) the procedures described in Eqs. (11-13), (ii) the ecdf computed on the first n ranked elements of the combined data set \mathbf{w} , and finally (iii) the ecdf computed on the output of the antithetic variates approach (14), constitute all legitimate and equivalent estimators for K . In the next section, we will use these specific features of the reconstructed cdf K to suggest a statistical test procedure that accounts for its truncation property in comparing the empirical distribution functions in the RST and SST conditions.

2.3 The test procedure

The null hypothesis (5) entails that, under the RMI, the cdf F_Z of the RST condition never dominates the reconstructed cdf K . As a consequence, under the null hypothesis we generally expect that $\widehat{F}_{Z,m}(t)$ should have smaller or equal values than $\widehat{K}(t)$ for all values $t \geq 0$. In line with Maris and Maris (2003), we opted for a signed Kolmogorov-Smirnov (KS) test approach to evaluate this condition. In general, the KS test can be used to compute a distance between the two ecdfs as follows:

$$\begin{aligned}
 d^+ &= D^+(\mathbf{z}, \mathbf{w}) \\
 &= \max_t [\widehat{F}_{Z,m}(t) - \widehat{K}(t)] \\
 &= \max_t [\widehat{F}_{Z,m}(t) - \min\{1, 2\widehat{A}(t)\}] \\
 &= \max_t [\widehat{F}_{Z,m}(t) - \widehat{A}_{0.5}(t)]
 \end{aligned} \tag{16}$$

However, because under the maximal statistical facilitation assumption the reconstructed distribution K is always characterized by a truncated feature, we need to modify the KS test to account for this specific distributional feature (Tsao, 1954; Conover, 1967; Barr and Davidson, 1973; Koziol and Byar, 1975; Grover, 1977; Dufour and Maag, 1978; Michael and Schucany, 1979). In particular, a simple adaptation of the KS test which accounts for truncated distributions is as follows:

$$d_{tru}^+ = \max_{t \leq t^*} [\widehat{F}_{Z,m}(t) - \widehat{K}(t)] \tag{17}$$

where $t^* = \min\{w_{(n)}, z_{(m)}\}$ is the common threshold value for the two ecdfs and $z_{(m)}$ is the maximum value for \mathbf{z} . Moreover, because under the point null hypothesis (6) the ecdf $\widehat{F}_{Z,m}$ tends to be largest

everywhere, we also expect that under this condition also d_{tru}^+ should reflect largest deviation values. Note that, Eq. (17) highlights another important difference with the mixture representation by Maris and Maris (2003). In particular, in their proposal the statistical test did not require any kind of truncation correction as evident from their Kolmogorov-Smirnov distance equation:

$$d_{mix}^+ = \max_t [\widehat{F}_{Z,m}(t) - \min\{1, 2\widehat{M}(t)\}].$$

Moreover, in the computation of the distance measure the ecdf of the mixture distribution M was adopted in place of the ecdf of the average distribution A .

Since the ordinary KS statistic, as well as its truncated (resp. mixture) version are invariant under the probability integral transformation of the underlying data, there is no loss of generality in limiting the discussion of these tests to the case of the uniform transformations (Angus, 1994; Barr & Davidson, 1973). In particular, in the truncated version of the KS statistic, the sample distribution functions can be viewed as random walks from the origin $(0, 0)$ to the truncation point t^* once it has been rescaled under the uniform distribution. Finally, we can easily approximate the probability distribution of D_{tru}^+ under the point null hypothesis by adopting the following simple simulation procedure.

2.3.1 Generating the null distribution

The main idea consists in generating the transformed sample arrays for the SST and RST conditions under the point null hypothesis (6) on the basis of the simple algorithm described in Table 2. The data generation algorithm is then used to approximate the probability distribution of D_{tru}^+ under the point null hypothesis to any degree of precision. In particular, to generate the point null distribution for the truncated version of the KS statistic we revolve in setting the following three steps:

1. Simulate a large number N (e.g., $N = 10,000$) of transformed uniform samples, \mathbf{u}^* and \mathbf{v}^* , for the SST and RST conditions as detailed in Table 2.
2. For each pair of transformed uniform samples $(\mathbf{u}^*, \mathbf{v}^*)$ compute \widehat{K}_n on \mathbf{u}^* and $\widehat{F}_{Z,m}$ on \mathbf{v}^* . Finally, compute the simulated truncated KS test c_{tru}^+ according to (17) with $t^* = \min\{u_{(n)}, v_{(m)}\}$.
3. Derive an approximated p -value by using the proportion of generated N values, c_{tru}^+ , that exceeds the empirical sample value d_{tru}^+ .

