

# A tutorial on testing the race model inequality

Matthias Gondan<sup>1</sup> · Katsumi Minakata<sup>1</sup>

Published online: 4 December 2015  
© The Psychonomic Society, Inc. 2015

**Abstract** When participants respond in the same way to stimuli of two categories, responses are often observed to be faster when both stimuli are presented together (redundant signals) relative to the response time obtained when they are presented separately. This effect is known as the redundant signals effect. Several models have been proposed to explain this effect, including race models and coactivation models of information processing. In race models, the two stimulus components are processed in separate channels, and the faster channel determines the processing time. This mechanism leads, on average, to faster responses to redundant signals. In contrast, coactivation models assume integrated processing of the combined stimuli. To distinguish between these two accounts, Miller (*Cognitive Psychology*, 14, 247–279, 1982) derived the well-known race model inequality, which has become a routine test for behavioral data in experiments with redundant signals. In this tutorial, we review the basic properties of redundant signals experiments and current statistical procedures used to test the race model inequality during the period between 2011 and 2014. We highlight and discuss several issues concerning study design and the test of the race model inequality, such as inappropriate control of Type I error, insufficient statistical power, wrong treatment of omitted responses or anticipations, and the interpretation of violations of the race model inequality. We make detailed recommendations on the

design of redundant signals experiments and on the statistical analysis of redundancy gains. We describe a number of coactivation models that may be considered when the race model has been shown to fail.

**Keywords** Attention: divided attention and inattention · Reaction time methods · Multisensory processing

## Redundant signals effect

Redundant signals are typically utilized to ensure that important, broadcast information is conveyed to an operator or user. For example, auditory-visual warning signals (i.e., flashing lights and sirens) are used in ambulances and police vehicles to make certain that drivers stop or yield as they head to their destinations. The use of multimodal redundant signals for alerting systems is imperative because they result in faster and more accurate responses than alerting systems without redundant coding. How does the human mind integrate the information from the different senses? By testing specific predictions of certain models one can infer the organization and characteristics of human information processing. A classical study of redundant signaling was performed by Hershenson (1962; the very first study with redundant signals has been reported by Todd, 1912, Section IV.2). Hershenson implemented a speeded response time (RT) task where observers had to respond by a keypress whenever they were presented with visual, auditory, or audio-visual stimuli. The redundant, audio-visual signals elicited the lowest mean RT relative to the single-signal conditions; this phenomenon is the “redundant signals effect.” Such redundant signals effects have been found across different response demands such as go/no-go and choice RT tasks (Grice & Canham, 1990; Grice,

---

**Electronic supplementary material** The online version of this article (doi:10.3758/s13414-015-1018-y) contains supplementary material, which is available to authorized users.

---

✉ Matthias Gondan  
Matthias.Gondan@psy.ku.dk

<sup>1</sup> Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 Copenhagen K, Denmark

Canham, & Boroughs, 1984; Grice, Canham, & Gwynne, 1984; Grice & Reed, 1992).

While Hershenson's results refer to simple responses to bimodal stimuli (i.e., of two sensory modalities), redundant signals effects have also been reported for stimuli presented within one sensory modality. If participants are instructed to respond whenever they detect a visual signal, they are faster if two visual signals are presented simultaneously compared to a single-signal presentation (e.g., Corballis, 2002; Miller & Adam, 2006). In the auditory modality, the pattern is more complex (e.g., no redundancy gains with pure tones of equal frequency; Schröter, Ulrich, & Miller, 2007).

Several models explain such redundancy gains. In the so-called race model (Miller, 1982) or parallel, first-terminating model (Townsend & Ashby, 1983), the redundancy gain is a consequence of "statistical facilitation" (e.g., Raab, 1962). Detection times for single target stimuli vary from trial to trial based on a statistical distribution. Whenever two target stimuli are presented, the stimulus that is faster processed triggers the response (i.e., this stimulus wins the "race"). As the processing time distributions for the two stimuli usually overlap, slow processing in one channel is compensated by faster processing in the other channel. Consequently, mean RTs for redundant stimuli are faster than mean RTs for single stimuli.

Instead of separate processing of the two signals, coactivation models assert some kind of integrated processing of the two stimuli. For example, Miller (1982, Appendix) proposed that the activations induced by both stimuli add up and the summed activation enables faster response initiation ("coactivation model"; see also Blurton, Greenlee, & Gondan, 2014; Diederich, 1995; Miller, 2004; Schwarz, 1989, 1994; Townsend & Nozawa, 1995). In order to distinguish between separate processing (race model) and integrated processing (coactivation model), Miller (1982) derived the well-known race model inequality (RMI), which has become a routine test for behavioral data in experiments with redundant signals. If a given data set violates the RMI, separate processing cannot explain the redundancy gain (for details, see below). Consequently, most researchers would then assume the alternative explanation for the given data set (i.e., integrated/"co-active" processing of the two stimuli).

This article begins with a derivation of the RMI. We then provide a short review of the methods used to test for a violation of the RMI and highlight some of the frequently encountered issues when testing the RMI. We close with a description of coactivation models and two generalizations. For simplicity, the example of redundant signals from the visual and auditory sensory modalities shall be used throughout this article with A, V, and AV denoting the auditory, visual, and audio-visual conditions, respectively. Of course, the derivations hold for any other combination of stimuli within or across modalities.

## Race model inequality

The race model is utilized to describe and predict how two or more stimuli may be processed in a redundant signals task. In a typical redundant signals task, three sets of RTs are involved; one for each single-signal stimulus and one for the redundant-signal stimulus. Miller's (1982) original race model assumes that the information that gets passed through these channels is processed separately. In other words, activation from one channel does not accumulate with the activation of the other channel. From this assumption (and another assumption, "context invariance"; see below), the RMI can be derived (Miller, 1982, Inequality 2) that describes the upper limit of the redundancy gain.

More formally, the race model assumes two channel-specific processing times  $\mathbf{D}_A$ ,  $\mathbf{D}_V$  in the bimodal context. The faster of the two channels is the winner of the race, so the smaller of the two processing times determines the processing time  $\mathbf{D}_{AV}$  for the redundant stimulus:

$$\mathbf{D}_{AV} = \min(\mathbf{D}_A, \mathbf{D}_V). \quad (1)$$

Take note that all of these times ( $\mathbf{D}_A$ ,  $\mathbf{D}_V$  and  $\mathbf{D}_{AV}$ ) refer to the processes within the context of a redundant, audio-visual signal. Most authors use the notation  $\{\mathbf{D} \leq t\}$  to describe the event of "fast processing" (i.e., a processing time below or equal to  $t$ ; e.g., Miller, 1982; Luce, 1986). As easily seen in Eq. 1, under the race model assumption,  $\mathbf{D}_{AV}$  is below  $t$  if either the processing time for the auditory component OR ( $\cup$ ) for the visual component is below  $t$ :

$$\begin{aligned} \{\mathbf{D}_{AV} \leq t \mid AV\} &= \{\min(\mathbf{D}_A, \mathbf{D}_V) \leq t \mid AV\} \\ &= \{\mathbf{D}_A \leq t \mid AV\} \cup \{\mathbf{D}_V \leq t \mid AV\}, \end{aligned} \quad (2)$$

with " $\mid AV$ " reiterating that in Eqs. 1 and 2,  $\mathbf{D}_A$ , and  $\mathbf{D}_V$  refer to the auditory and visual component processing times within the context of a redundant, audio-visual signal.

We now consider the probability  $P(\mathbf{D} \leq t)$  for fast processing. An upper bound for the probability of a union of two events as in Eq. 2 is given by Boole's inequality,  $P(X \cup Y) \leq P(X) + P(Y)$ , which states that the probability for the union of  $X$  and  $Y$  is always less than or equal to the sum of the probabilities for the individual events. Applying Boole's inequality to the union of the two events in Eq. 2 results in

$$P(\mathbf{D}_{AV} \leq t \mid AV) \leq P(\mathbf{D}_A \leq t \mid AV) + P(\mathbf{D}_V \leq t \mid AV), \quad (3)$$

for all  $t$ .

As already noted above, all the probabilities in Inequality 3 refer to the (latent) processing times within the context of a redundant signal. To relate the terms on the left and on the right side of Inequality 3 to observable RT distributions, the additional assumption of "context invariance" (or context independence; Luce, 1986, p. 129) is needed. Context invariance states that the processing times for the auditory

component of the redundant signal follow the same distribution as the processing times elicited from a single auditory signal. Likewise, the processing times for the visual component of AV follow the same distribution as the processing times elicited from a single visual signal:

$$P(\mathbf{D}_A \leq t | AV) = P(\mathbf{D}_A \leq t | A), \text{ for all } t, \tag{4A}$$

$$P(\mathbf{D}_V \leq t | AV) = P(\mathbf{D}_V \leq t | V), \text{ for all } t. \tag{4B}$$

Inserting Eqs. 4A and 4B into Inequality 3 yields

$$P(\mathbf{D}_{AV} \leq t | AV) \leq P(\mathbf{D}_A \leq t | A) + P(\mathbf{D}_V \leq t | V), \tag{5}$$

for all  $t$ .

Inequality 5 only describes the modality-specific processing times  $\mathbf{D}$  under the race model assumption. The observable response time  $\mathbf{T}$  also includes residual processes  $\mathbf{M}$ , for example, the finger movement, or, more generally, processes related to response execution, so that  $\mathbf{T} = \mathbf{D} + \mathbf{M}$ . Assuming that Inequality 5 holds for all  $\mathbf{M} = m$  (Dzhafarov, 2003; Dzhafarov, Schweickert, & Sung, 2004),

$$P(\mathbf{D}_{AV} \leq t | AV, \mathbf{M} = m) \leq P(\mathbf{D}_A \leq t | A, \mathbf{M} = m) + P(\mathbf{D}_V \leq t | V, \mathbf{M} = m), \text{ for all } t, m,$$

the inequality is directly translated to observable RT distributions,

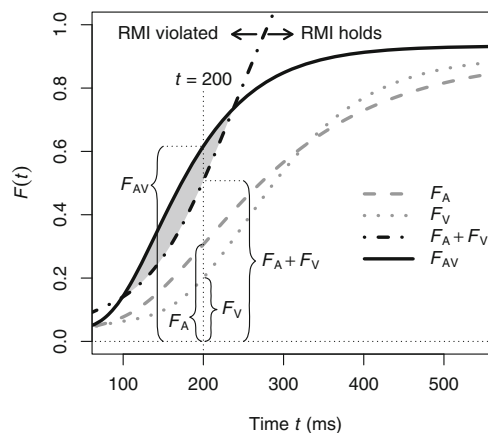
$$P(\mathbf{T}_{AV} \leq t) \leq P(\mathbf{T}_A \leq t) + P(\mathbf{T}_V \leq t), \text{ for all } t. \tag{6A}$$

The upper bound Inequality 6A is the well-known RMI (Miller, 1982; for a lower bound, see Grice, Canham & Gwynne, 1984). Inequality 6B is usually written compactly using the notation for the cumulative RT distribution function  $F(t) = P(\mathbf{T} \leq t)$ ,

$$F_{AV}(t) \leq F_A(t) + F_V(t), \text{ for all } t. \tag{6B}$$

Hence, under the race model, the cumulative RT distribution for AV is always either equal to or below (see Fig. 1) the summed distributions from the single auditory signal (A) and the single visual signal (V). Because cumulative distributions are monotonically increasing, an equivalent prediction is that the RT distribution for AV is equal to or to the right of the summed distributions for the single signals A and V. When the RMI is violated at some  $t$ , it is because the race model is the wrong model (Eq. 1), the context invariance assumption is untenable (Eqs. 4A and 4B), or both. In many studies context invariance is assumed to hold, and a violation of the RMI is used as a synonym for “coactivation.”

Note that Boole’s law (Inequality 3) does not require the channel processing times  $\mathbf{D}_A, \mathbf{D}_V$  to be stochastically independent. Therefore, the RMI is consistent with a scenario in which the participant concentrates on the auditory channel in a given trial ( $\mathbf{D}_A$  will be lower, at the expense of  $\mathbf{D}_V$ ), and on the visual channel in another trial. If the channel processing times were independent, the race model would predict an equality,



**Fig. 1** Cumulative distribution functions  $F(t)$  of RTs for auditory, visual, and auditory-visual stimuli, as well as the sum of the distributions for A and V ( $F_A + F_V$ ). According to the race model inequality,  $F_{AV}$  should always be below/to the right of  $F_A + F_V$ . In this example, the inequality is violated, for example, at  $t = 200$  ms, because  $F_{AV}(200)$  is above  $F_A(200) + F_V(200)$ . The violation occurs within the range illustrated by the shaded area. Two features of the curves should be noted: Because there might be a tendency to guess the onset of the stimulus especially in simple response tasks, a number of very fast “responses” might occur, so that  $F(t)$  is already above zero for  $t = 100$  ms. Similarly, the  $F(t)$  do not seem to tend to one at the upper limit of the RT window; this happens because some stimuli might be overlooked

$$F_{AV}(t) = F_A(t) + F_V(t) - F_A(t) \cdot F_V(t), \text{ for all } t. \tag{7}$$

We mention this here because Eq. 7 is sometimes used in studies with redundant signals (e.g., Stevenson et al., 2014; see also Current practice section below). We do not recommend using it because we think that the assumption of channel independence is poorly motivated.

The cumulative distribution functions in Inequalities 6A and 6B relate to theoretical probabilities that are not seen in an experiment. However, experimental data can be used to obtain empirical estimates  $\hat{G}(t)$  for these probabilities.  $\hat{G}(t)$  is given by the proportion of RTs below  $t$  in a specific experimental condition. These estimates can then be used to test whether the RMI holds or not in a given experimental task.

### Testing the RMI

In RT studies in experimental psychology hypothesis tests generally refer to differences in the mean RT observed under two or more experimental conditions. Most often, the “mean correct RT” is determined for the subset of correct responses that fall within some predefined response window, and a paired  $t$  test or an analysis of variance is used to test whether response speed in one condition differs from the other condition. The theoretical prediction usually corresponds to the alternative hypothesis, so that a significant difference in mean RTs is taken as support for the theory.