Note that, under the probability integral transformation the derivation of the distribution function F_{T^*} for the common threshold $T^* = \min\{U_{(n)}, V_{(m)}\}$ is straightforward (see also Table 2). In the following proposition we will derive the analytic representation for the distribution of T^* .

Proposition 9. Under the point null hypothesis, the distribution function F_{T^*} for the common threshold variable T^* , computed on arrays of independent and uniformly distributed r.v.s (U_1, \dots, U_{2n}) and (V_1, \dots, V_{2m}) , takes the following form:

$$F_{T^*}(x) = I_x(n, n+1) + I_x(m, m+1) - I_x(n, n+1)I_x(m, m+1), \quad (18)$$

with $0 < x < 1$ and where $I_x(a, a+1)$ is the regularized beta function

$$I_x(a, a+1) = \frac{1}{B(a, a+1)} \int_0^x c^{(a-1)}(1-c)^a dc, \quad (19)$$

with B being the ordinary beta function and $a > 0$.

Proof. For the basic theory on order statistics (e.g., Arnold, Balakrishnan, Nagaraja, 2008) we have that the distribution function for T^* is given by

$$\begin{aligned} F_{T^*}(x) &= P(T^* \leq x) \\ &= 1 - P(T^* > x) \\ &= 1 - [1 - F_{U_{(n)}}(x)][1 - F_{V_{(m)}}(x)] \end{aligned}$$

with $F_{U_{(n)}}$ and $F_{V_{(m)}}$ being the distribution functions for the order statistics $U_{(n)}$ and $V_{(m)}$, respectively. Moreover, for independent and uniformly distributed random variables it is known that

$$F_{U_{(n)}}(x) = I_x(n, n+1) \quad (20)$$

$$F_{V_{(m)}}(x) = I_x(m, m+1) \quad (21)$$

and therefore

$$\begin{aligned} F_{T^*}(x) &= 1 - [1 - I_x(n, n+1)][1 - I_x(m, m+1)] \\ &= I_x(n, n+1) + I_x(m, m+1) - I_x(n, n+1)I_x(m, m+1), \end{aligned}$$

by simple algebra. \square

Note that the expected values and variances for the order statistics $U_{(n)}$ and $V_{(m)}$ are, respectively,

$$\begin{aligned} E(U_{(n)}) &= \frac{n}{2n+1} & \text{Var}(U_{(n)}) &= \frac{n^2+n}{(2n+1)^2(2n+2)}, \\ E(V_{(m)}) &= \frac{m}{2m+1} & \text{Var}(V_{(m)}) &= \frac{m^2+m}{(2m+1)^2(2m+2)}, \end{aligned}$$

which make $U_{(n)}$ and $V_{(m)}$ consistent but biased estimators of the uniform population quantile $\xi_{0.5}$ (= 0.5). The latter entails that also T^* is a consistent but biased estimator of $\xi_{0.5}$, where in particular it underestimates the population median 0.5 of the uniform distribution. However, it is worthwhile noticing that this biased feature is typical for any random sample with an even number of observations and it does not affect the overall functioning of the truncated statistic. More information about the basic properties of the truncated KS statistic can be found in the nonparametric statistics literature (e.g. Tsao, 1954; Conover, 1967; Barr and Davidson, 1973; Koziol and Byar, 1975; Grover, 1977; Dufour and Maag, 1978; Michael and Schucany, 1979).

2.3.2 R script functions

The testing procedure based on the truncated Kolmogorov-Smirnov test is implemented using a set of R script functions which only require the base R package (R Core Team, 2017) to be installed. The R script functions work with raw observed reaction times for the three samples of observations, \mathbf{x} , \mathbf{y} , and \mathbf{z} , and can be collected with any recording software capable to export data in .csv format. The R scripts are freely available at [TO ADD URL] together with an associated documentation illustrating the application of the R scripts via simple examples of data analysis.

3 Simulation studies

In this section we will test the behavior of the proposed algorithm by means of two simulation studies and compare its results with the original nonparametric test introduced by Maris and Maris (2003). The first simulation will investigate the type I error rate as a function of factors such as sample size and distributional features. The second simulation will be focused on some statistical power aspects of the truncated Kolmogorov-Smirnov test.