For the RMI (Inequality 6B), the situation differs from this standard scenario in several aspects: (A) The RMI does not refer to mean RT or mean correct RT within a specific time window, but to the RT distribution  $F(t)$ , that is, the probability that a response occurs within time  $t$  under the respective experimental condition. (B) The RMI refers to RTs for three experimental conditions instead of two, and the inequality is asymmetric, with one term on the left hand side, and two terms on the right hand side of the inequality. (C) The RMI is defined for all  $t$ , so violations may occur at any  $t$ . (D) The race model describes a mechanism, namely, separate, context-invariant processing of the redundant information provided by the two channels. Violation of the RMI rules out this mechanism but does not automatically support a specific alternative cognitive architecture. As a consequence of these four aspects, the test of the RMI is somewhat special—with special requirements for study design, preprocessing of the data, and statistical testing as well as the interpretation of the results. Conversely, blind application of standard methods can lead to biased results and wrong conclusions.

In this section we describe the most commonly used method that has been proposed for testing the RMI in a group of participants (Miller, 1982). For tests of single-participant data, see Miller (1986) and Maris and Maris (2003). We then focus on whether and how the method is affected by Issues A–D mentioned above and suggest ways to improve study design, analysis, and interpretation of the results.

The RMI can be tested by applying paired  $t$  tests to compare the empirical cumulative distribution functions  $\hat{G}(t)$  of RTs for A, V, and AV at prespecified time points or quantiles. While the RMI refers to the theoretical probabilities  $F(t)$ , the observed proportion of responses below  $t$  is denoted by  $\hat{G}(t)$ . The test involves three steps (Ulrich, Miller, & Schröter, 2007): (1)  $\hat{G}_A(t)$ ,  $\hat{G}_V(t)$  and  $\hat{G}_{AV}(t)$  are determined as the number of RTs below or equal to  $t$ , divided by the number of trials. (2) Then the sum of the distributions for auditory and visual stimuli is computed,  $\hat{G}_{A+V}(t) = \hat{G}_A(t) + \hat{G}_V(t)$ . (3) Each participant now provides a pair of proportions  $\hat{G}_{AV}(t)$  and  $\hat{G}_{A+V}(t)$  for a given  $t$ . For reasonably large numbers of trials, such proportions tend to Normal distributions, and so does their difference, therefore, the proportions can be compared using paired  $t$  tests. If  $\hat{G}_{AV}(t)$  is significantly *above*  $\hat{G}_{A+V}(t)$  at any  $t$ , the race model is rejected (“vertical” test; see Fig. 1 for an illustration at  $t = 200$  ms).

Most often, a slightly more complicated procedure (“horizontal test”) is chosen that involves one additional step. Remember that the RMI states that  $F_{AV}(t)$  is *below*  $F_A(t) + F_V(t)$ , for all  $t$ . However, because the  $F(t)$  are monotonically increasing, the RMI equivalently states that  $F_{AV}$  is to the *right* of  $F_A + F_V$  on its entire course (see Fig. 1 for an illustration). The latter statement refers to the inverse distributions (quantile functions).  $F$  and  $\hat{G}$  are functions of time that return probabilities/proportions, for example,  $F_V(200 \text{ ms}) \approx 20\%$  in Fig. 1.

The inverses  $F^{-1}$  and  $\hat{G}^{-1}$  are functions of probabilities/proportions that return times,  $F_V^{-1}(20\%) \approx 200$  ms. Therefore,  $\hat{G}_{AV}^{-1}(p)$  and  $\hat{G}_{A+V}^{-1}(p)$  are estimated for some prespecified set of  $p$ . Each participant then provides a pair of quantiles  $\hat{G}_{AV}^{-1}(p)$  and  $\hat{G}_{A+V}^{-1}(p)$  for a given  $p$ . If the sample size is large enough, the quantiles follow approximate Normal distributions (Bahadur, 1966), so that the pairs can be compared using paired  $t$  tests. If  $\hat{G}_{AV}^{-1}(p)$  is significantly to the left of  $\hat{G}_{A+V}^{-1}(p)$  at any  $p$ , the race model is rejected.

It is unclear whether the horizontal test or the vertical test is more powerful. Both variants can be adjusted to the response speed of the individual participants (see below), and other measures of violation have been suggested (e.g., Colonius & Diederich, 2006). Regardless of the specific procedure, the probabilities or time points at which the test is performed should be carefully selected. For example, Miller (1982) used a horizontal test with  $p = 5\%$ ,  $15\%$ ,  $25\%$ , and so forth.

## Analysis issues

In the following, we discuss the prerequisites for a proper test of the RMI, and we outline solutions for some typical problems in experimental research with redundant signals. As already mentioned above, the test of the RMI is special in the senses that (A) it refers to RT distributions instead of mean RTs, (B) it is asymmetric with only one term on the left but two terms on the right, and (C) it includes significance tests at multiple time points or quantiles. We now turn to these aspects in more detail.

**Estimation of distribution functions and quantiles** We first consider a scenario in which accuracy is 100%. When the time point  $t$  is held fixed at a certain value (e.g.,  $t = 200$  ms), the proportion  $\hat{G}(t)$  of RTs below or equal to  $t$  (i.e., the number of RTs  $\leq t$  divided by the number of trials  $n$  in the respective condition) is an unbiased estimate of the underlying probability  $F(t) = P(T \leq t)$ . It is, therefore, relatively easy to obtain the numerical estimates  $\hat{G}_A(t)$ ,  $\hat{G}_V(t)$ , and  $\hat{G}_{AV}(t)$  for the different  $t$ . For the horizontal test, it is slightly more difficult to obtain unbiased percentile estimates  $\hat{G}^{-1}(p)$ : basically, one needs to find the time  $t$  at which  $\hat{G}(t)$  crosses a certain level  $p$  (see Hyndman & Fan, 1996, for an overview of software solutions). Bias might occur in small trial numbers, but simulations have shown that the bias is minimal and manifests only if the sample is exceptionally small (Kiesel, Miller, & Ulrich, 2007).

**Errors and fast guesses** In many RT experiments, it is possible to anticipate the onset of the next upcoming stimulus to some extent. Especially in simple response tasks, it might happen that highly motivated participants press the response button before the stimulus actually appeared. In such a

situation, the RT is negative and, of course, such fast guesses are easily identified. However, for very fast responses of, say, 120 ms after stimulus onset, it is actually unclear whether they reflect informed responses or just fast guesses that happily coincided with the onset of the next stimulus. Eriksen (1988) noticed that due to the asymmetry of the RMI, fast guesses can bias the test of the RMI. Although he described a procedure for choice responses (see Inequality 10), his argument applies for both simple and choice response tasks: Assuming that fast guesses occur equally in all conditions (A, V, AV), they lead to an increase of all distribution functions involved (cf. Fig. 1). However, on the left-hand side of Inequality 6B, only  $F_{AV}(t)$  is increased. In contrast, on its right-hand side, both  $F_A(t)$  and  $F_V(t)$  are increased. Because the right-hand side is more affected by fast guesses than the left-hand side, it is more difficult to identify violations of the RMI. In other words, with fast guesses, the test of the RMI is biased, which might result in a considerable loss of statistical power (Miller & Lopes, 1991). Eriksen assumed a simple two-state model of fast guesses and informed responses (“guess” with probability  $g$ , “response” with probability  $1 - g$ ; Ollman, 1966). Based on this argument a refined test of the RMI can be constructed by inserting catch trials (C) in the stimulus sequence in which the RTs to no-stimulus events are recorded (these must be fast guesses). The onset of this event ( $t = 0$ ) is defined as for the other stimulus; it is the time point at which a stimulus would have been presented. The distribution of fast guesses  $F_G(t)$  equally contaminates the RT distributions for A, V, AV and C. Denote the contaminated distributions by  $F^*(t)$ .<sup>1</sup> Then,

$$\begin{aligned} F_A^*(t) &= g F_G(t) + (1 - g) F_A(t) \\ F_V^*(t) &= g F_G(t) + (1 - g) F_V(t) \\ F_{AV}^*(t) &= g F_G(t) + (1 - g) F_{AV}(t) \\ F_C^*(t) &= g F_G(t). \end{aligned}$$

Therefore, a refined test of the RMI is obtained if the RT distribution for catch trials  $F_C^*(t)$  is subtracted from the RT distributions for A, V and AV:

$$[F_{AV}^*(t) - F_C^*(t)] \leq [F_A^*(t) - F_C^*(t)] + [F_V^*(t) - F_C^*(t)],$$

for all  $t$ .

The above expression is easily seen to reduce to

$$F_{AV}^*(t) + F_C^*(t) \leq F_A^*(t) + F_V^*(t), \text{ for all } t. \tag{8}$$

Inequality 8 is the so-called kill-the-twin correction (see Eriksen, 1988, for more details). Gondan and Heckel (2008) showed that Inequality 8 is not restricted to a two-state model of guess and response but also follows from a more general “deadline model” (Yellott, 1971) in which guess and response tendencies coexist in parallel. Because of the symmetry of

Inequality 8, biases due to fast guesses cancel out, and the power to detect a violation of the RMI is increased (Miller & Lopes, 1991). For go/no-go and choice responses, the kill-the-twin correction looks slightly different (see Inequality 10).

**Omitted responses and outliers** In most RT experiments, there is a minimum and a maximum allowed RT for each stimulus (e.g., between  $t_1 = 100$  ms and  $t_2 = 1,000$  ms after stimulus onset). Responses outside of this window are considered anticipations and omitted responses, respectively, and are generally excluded from the RT analysis. For the test of the RMI, this procedure is wrong and results again in a loss of statistical power. The RMI refers to the probability that a response occurred within  $t$ , unconditionally. It does *not* refer to conditional probabilities  $F(t | t_1 \leq T_{AV} \leq t_2)$ ; and it is not possible to derive a prediction like

$$F_{AV}(t | 100 \leq T_{AV} \leq 1000) \leq F_A(t | 100 \leq T_A \leq 1000) + F_V(t | 100 \leq T_V \leq 1000)$$

from the race model assumption (see also Miller, 2004, Appendix A). We have dealt with anticipations in the previous paragraph and concluded that they should not be excluded from the data, but instead, the kill-the-twin correction should be applied. Similarly, slow responses and omitted responses should not be excluded from the analysis. Consider the following example data: three RTs below 500 ms, five RTs between 501 ms, and 1,000 ms, and two omitted responses. In 3 out of 10 trials, the participant responded within 500 ms, so  $\hat{G}(500 \text{ ms}) = 30\%$ . It would be wrong to exclude the omitted responses and estimate this probability to  $3/8 = 37.5\%$ .

By default, omitted responses indicate that the participant did not detect the stimulus within some time limit. Data cleaning should be limited to those trials of an experiment in which the participant was distracted (this may include trials with omitted responses as well as trials with valid RTs). When estimating  $P(T \leq t)$ , omitted responses should be assigned an infinitely long RT (or any value above the maximum of the RT window). As a consequence, the distribution functions  $F(t)$  and their estimates  $\hat{G}(t)$  do not necessarily tend to 0% and 100% at the boundaries of the RT window (see Fig. 1). This holds for  $\hat{G}_A(t)$ ,  $\hat{G}_V(t)$ ,  $\hat{G}_{AV}(t)$ , and, of course, for the distribution of spontaneous responses in catch trials  $\hat{G}_C(t)$ .

**A permutation test for multiple time points** The RMI is predicted to hold for all  $t$ , so it may be violated at any  $t$ . Therefore, multiple  $t$  tests are used to test the RMI at different time points or quantiles, and the race model is rejected if the test is significant for at least one of them. Such a multiple test procedure requires some control of the significance level. Kiesel et al. (2007) showed that running multiple  $t$  tests results in a serious accumulation of the Type I error. On the other hand, a

<sup>1</sup> Strictly speaking, only these contaminated RTs distributions are observable in an experiment. To avoid cumbersome notation, the stars are omitted in the sections that follow.

Bonferroni correction would be too conservative because the tests at the different time points are positively correlated.

Permutation tests provide an elegant way to control for the overall Type I error in multiple significance tests (e.g., Pesarin, 2001). Here, we describe a permutation test/randomization test for the RMI proposed by Gondan (2010). Suppose that data has been collected for  $N$  participants, and Inequality 8 is tested at  $t_0$  (e.g.,  $t_0 = 200$  ms); this will be generalized to multiple time points below. In the paired  $t$  test, the test statistic reads as

$$T = \bar{d}(t_0) / \text{se}[\bar{d}(t_0)], \quad (9)$$

with  $\bar{d}(t_0) = 1/N \times \sum_i d_i(t_0)$  and  $\text{se}[\bar{d}(t_0)]$  denoting the mean and the empirical standard error of the differences  $d_i(t_0) = \hat{G}_{AV,i}(t_0) + \hat{G}_{C,i}(t_0) - \hat{G}_{A,i}(t_0) - \hat{G}_{V,i}(t_0)$  for each participant  $i$ . At the boundary between the null hypothesis and the alternative hypothesis, the RMI turns into an equality,  $F_{AV}(t_0) + F_C(t_0) = F_A(t_0) + F_V(t_0)$ , so that the left side and the right side are *exchangeable*. In other words, under the null hypothesis, the sign of  $d_i(t_0)$  is random in each participant. This property is used in a permutation test to generate the distribution of the test statistic under the null hypothesis. In each simulation, the  $d_i(t_0)$  are randomly signed,  $d_i^*(t_0) = \pm d_i(t_0)$ , and the test statistic is calculated:  $T^* = \bar{d}^*(t_0) / \text{se}[\bar{d}^*(t_0)]$ . This procedure is repeated, for example, 10,001 times, resulting in a distribution of 10,001  $T^*$  values simulated under the assumption of the race model. Similar to a standard one-sided significance test, the race model is rejected at  $\alpha = 5\%$  if the observed  $T$  is greater than 95% of the simulated  $T^*$ .