3.1 Type I Error Rate

To evaluate the performance of the new statistical procedure, we ran a Monte Carlo (MC) simulation study to investigate the Type I Error rate. The analysis was conducted within the *point null hypothesis* framework: $H_0 : F_Z(t) = K(t), \quad \forall t \geq 0$. In this context, both $\widehat{F}_{Z,m}$ and \widehat{K}_n were constructed in order to correspond to the maximal statistical facilitation attainable for a race model. Both single target conditions and redundant target condition were modeled according to a Weibull distribution (Van Zandt, 2000, Palmer, Horowitz, Torralba, & Wolfe, 2011). We recall that the Weibull distribution is characterized by two parameters: scale (λ) and shape (k). In this and in the following simulation study, only non-scaled observations were considered (**as the Maris and Maris's KS test as well as the truncated KS test are known to be scale invariant**). The simulation was performed by systematically varying two factors in a complete 4×5 two-factorial design (see also Rach, Diederich, Steenken, and Colonius, 2010). In particular, the factors were:

- (i) the *distance rate* (δ) at four levels: 0, 0.1, 0.2, 0.3. Factor δ regulates the relation between the distributions of the two channels (X and Y) for the two single target conditions. This relation is modeled by increasing an initial parameters array $\Theta_0 = (\lambda_0, k_0)$ according to $\Theta_i = \Theta_0 + \delta_i \Theta_0$, $i = 1, 2, 3, 4$. The single target conditions are modeled by keeping fixed Θ_0 for X , and by letting Θ_i to vary along the levels of the factor, for Y . Initial parameter values were fixed at $\lambda_0 = 1.5$ and $k_0 = 100$ (Figure 1);
- (ii) the *sample size* (s) at five levels: 20, 50, 100, 250, 500.

For each of the four distributional patterns (conveyed by the levels of δ), 10,000 simulations were performed for each level of s (here we assume that $m = n = s$). On the basis of these simulations, we computed the proportion of rejections of the null hypothesis for all the combinations of the factor levels. Table 3 reports the results of this first simulation study, that is, the Type I Error rate as a function of distance rate and sample size **for the two statistical procedures**. To summarize, under the framework of the point null hypothesis no particular effects of the factors on the Type I Error rate were observed. Overall, for the truncated KS test the error rate approximately settled on the expected value (5%), whereas the Maris and Maris KS test showed sub-optimal performances. **In particular, for the mixture KS test the H0 rejection rates were above 10 percent in many simulated conditions.**

3.2 Power Analysis

In this section, we present a more realistic scenario in which the (non-scaled) RT distribution F_Z for the redundant target condition is modeled in a more natural and plausible way. In this second simulation study, we considered several redundant target conditions in which the distributional patterns were not constructed to reflect a maximal statistical facilitation (and therefore a truncated distribution for F_Z). Also in this second simulation, the observed responses were modeled according to a Weibull distribution and assumed equal sample size ($n = m = s$) for F_Z and K (Rach, Diederich, Steenken, and Colonius, 2010). In the simulation design we varied three factors in a complete $4 \times 5 \times 10$ factorial design. The factors were:

- (i) the *distance rate* (δ) at four levels: 0, 0.1, 0.2, 0.3;
- (ii) the *sample size* (s) at five levels: 20, 50, 100, 250, 500;
- (iii) the *shape scaling factor* (η) at ten levels: 0.0159, 0.0169, 0.0179, 0.0189, 0.0199, 0.0209, 0.0219, 0.0229, 0.0239, 0.0249. Factor η regulates the shape parameter (k) of the redundant target condition distributions according to $k_r = \frac{1}{\eta_r}$ for $r = 1, 2, \dots, 10$.

The aim of this new simulation study was to analyze the proportion of rejection of H_0 under different conditions, that is, different distributional patterns of single and redundant target conditions in which specific violations may occur. In particular, given δ , four overall patterns of RMI violation magnitude blocks were generated, one for each level of δ . For each δ -block, the reconstructed RST distribution K was generated according to the maximal statistical facilitation hypothesis for a race model. By contrast, F_Z was generated as the cdf of Z with $Z \sim Weibull(\lambda_i, k_r = \frac{1}{\eta_r})$ where $\lambda_i = \lambda_0 + \delta_i \lambda_0$ and $\lambda_0 = 1.5$. These redundant target condition distributions are meant to elicit a violation of the race model inequality to varying magnitudes. The higher the value of η_r , the greater the magnitude of the violation.

Four power analyses were conducted, one for each of the four blocks of violation magnitude patterns. A total of 10,000 simulations were performed for each combination of levels of δ , s , and η^5 .

⁵We rescaled η_r levels, after the end of each δ block. The rescaling was obtained by setting $\eta_r = \eta_{r-1} - 0.0008$. This ensured that the magnitudes of violations of the race model inequality, at a population level, settled on similar values for each distributional pattern across the δ_i levels.

The dependent variable was the proportion of rejection of the null hypothesis ($H_0 : F_Z \leq K$) for the two statistical tests.