Generalization of this permutation test to multiple time points is straightforward. We illustrate this for three time points,  $t_1$ ,  $t_2$ , and  $t_3$ , resulting in three differences  $d_i(t_1)$ ,  $d_i(t_2)$ ,  $d_i(t_3)$  for each participant  $i$ , as well as three test statistics  $T_1$ ,  $T_2$ ,  $T_3$ . In each simulation, a single multiplier,  $s_i^* = \pm 1$ , is chosen at random for each participant  $i$  and is multiplied with the difference values,  $d_i^*(t_1) = s_i^* \times d_i(t_1)$ ,  $d_i^*(t_2) = s_i^* \times d_i(t_2)$ , and  $d_i^*(t_3) = s_i^* \times d_i(t_3)$ . Separate  $T^*$  statistics are then calculated for the three time points, and the largest of the three  $T^*$  is chosen,  $T_{\max}^* = \max(T_1^*, T_2^*, T_3^*)$ . This is the so-called  $T_{\max}$  statistic. The race model is rejected at the one-sided significance level  $\alpha = 5\%$  if the maximum of the observed  $T$  values,  $T_{\max} = \max(T_1, T_2, T_3)$ , is greater than 95% of the simulated  $T_{\max}^*$  values.

This permutation procedure is known to control the Type I error in correlated significance tests. It has been repeatedly used for multivariate significance tests, for example, to test for differences in EEG waveforms (Blair & Karniski, 1993). Gondan (2010, Table 1) simulated different scenarios to investigate the power of the permutation test. Because the tests at the different time points are highly correlated, the better control of the Type I error does not come at the price of a substantially reduced statistical power.

The time points  $t_1$ ,  $t_2$ ,  $t_3$ , and so forth can be adjusted to the participant's individual response speed. To this end, the percentiles  $t_{1,i} = \hat{G}_i^{-1}(5\%)$ ,  $t_{2,i} = \hat{G}_i^{-1}(10\%)$ ,  $t_{3,i} = \hat{G}_i^{-1}(15\%)$ , and so forth, are determined separately for each participant  $i$ , using the RTs of all conditions mixed together. The permutation test is then performed using the differences  $d_i(t_{1,i})$ ,  $d_i(t_{2,i})$ ,  $d_i(t_{3,i})$ , obtained at these individually determined time points. We recommend to test the RMI in the lower percentile range. For large  $t$ , the left-hand side of the RMI tends to 1, whereas the right-hand side tends to 2, so the RMI cannot be violated for large  $t$ . The [Supplementary material](#) to this article includes scripts written in R (R Core Team, 2015) which perform these computations at quantiles 5%, 10%, 15%, . . . , 45%. The choice of the specific quantiles is somehow arbitrary; whichever quantiles are chosen, they should be specified a priori. If the quantiles are chosen after inspection of the data, the significance level will not be protected.

## Study design

Redundant signals experiments should be carefully designed to enable efficient analysis of the processing architecture. Of course, the experimental setup is determined by the primary research hypothesis. Here, we assume that the primary interest is whether the RMI holds or not, and we outline a minimalistic experimental setup with a simple response task that allows for a reliable and valid estimation of the respective RT distributions.

**Random presentation** The stimuli of the different modalities should always be presented in random order; they should never be presented in separate blocks unless there are really good reasons for it. In nontechnical terms, the context invariance assumption (Eq. 4A) states that the participants “do the same” in all conditions. When stimuli of different modalities are presented in separate blocks, the participants are encouraged to do different things. For example, a participant might decide to close her eyes in a block with only A or AV stimuli in order to better “concentrate” on the auditory stimuli. The same participant would obviously follow another strategy in the block with the visual stimuli. This exaggerated, but plausible, example makes clear that blocked stimulus presentation almost surely leads to a violation of the context invariance assumption, which invalidates the whole derivation of the RMI. A similar problem arises in so-called pseudorandomized stimulus sequences: Many studies use stimulus sequences in which specific conditions are not presented more than three times in a row (e.g., mandatory shift of the modality after V—V—V). In such sequences it cannot be ruled out that motivated participants will become aware of this experimental strategy and prepare for modality shifts, which again violates the

assumption of context invariance, and makes it difficult to interpret a violation or non-violation of the RMI.

On the other hand, random stimulus sequences might produce artificial redundancy gains. In random stimulus sequences, modality shifts are known to produce longer response latencies (e.g., Spence, Nicholls, & Driver, 2001). Because modality shifts affect only the unimodal conditions they might result in apparent coactivation because they do not occur in redundant stimuli. Therefore, several authors (e.g., Mozolic, Hugenschmidt, Pfeiffer, & Laurienti, 2008; Otto & Mamassian, 2012; Van der Stoep, Van der Stigchel, & Nijboer, 2015) presented sensitivity analyses that are based on the subset of stimuli that are free of modality shifts (e.g., A—A, V—V, AV—AV; e.g. Gondan, Lange, Rösler, & Röder, 2004), or stimuli that follow a specific modality (A—A, A—V, A—AV; e.g., Miller, 1986).

**Minimizing fast guesses and controlling for them** In the standard redundant signals task with simple responses, there are three stimulus conditions (A, V, and AV). A response is executed on each trial, which induces a tendency to anticipate the stimuli. Therefore, the target stimuli should not be presented in a fixed rhythm. Instead, the prestimulus period should be random, with the duration preferably sampled from an exponential distribution (see Luce, 1986, for an excellent introduction to this topic). Many studies sample from a uniform distribution instead. A uniform distribution, however, raises expectations, especially at its end: If the duration of the prestimulus period is sampled from a uniform distribution between 1,000 and 2,000 ms, and a stimulus has not appeared for 1,900 ms, the participant can be sure that the stimulus will appear immediately and will be tempted to anticipate it. As a general rule, everything that makes the stimulus sequence foreseeable should be avoided unless the researcher is particularly interested in sequence effects and expectations.

Researchers should always consider using catch trials (i.e., trials with no stimulus, C) in redundant signals tasks with simple responses and randomly mix them in the stimulus sequence. Catch trials, by instruction, discourage participants from fast guesses. If fast guesses occur nevertheless, it is possible to estimate the guessing tendency from the “responses” to catch trials, that is  $\hat{G}_C(t)$ , and use Inequality 8—thereby automatically applying the aforementioned kill-the-twin correction to eliminate guessing bias.

**Power and sample size** Consider a test of the RMI at a single time point  $t_0$ . If the number of trials  $n$  is large enough, the proportion of responses below  $t_0$ ,  $\hat{G}(t_0)$  follows an approximate Normal distribution with expectation  $\mu = F(t_0)$  and variance  $\sigma^2 = F(t_0) \times [1 - F(t_0)] / n$ . Let  $n = 50$  trials per condition. We assume that the RMI is violated, that is, the alternative hypothesis holds, with  $F_{AV}(t_0) = 0.20$ ,  $F_A(t_0) = 0.10$ , and  $F_V(t_0) = 0.05$ . For simplicity, we assume that the participants

do not anticipate, so that  $F_C(t_0)$  is zero. The difference between the three proportions  $d(t_0) = \hat{G}_{AV}(t_0) - \hat{G}_A(t_0) - \hat{G}_V(t_0)$  is then approximately normally distributed,

$$\begin{aligned} \hat{G}_{AV}(t_0) &\sim N(\mu_{AV} = 0.20, \sigma_{AV}^2 = 0.20 \times 0.80 / 50 = 0.003) \\ \hat{G}_A(t_0) &\sim N(\mu_A = 0.10, \sigma_A^2 = 0.10 \times 0.90 / 50 = 0.002) \\ \hat{G}_V(t_0) &\sim N(\mu_V = 0.05, \sigma_V^2 = 0.05 \times 0.95 / 50 = 0.001) \\ d(t_0) &\sim N\left(\begin{array}{l} \mu_{\text{diff}} = \mu_{AV} - \mu_A - \mu_V = 0.05, \\ \sigma_{\text{diff}}^2 = \sigma_{AV}^2 + \sigma_A^2 + \sigma_V^2 = 0.006 \end{array}\right) \end{aligned}$$