For the sake of space we limit here the presentation of our results to the $\delta = 0.2$ condition only⁶.

Table 4 reports the results concerning the four blocks of violation magnitude patterns as a function of the sample size and *median gain* MG (resp., *maximum theoretical distance* MTD) for $\delta = 0.2$. The median gain can be computed as the difference between the median of K and the median of F_Z for each separate block. As expected, negative values for the median gain were more associated with larger approximated p -value (Figure 2). By contrast, positive median gains are associated with a violation of the race model inequality (Figure 2).

However, the median gain only provides an horizontal measure of distance and does not provide an effective measure of theoretical violation which is, instead, yielded by the maximum theoretical distance (MTD) between the two distributions. Both median gain and maximum theoretical distance were computed in order to compare the rejection rates according to readable indices of violation. The results showed that, indeed, by increasing the positive median gain (resp. maximum theoretical violation), the proportion of rejections for H_0 also increased in both statistical tests (truncated KS test and mixture KS test). Finally, as expected, rejection rates were also clearly affected by the sample size factor with larger samples being associated with larger probabilities of rejecting the null model when H_0 was false (Figure 3). However, one clear difference emerged between the two statistical procedures. The truncated KS test was more conservative for small samples but resulted more powerful with medium or large samples as compared to the mixture KS test. This is not a surprise as in the previous simulation study, the mixture KS test failed in controlling the type I error rate. The same results were observed for the remaining conditions of the simulation study which are reported in the supplementary material.

Taken together the two simulations studies shows that the truncated KS test efficiently controls for type I error and it is also characterized by a reasonable power. By contrast, the statistical test based on the mixture representation (Maris and Maris, 2003) appears somehow flawed. In particular, it does not seem to adequately control for the rejection rate under H_0 . It also shows (at some levels) less power as compared to our proposal, specifically in the asymptotic conditions where large sample

⁶The complete set of results for the two statistical tests are reported in the supplementary material.

sizes are considered. At the same time, the mixture KS test seems to be too sensitive in small samples scenarios which, in turn, should naturally be considered less reliable. In the supplementary material, we finally reported the results of an additional simulation study based on the well known Diffusion Superposition model for reaction times (Schwarz, 1994) which generally confirmed the results of the power simulation study described in this section.

4 Discussion and conclusions

In this paper we first studied the truncated property of the reconstructed distribution K for the redundant condition and, subsequently, proposed a new nonparametric procedure (based on a modified version of the KS test) to evaluate violation of the race model inequality in the single-participant analysis context. To our knowledge, this is the first time that this property has been explicitly studied in the RMI framework. In particular, we showed that under an equal sample size assumption for the two SST conditions, estimation of the reconstructed distribution K can be obtained using a variety of perfectly equivalent procedures (e.g., antithetic variates method, ecdfs transformations, truncated sample derivation) which all yield the same consistent estimator for K . In addition, by using the averaged representation for the combined set \mathcal{W} , we were also able to propose a new, simple simulation algorithm (based on the probability integral transformation) to approximate, to any level of desired precision, the distribution of the test statistic under the null hypothesis. The corresponding sampling procedure nicely overcame some inherent limitations of applying the mixture representation as originally proposed by Maris & Maris (2003). Indeed, in experimental contexts participants are typically presented a fixed number of trials from each of the two SST conditions, instead of a random assignment as required by the mixture sampling procedure.

Regarding the performance of the new statistical test, the results of two Monte Carlo simulation studies suggested that our procedure efficiently controlled for type I error when a point null hypothesis $F_Z = K$ was true. By contrast, the mixture KS-test failed in controlling the type I error as it showed a larger proportion of H_0 rejections (> 0.05). As for power analysis, the truncated version of the KS test was in general able to reject the null hypothesis when a violation of the race model inequality occurred at the population level. As expected, the power of the statistic was positively related with sample size. Overall, the results of these preliminary simulation studies seem to support the validity

of the new statistical proposal.

To conclude this section, we now discuss some potential concerns associated with our statistical framework. First, we consider the consequences of relaxing the equal sample size assumption for the two SST conditions in testing the RMI. In some circumstances, one may face the problem of unbalanced data for the two SST conditions due, for example, to missing observations caused by omitted responses, errors, or outliers. Alternatively, in some experiments we may observe unbalanced data because of specific experimental designs or requirements. In all these situations, a tempting solution could be estimating K according to:

$$\widehat{K}(t) = \min\{1, \widehat{F}_{X,n_1}(t) + \widehat{F}_{Y,n_2}(t)\}, \quad t \geq 0 \quad (22)$$

with $n_1 \neq n_2$. However, although the two ecdfs \widehat{F}_{X,n_1} and \widehat{F}_{Y,n_2} are consistent and unbiased estimators of F_X and F_Y , the reconstructed distribution \widehat{K} , as defined in (18), generally is no more an ecdf. In particular, the resulting $\widehat{K}(t)$ would not necessarily be a step function with equal fixed jumps which, in turn, represents a mandatory property for any well-formed ecdf⁷. Note that this problem disappears when sample sizes are equal. For $n_1 = n_2 = n$, the reconstructed \widehat{K} is always an ecdf (i.e., a step function that jumps up by $1/n$ at each of the n data points), and hence is also a consistent estimator of K . Similarly, it would not be clear how to estimate K using directly the combined sample \mathbf{w} or the antithetic variates method. In general, for unbalanced data a possible way out could be the adoption of a resampling procedure with which we repeatedly sampling a subset of $n^\circ = \min\{n_1, n_2\}$ observations from the largest SST sample. For each of these subsets, one can rerun the analysis (on the new balanced data set) to obtain a distribution of approximated p -values on which to compute some desired summary statistics. However, a first drawback of this resampling approach would be a potential loss of power due to the resulting data reduction in the largest sample. Another limitation pertains the possibility that *a posteriori* data manipulations might raise the chance of introducing some undesired bias in the analysis (e.g., Eriksen, 1988; Gondan and Minakata, 2016).

Second, we may also have methodological limitations associated with the equal sample size $n_1 = n_2$ assumption. In some contexts, setting equal trial numbers in the different experimental conditions could potentially impact the way participants perform the task and eventually lead them to use expectations to anticipate the upcoming stimulus. For example, if one of the two stimulus has

⁷We recall that an ecdf is a step function that jumps up by $1/n$ at each of the n observed data points.

been observed more frequently (as compared to the other), then the participant could be more biased toward the less frequent stimulus and eventually anticipate its associated response. However, this drawback can be easily minimized by running experiments with a sufficiently large number of trials in each of the experimental conditions.

Third, we consider the case of observed data with ties. Even if nowadays, with the availability of high resolution RT clocks, ties are barely present, it would be still desirable to have a statistical procedure for dealing with them. In our framework, a possible solution could consist in first transforming the full data set into ranks and next derive all possible total linear orders (*tlo*) by reordering the ties in the equivalent classes. So, for example, if we denote by n_b the number of ties in cluster b , with $b = 1, \dots, k \ll m + 2n$, then the total number of reconstructed *tlos* is given by

$$e = \prod_{b=1}^k n_b!$$

For each of the e *tlos* we can reconstruct a new dataset with no ties on which we can finally test the RMI. The average p -value computed across all the e distinct p -value would then be considered as a measure of violation of the race model. Of course, the former combinatorial procedure would be meaningful only when applied to samples with a small proportion of ties (i.e., less than 3 – 4%).

Finally, fourth, we stress the potential limitation of assuming independent and identically distributed observations (resp. independent and non-identically distributed observations) in the samples. As one of the reviewer noted, this assumption simply ignores the eventual presence and impact of known phenomena in RT experiments such as, for example, sequence effects and modality shift costs (e.g., Spence, Nicholls, and Driver, 2001). Under some conditions, these uncontrolled effects could cause our statistical test (as well as all the statistical procedures based on this assumption) to become anticonservative (Gondan and Minakata, 2016). However, we notice that the removal of the independence assumption in RMI data would be problematic even for the basic definition of context invariance itself.

References

Angus, J. E. (1994). The probabiliry integral transform and related result. *SIAM review*, **36**, 652-654

- Arnold, B. C. , Balakrishnan, N., & Nagaraja, H. N. (2008) *A first course in order statistics*. SIAM, Philadelphia, PA.
- Barr, D. R. & Davidson, T. (1972). A Kolmogorov-Smirnov test for censored samples. *Technometrics*, **15**, 739-757
- Colonus, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology*, **34**, 253-275
- Colonus, H. (2016). An invitation to coupling and copulas: With applications to multisensory modeling. *Journal of Mathematical Psychology*, **74**, 2-10
- Colonus, H. & Diederich, A. (2006). The Race Model Inequality: interpreting a geometric measure of the amount of violation. *Psychological Review*, **113**, 148-154
- Cormen, T. H. , Leiserson, C. E., Rivest, R. L., & Stein, C. (2001) *Introduction to Algorithms, Second Edition*. The MIT Press, Cambridge, MA
- Conover, W. J. (1967). The distribution functions of Tsao's truncated Smirnov statistics. *The Annals of Mathematical Statistics*, **38**, 1208-1215
- Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, **41**, 260-274
- Dufour, R. & Maag, U. R. (1978). Distribution results for modified Kolmogorov-Smirnov statistics for truncated or censored samples. *Technometrics*, **20**, 29-32
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, **7**, 1-26
- Eriksen, C. W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception & Psychophysics*, **44**, 191-193
- Gondan, M. (2010). A permutation test for the race model inequality. *Behavior Research Methods*, **42**, 23-28
- Gondan, M. & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, & Psychophysics*, **78**, 723-735