In the paired  $t$  test all participant-specific RMI differences are essentially added up, standardized and compared to zero. If the experiment is run with  $N = 20$  participants (indexed by  $i$ ), the approximate power of the RMI test at the one-tailed  $\alpha = 5\%$  is easily obtained,

$$\begin{aligned} \sum_i d_i(t_0) &\sim N(\mu_{\text{all}} = N\mu_{\text{diff}} = 1.00, \sigma_{\text{all}}^2 = N\sigma_{\text{diff}}^2 = 0.119), \\ \sum_i \frac{d_i(t_0)}{\sigma_{\text{all}}} &\sim N(2.90, 1). \end{aligned}$$

This number is above  $z_{0.95}$  with a probability of about 90%. The power estimate is slightly too optimistic (e.g.,  $z$  approximation instead of the  $t$  distribution, absence of fast guesses, test at a single percentile instead of permutation test, assumption of a quite large RMI violation at the same quantile or time point in each participant), but it illustrates that redundant signals experiments should be undertaken with enough participants and enough trials per condition.

Whereas the above design aspects could be considered as mandatory, we now outline some ways to make the experiment more “interesting” from the researcher’s point of view. By manipulating the strength of the stimuli, their relative onset, and the task requirements, one can get insight into the mechanisms underlying the integration of redundant information.

**Go/no-go and choice response tasks** In many redundant signals experiments, participants make simple responses to A or V stimuli or AV stimuli (e.g., Miller, 1986). The conclusions that can be drawn from such a basic experimental setup are, of course, limited (Miller, 1982). For example, researchers might wish to discern if redundancy gains indeed reflect some kind of “integration” of the response mappings at the decisional stage, as proposed by some coactivation models (e.g., Feintuch & Cohen, 2002). Alternatively, do redundancy gains merely result from a speeding of RTs because the increased stimulus energy with multiple stimuli leads to higher “arousal” (Nickerson, 1973; Posner, Nissen, & Klein, 1976)? The simple response task cannot distinguish these two underlying mechanisms.

To eliminate the possibility of generally higher arousal due to multiple stimuli in the redundant signal condition, some

studies have used target (+) and nontarget (−) stimuli in each modality, yielding the four stimulus types  $A^-V^-$ ,  $A^+V^-$ ,  $A^-V^+$ ,  $A^+V^+$ , with  $A^-V^-$  acting as the catch trial and  $A^+V^+$  being the redundant stimulus (e.g., Gondan, Götze, & Greenlee, 2010; Gondan, Niederhaus, Rösler & Röder, 2005; Grice & Reed, 1992; Experiment 4 in Miller, 1982; Otto & Mamassian, 2012; Schröger & Widmann, 1998). The intensity of the overall stimulation can be kept constant across the experimental conditions. The redundancy gain corresponds to the difference between mean RTs for  $A^+V^+$  and the mean RTs for  $A^+V^-$  and  $A^-V^+$  alone. However, the interpretation of the redundancy gains is difficult because response conflicts may occur in  $A^+V^-$  and  $A^-V^+$  (respond to  $A^+$  but withhold the response in  $V^-$ ) and may slow down the responses in these two conditions. Artificial redundancy gains may arise because  $A^+V^+$  is obviously free of response conflicts (Fournier & Eriksen, 1990). For example, Grice and Reed (1992) reported slower responses for target stimuli that were accompanied by a nontarget distractor ( $A^+V^-$ ), compared to single targets ( $A^+$ ).

An interesting experimental solution to this problem has been implemented by Grice, Canham, and Boroughs (1984). In addition to go- and no-go stimuli, Grice et al. introduced a neutral stimulus (0) in each modality, yielding three target combinations ( $A^+V^0$ ,  $A^+V^+$ ,  $A^0V^+$ ), and three nontarget combinations ( $A^-V^0$ ,  $A^-V^-$ ,  $A^0V^-$ ). Note that the neutral stimulus was always paired with either a target or a nontarget, so there were no ambiguous combinations such as  $A^0$ ,  $V^0$ , or  $A^0V^0$ . When comparing  $A^+V^+$  (redundant condition) to  $A^-V^0$  and  $A^0V^+$  (single target conditions), the intensity of the stimulation is kept constant in all conditions, and response conflicts in the unisensory conditions are avoided (e.g., Blurton et al., 2014; Feintuch & Cohen, 2002). For choice reaction tasks with two response alternatives, the design is identical, with the “−”-stimuli being associated with the second response alternative.

In go/no-go tasks or choice reaction tasks, the tendency to anticipate might be reduced; nevertheless, participants might still be tempted to produce “uninformed” responses (i.e., guess the correct response). Because of the multiple response alternatives, the kill-the-twin correction is slightly more complicated in go/no-go and choice response tasks. For the above setup with  $A^+V^+$ ,  $A^+V^0$ , and  $A^0V^+$  as go trials, the overall tendency to give an uninformed go response can be estimated by the false alarm rates in  $A^-V^-$ ,  $A^-V^0$ , and  $A^0V^-$ , respectively. This yields the following corrected inequality (see the original article by Eriksen, 1988, for a derivation and justification),

$$[F_{A^+V^+}(t) - F_{A^-V^-}(t)] \leq [F_{A^+V^0}(t) - F_{A^-V^0}(t)] + [F_{A^0V^+}(t) - F_{A^0V^-}(t)], \text{ for all } t. \quad (10)$$

Omitting the kill-the-twin correction can again bias the test in favor of the race model.

**Stimulus onset asynchrony manipulation** The RMI can be generalized to describe the upper bound for the redundancy gain in stimuli presented with onset asynchrony. For illustration, consider an auditory stimulus that is presented after the visual stimulus with a delay of 100 ms ( $V100A$ ). The following upper bound can be derived as a generalization of Inequality 8:

$$[F_{V100A}(t) - F_C(t)] \leq [F_A(t-100) - F_C(t-100)] + [F_V(t) - F_C(t)], \text{ for all } t,$$

which is easily seen to simplify to

$$F_{V100A}(t) + F_C(t-100) \leq F_A(t-100) + F_V(t), \quad (11)$$

for all  $t$ .

The expression  $F(t-100) = P(\mathbf{T} \leq t-100) = P(\mathbf{T} + 100 \leq t)$  implies that 100 ms are added to the RTs observed for A and C, consistent with the delayed onset of the auditory component in  $V100A$  (Miller, 1986) and with the shifted correction for fast guesses of the responses to unimodal auditory stimuli (for details, see Inequality 7 in Gondan & Heckel, 2008).