- Grover, N. B. (1977). Two-sample Kolmogorov-Smirnov Test for truncated data. *Computer Programs in Biomedicine*, **7**, 274-250
- Koziol, J. A. & Byar, D. P. (1975) - Percentage points of the asymptotic distributions of one and two sample K-S statistics for truncated or censored data. *Technometrics*, **17**, 507-510
- Luce, R.D. (1986) *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford University Press, New York.
- Maris, G. & Maris, E. (2003). Testing the Race Model Inequality: A Nonparametric Approach. *Journal of Mathematical Psychology*, **47**, 507-514
- Michael, J. R. & Schucany, W. R. (1979). A new approach to testing goodness of fit for censored samples. *Technometrics*, **21**, 435-441
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, **14**, 247-279
- Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & psychophysics*, **40**, 331-343
- Palmer, E. M. , Horowitz, T., Torralba, A. & Wolfe, A. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance*, **37**, 58-71
- Pratt, J. W. & Gibbons, J. D. (1981). *Concepts of nonparametric theory*. Springer Verlag, New York
- R Core Team (2018). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL <https://www.R-project.org/>.
- Raab, D. (1962). Statistical facilitation of simple reaction time. *Transactions of the New York Academy of Sciences*, **43**, 574-590
- Rach, S. , Diederich, A., Steenken, R., & Colonius, H. (2010). The race model inequality for censored reaction time distributions. *Attention, Perception, & Psychophysics*, **72**, 839-847
- Serfling, R. J. (1980). Approximation theorems of mathematical statistics. *New York: Wiley*

- Schwarz, W. (1994). Diffusion, superposition, and the redundant targets effect. *Journal of Mathematical Psychology*, **38**, 504-520
- Spence, C. , Nicholls, M.E.R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, **63**, 330-336
- Thompson, J. R. (2000). *Simulation: a modeler's approach*. John Wiley and Sons, Inc., New York
- Todd, J. W. (1912). Reaction to multiple stimuli. *New York: Science Press*
- Tsao, C. K. (1954). An extension of Massey's distribution of the maximum deviation between two sample cumulative step functions. *The Annals of Mathematical Statistics*, **25**, 587-592
- Ulrich, R. , Miller, J., & Schröter, H. (2007). Testing the race model inequality: An algorithm and computer programs. *Behavior Research Methods*, **39**, 291-302
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, **7**, 424-465
- Van Zuijlen, M. C. A. (1976). Some properties of the empirical distribution function in the non-i.i.d. case. *The Annals of Statistics*, **4**, 406-408
- Van Zuijlen, M. C. A. (1978). Properties of the empirical distribution function for independent non-identically distributed random variables. *The Annals of Probability*, **6**, 250-266
- Vorberg, D. (2008, September). Exact statistical tests of the race model and related distribution inequalities. *Paper presented at the meeting of the European Mathematical Psychology Group, Graz, Austria*

Tables and figures

rank	\mathbf{w}^*	\mathbf{x}^*	\mathbf{y}^*	\hat{F}_5	\hat{G}_5	\hat{A}	$\hat{A}_{0.5}$	\hat{K}
1	0.165	-	0.165	0	1/5	1/10	1/5	1/5
2	0.346	0.346	-	1/5	1/5	2/10	2/5	2/5
3	0.395	-	0.395	1/5	2/5	3/10	3/5	3/5
4	0.470	-	0.470	1/5	3/5	4/10	4/5	4/5
5	0.501	0.501	-	2/5	3/5	5/10	1	1
6	0.577	0.577	-	3/5	3/5	6/10	1	1
7	0.673	0.673	-	4/5	3/5	7/10	1	1
8	0.747	-	0.747	4/5	4/5	8/10	1	1
9	0.753	0.753	-	1	4/5	9/10	1	1
10	0.783	-	0.783	1	1	1	1	1

Table 1: Hypothetical data based on a uniform model for F and G . The superscript * denotes that the samples are sorted according to an increasing order. Note that $\hat{\xi}_{0.5} = 0.501$ which is also the 5-th order statistic $w_{(5)}$ for the combined sample \mathbf{w} .