SOA manipulation is recommended for two reasons: If the processing times for the unimodal stimuli differ a lot, the redundancy gain might be very limited under any cognitive architecture, be it a race model or a coactivation model. If A is processed much faster than V, there is no reason to “wait” for the V stimulus in a redundant signals task. The RMI is not a powerful tool in such a situation. SOA variation might reestablish “physiological synchrony” (Hershenson, 1962), by equalizing the channel-specific processing times. Gondan (2009) provided a solution to the problem of multiple testing that arises from testing multiple SOAs. SOA variation can also be diagnostic regarding the mechanisms involved in coactivation (see Inequalities 13 and 14). Moreover, a number of coactivation models make testable predictions about the relationship between SOA and mean RT (e.g., Diederich, 1995; Colonius & Diederich, 2004; Gondan & Blurton, 2013; Miller & Ulrich, 2003; Schwarz, 1989, 1994; Ulrich & Miller, 1997).

**Generalized RMIs** Townsend and Nozawa (1995) investigated  $2 \times 2$  factorial experiments in which Factor A selectively influences the processing time distribution of Process A, and Factor V selectively influences the processing time distribution of Process V (see Dzhamfarov, 2003, for a formal definition of selective influence, and note the equivalence to the assumption of context invariance). Consider an experiment in which participants respond to auditory-visual stimuli, with the stimulus components being presented in different intensities, for example, weak (a, v) and strong (A, V). If the processing time distributions for the different levels of the two factors are ordered (e.g., faster processing of A than a, faster processing



of V than v), Townsend and Nozawa demonstrated that the race model predicts

$$F_{AV}(t) + F_{av}(t) \leq F_{Av}(t) + F_{aV}(t), \text{ for all } t. \quad (12)$$

Inequality 12 is easily seen to be a generalization of the RMI with kill-the-twin correction. Inequality 8 results when the intensity of the weak stimuli is zero. The generalization is important for two reasons: First, it allows for much more experimental flexibility because the basic setup (A, V, AV, catch trial) is being extended to a more realistic setting which includes only redundant stimuli. For example, in multisensory research, it is much more realistic to study the details of speech integration with audio-visual speech samples instead of comparing “natural” AV stimuli to “artificial” samples of A and V (e.g., Chandrasekaran, Lemus, Trubanova, Gondan, & Ghazanfar, 2011).

Even more importantly, Townsend and Nozawa (1995) derived predictions for the sign of the interaction contrast  $F_{AV}(t) + F_{av}(t) - F_{Av}(t) - F_{aV}(t)$  not only for the race model but also for other architectures, including serial self-terminating models, serial exhaustive models, parallel exhaustive models, and coactivation models. Applying the technique of Townsend and Nozawa, Gondan and Blurton (2013) derived similar predictions for stimuli presented with onset asynchrony and derived a further generalization of the RMI. Let AV and 100AV denote synchronous stimuli presented at  $t = 0$  and 100 ms after the “usual” stimulus onset (the meaning of this depends on the specific experiment), and A100V and V100A denote asynchronous stimuli. Then, under the usual assumptions, the race model predicts

$$F_{AV}(t) + F_{100AV}(t) \leq F_{A100V}(t) + F_{V100A}(t), \text{ for all } t. \quad (13)$$

When the onset asynchrony is increased further and further, Inequality 13 converges again to the RMI with kill-the-twin correction (see Inequality 8). Again, predictions can be derived for the sign of the interaction contrast  $F_{AV}(t) + F_{100AV}(t) - F_{A100V}(t) - F_{V100A}(t)$  for several processing architectures (serial, parallel, coactive; summarized in Gondan & Blurton, 2013).

This list is necessarily nonexhaustive and mainly reflects the authors’ experience and personal taste. Other very promising approaches exist that acquire response force, RTs for bimanual responses, and get important insights by manipulating the contingencies between the stimulus components (e.g., Diederich & Colonius, 1987; Giray & Ulrich, 1993; Miller, 1991; Mordkoff & Yantis, 1991, 1993).

### When the race model fails . . .

We have shown above that the RMI is a direct consequence of a parallel, first-terminating model (i.e., race model) and the

assumption of context invariance. Violation of Inequality 8 or Inequality 10 indicates that either the race model is wrong, the context invariance assumption is untenable, or both (see Miller, 2015, for a discussion). In many studies violation of the RMI is considered as evidence for the validity of “the coactivation model” which assumes integrated processing of the redundant information. Because of the great variety of coactivation models, this line of reasoning is incomplete as long as the specific mechanism of integration is not specified and tested.

Since Miller (1982), a number of coactivation models have been developed. Some of these models are very general (e.g., Miller, 1986, Equation 3) while others predict specific shapes of the RT distribution (e.g., Otto & Mamassian, 2012). In the following we list an incomplete set of candidates that might be considered if the race model has been shown to fail, and the researcher is interested in what happened instead. It is very likely that there is no universal coactivation model that accounts for the redundancy gains observed in all the studies of the field—studies that differ with respect to stimuli, task demands, and participants.

In the history-free “exponential” coactivation model (Miller, 1986, Inequality 3) stimulus detection occurs spontaneously, in an all-or-none fashion, comparable to the spontaneous decay of a radioactive particle. In the bimodal stimulus, detection is assumed to occur at a higher rate, such that the expected detection time is shorter than the detection time for unimodal stimuli. What happens with asynchronous stimuli (e.g., V100A)? Miller argued that in the exponential coactivation model, detection of the asynchronous signal V100A occurs at the same rate as for V, for the first 100 ms. Starting at 100 ms after stimulus onset, both stimuli are active, and detection occurs at the increased bimodal detection rate. This is comparable to a race of V and a delayed AV, with the visual racer being stopped at 100 ms, so that  $\mathbf{D}_{V100A} \geq \min(\mathbf{D}_V, \mathbf{D}_{AV} + 100)$ . The following upper limit is derived (note  $F_{AV}$  instead of  $F_A$  on the right side of the inequality)<sup>2</sup>,

$$F_{V100A}(t) + F_C(t - 100) \leq F_V(t) + F_{AV}(t - 100), \quad (14)$$

for all  $t$ .

Violation of the upper bound defined by Inequality 14 rules out history-free coactivation models in favor of accumulator models in which sensory evidence is collected over time.

Specific examples of such accumulator models are the Poisson counter and diffusion superposition models proposed by Diederich (1995) and Schwarz (1989, 1994), and the parallel grains model (Miller & Ulrich, 2003). In the counter superposition model (Schwarz, 1989), buildup of

<sup>2</sup> Miller (1986) did not include a kill-the-twin correction, so his Inequality 3 does not include the  $F_C$  term. Because the kill-the-twin correction can be shown to hold for a standard race model for asynchronous stimuli (Gondan & Heckel, 2008), and because in the exponential coactivation model  $F_{V100A}(t)$  can only decrease in comparison to the standard race model, the correction term is justified.

sensory evidence is conceived as a series of impulses with independent, exponentially distributed intervals between two impulses. Detection occurs as soon as a criterion of  $c > 1$  impulses has been registered. In bimodal targets, the impulses are pooled in a common channel (linear superposition). As a consequence, the criterion is reached earlier, which causes the redundancy gain. Schwarz (1989, Figure 1) showed that the model accurately predicts the mean RTs for speeded manual responses to auditory-visual stimuli with different onset asynchronies (see discussion in Diederich & Colonius, 1991; Schwarz, 1994). In the diffusion superposition model (Diederich, 1995; Schwarz, 1994), buildup of sensory evidence is conceived as a time-homogeneous diffusion process with drift. Again, detection occurs when a criterion  $c > 0$  is reached for the first time. In bimodal stimuli, the activation of the two channels is pooled and the criterion is again reached earlier. Schwarz (1994) demonstrated that the diffusion superposition model accurately describes the means and variances of the RTs at different SOAs reported by Miller (1986). In the parallel grains model (Miller & Ulrich, 2003), buildup of sensory evidence is described by a discrete accumulation of information. Similar to Schwarz' (1989) model, in auditory-visual stimuli, the contributions of both sensory channels are pooled, so that a criterion of  $c$  impulses is reached earlier than in single signals. Prediction of the mean RTs in the standard redundant signals experiment seems to be quite satisfactory (Miller & Ulrich, 2003, Figures 7 and 8).