<i>Condition</i>	<i>Uniformly transformed data</i>	<i>Step description</i>
SST	$\mathbf{u} : u_i \sim U(0, 1), \quad i = 1, 2, \dots, 2n$	Data generation ($2n$ i.i.d. observations)
	$\mathbf{u}^* = (u_{(1)}, u_{(2)}, \dots, u_{(n)})$ sub-array of \mathbf{u}	Data selection (first n ranked obs.)
RST	$\mathbf{v} : v_j \sim U(0, 1), \quad j = 1, 2, \dots, 2m$	Data generation ($2m$ i.i.d. observations)
	$\mathbf{v}^* = (v_{(1)}, v_{(2)}, \dots, v_{(m)})$ sub-array of \mathbf{v}	Data selection (first m ranked obs.)

Table 2: Data generation algorithm. $U(0, 1)$ denotes the uniform distribution.

Table 3: Type I Error Rate (as percentage of rejections) for the truncated Kolmogorov-Smirnov test and the mixture Kolmogorov-Smirnov test (*in parenthesis*).

distance rate (δ)	sample size (s)				
	20	50	100	250	500
0	4.9 (15.7)	4.7 (11.4)	5.3 (13.9)	5.9 (12.6)	5.2 (13.6)
0.1	6.0 (15.3)	5.3 (10.9)	6.4 (14.1)	5.9 (12.4)	4.8 (13.7)
0.2	5.3 (15.1)	5.3 (11.5)	6.0 (10.0)	5.6 (12.7)	5.0 (13.4)
0.3	5.8 (15.4)	5.3 (11.2)	4.9 (14.2)	4.9 (12.6)	4.9 (13.1)

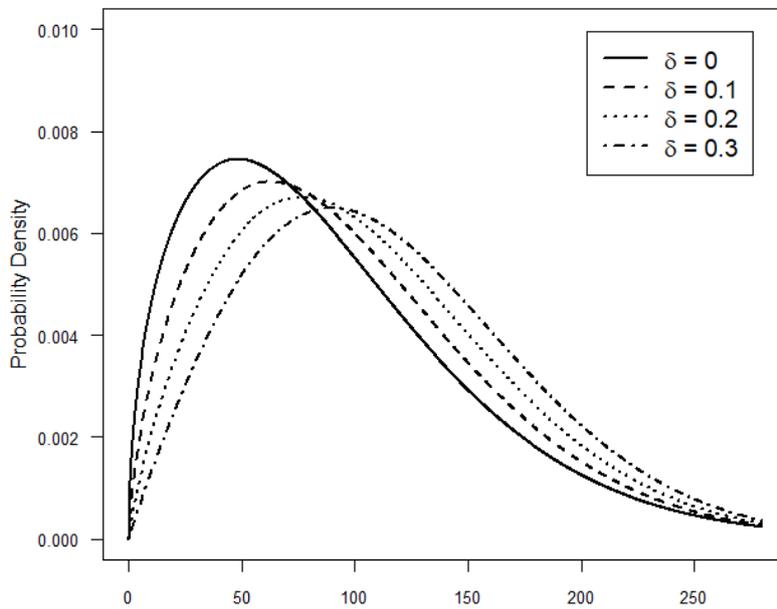


Figure 1: Distributions of the channels for single signal conditions when varying the levels of factor δ . Note that, when $\delta = 0$, the distribution of X and that of Y perfectly overlap.

Table 4: Power analysis (as percentage of rejections) for the truncated Kolmogorov-Smirnov test (*first line*), and the mixture Kolmogorov-Smirnov test (*second line*). MG: median gain; MTD: maximal theoretical distance.

sample size (s)	MG									
	-5.88	-2.24	1.12	3.92	6.44	8.97	10.93	12.89	14.57	16.25
20	1.2	2.0	3.2	5.3	7.4	12.2	16.5	21.1	27.0	32.3
	2.0	3.1	4.8	7.6	10.2	14.5	18.4	23.2	27.2	33.3
50	0.5	0.9	2.6	5.8	10.7	17.8	27.7	39.3	51.3	62.7
	0.7	1.9	4.3	7.3	13.0	19.1	26.5	35.1	44.6	52.7
100	0.1	0.7	2.0	6.5	15.6	29.8	47.8	65.4	79.0	88.7
	0.5	1.7	4.6	10.2	18.9	30.2	44.5	58.5	70.5	79.9
250	0.0	0.3	2.4	12.6	35.6	65.1	85.9	95.9	99.4	99.9
	0.0	0.1	4.4	14.6	31.8	56.2	75.8	90.2	96.1	98.4
500	0.0	0.1	3.6	24.6	65.3	93.0	99.3	100	100	100
	0.0	0.7	6.0	25.3	58.5	83.8	96.6	66.5	99.9	100
MTD	0	0.00	0.02	0.05	0.09	0.12	0.16	0.19	0.22	0.25

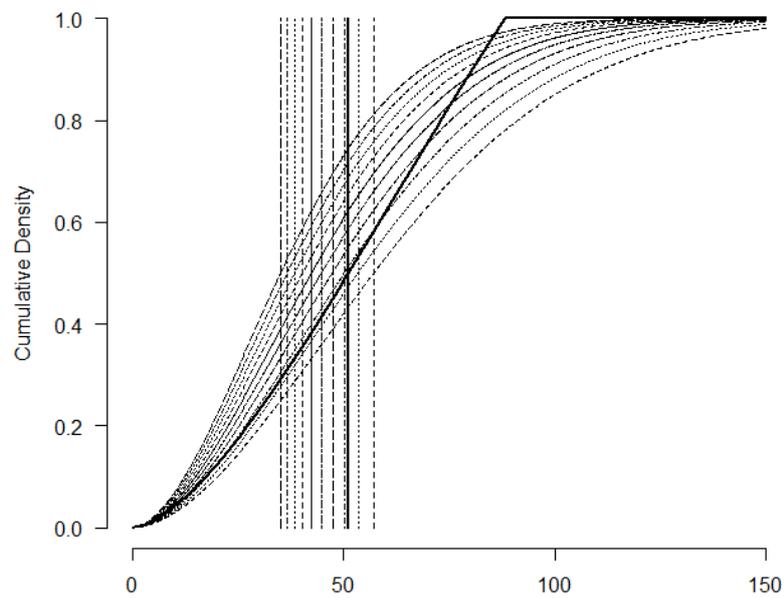


Figure 2: Distributional pattern block for the power analysis ($\delta = 0.2$). In the block, we generated 10 different distributions for H , one for each level of η . The truncated distribution (thick solid line) represents K . Dotted vertical lines represent the medians of the redundant target distributions H . The thick line represents the median of K .

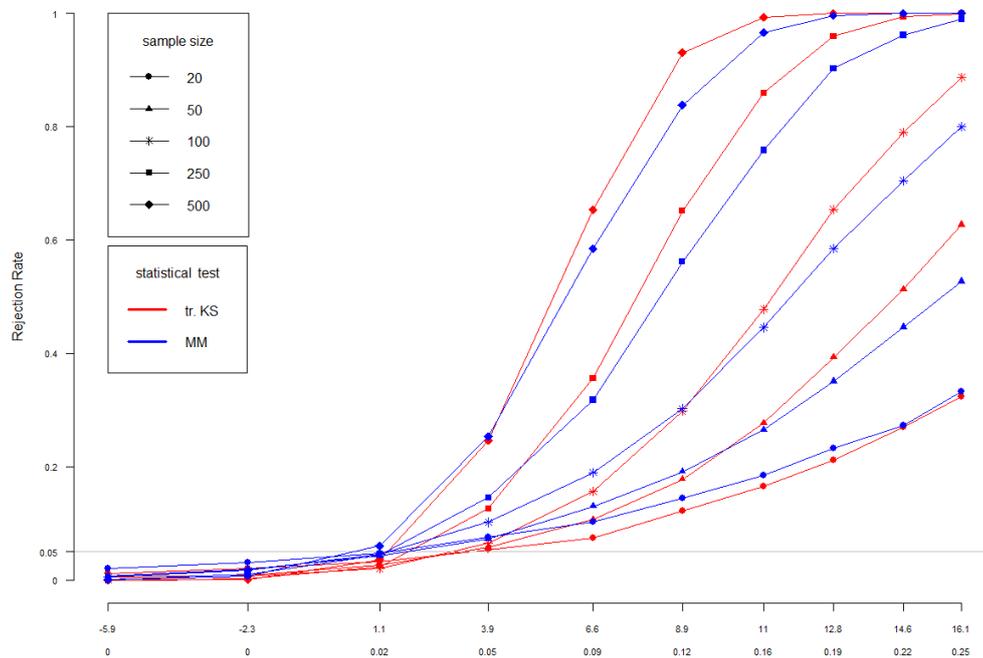


Figure 3: Power analysis as a function of sample size and distance information for block $\delta = 0.2$. The first row of the horizontal axis shows the median gain, whereas the second row shows the maximum theoretical distance. Black lines denote the power curves for the truncated KS test. Blue lines indicate the power curves for the mixture KS test.