A second class of coactivation models keeps the assumption of a race between channels (Eq. 1), but instead drops the notion of context invariance (Eq. 4A): For illustration, consider a minimalistic redundant signals task with 1/3 A stimuli, 1/3 V stimuli, and 1/3 AV stimuli. The prior probability for a stimulus in the auditory modality is then 2/3, because two out of the three experimental conditions include an auditory component. If, however, a visual stimulus is presented, the posterior probability for an auditory stimulus is decreased to 1/2, because one out of two conditions that include a visual component also includes an auditory component. Mordkoff and Yantis (1991) systematically manipulated such contingencies as well as contingencies of the response relevance of the different channels in a go/no-go task (e.g., with the auditory component predicting a target in the visual channel). The pattern of results indicated that the system adapts to such contingencies by excitatory interactions if the posterior probability is greater than the prior probability, and inhibitory interactions in the opposite case. Mordkoff and Yantis called the model an “interactive race model.”

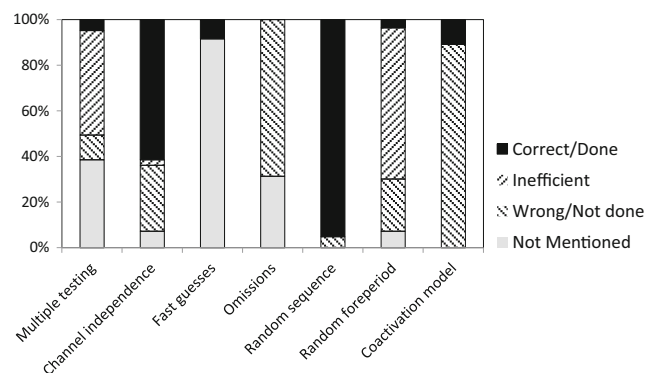
In a recent approach, Otto and Mamassian (2012) investigated audio-visual redundancy gains with a special focus on sequence effects. Effects of the stimulus sequence have also been reported by Miller (1986, Table 4), Gondan et al. (2004) and Van der Stoep et al. (2015). Whereas Miller and Gondan

et al. basically tested the RMI in subgroups of stimuli (e.g., only stimuli that follow an A trial; see above), Otto and Mamassian explicitly modeled the effects of the stimulus sequence. They then demonstrated that the race mechanism can explain the redundancy gain when allowing for increased noise in the system in bimodal stimuli. For further examples and taxonomies see Ulrich and Miller (1997), Colonius and Townsend (1997), and Townsend and Nozawa (1997).

## Current practice

To outline the current practice of testing the RMI, a review from 2011 to 2014 was conducted on research articles that cited Miller's 1982 paper (Google Scholar and ISI Web of Science). This yielded 181 papers, wherein 98 (54%) articles cited the redundant signals effect but did not explicitly utilize or test a race model. The methods of the remaining 83 studies in which the RMI was tested are summarized in Fig. 2.

Out of the 83 articles that tested a race model, 77 (93%) papers claimed to have obtained a violation of the RMI. Yet, out of the 83 articles reviewed, only 9 (11%) papers tested a coactivation model to describe and predict their data. Surprisingly, only 58 of these 83 studies (70%) did some kind of statistical analysis, mostly percentile-specific  $t$  tests, with 9 (10%) studies not correcting for multiple testing at all, and 23 studies (28%) using an overly conservative Bonferroni correction. Only 4 out of 83 studies (5%) used a permutation test that allows for an adequate Type I error control while preserving statistical power. In a surprisingly high number of studies (24 of 83), it was claimed that the race model inequality or “Miller inequality” was tested, but these studies actually used the so-called independent race model (Eq. 7), which results if the channel processing times are assumed to be stochastically



**Fig. 2** Current practice of RMI testing within 2010–2014. Only a few studies handled issues of multiple testing appropriately, and nearly all studies estimated the RT distributions involved in the RMI in an incorrect way. Many studies tested the “independent” race model instead of Miller's (1982) inequality. The study designs generally met the requirements for context invariance. Although the RMI is violated in many studies, coactivation models are rarely investigated

independent. The shortcomings of this additional assumption were only discussed in 2 (3%) out of 83 of these studies.

None of the studies (0%) explicitly incorporated a procedure that would account for omitted responses in the distribution functions. Fifty-seven (69%) studies ignored misses by excluding them from the analyses. The other 26 studies (31%) did not mention how misses were treated in the analysis. Most articles had a priori defined cutoff boundaries for the RT distributions and fast guesses and misses were not accounted for in the distribution functions. When plots of the distribution functions were given, the plots were rescaled such that  $\hat{G}(t)$  was 0% at the lower cutoff boundary and 100% at the upper cutoff boundary, which indicates that omitted responses and other problematic behavior were simply excluded. Only 7 out of the 83 studies (8%) did run a kill-the-twin procedure, which corrects the distribution functions from becoming artificially larger; in the other studies, the kill-the-twin procedure was not mentioned (and probably not performed).

Seventy-nine (95%) out of the 83 studies correctly implemented an intermixed presentation of their experimental stimuli (instead of presenting the visual, auditory, and audio-visual stimuli in separate blocks). Although exponentially distributed foreperiods have already been recommended by the seminal book on reaction times by Luce (1986), only three articles implemented this timing procedure (4%). Fifty studies (60%) sampled from uniform distributions instead, and 23 studies (28%) used a constant foreperiod.

This short review is intended to underline the need for careful study design and data analysis in research on redundant signals effects. Some of the aspects summarized in Fig. 2 can be considered to be specific to redundant signals tasks and to testing the RMI, but others are equally applicable to other RT experiments and analyses (e.g., random foreperiods and stimulus sequences, careful justification for data cleaning procedures routinely applied in RT research). Researchers should be aware that wrong or inefficient methodology may yield biased results (e.g., anticonservative decisions if issues of multiple testing are not taken serious, loss of power if the kill-the-twin correction is not performed) and may lead to wrong conclusions.

## Discussion

Hypothesis tests in RT studies in Experimental Psychology most often refer to differences in the mean RTs observed under two experimental conditions. Redundant signals tasks differ from this standard setup: (A) The test of the RMI refers to a comparison of three instead of two experimental conditions. (B) Hypotheses refer to cumulative RT distributions instead of mean RTs. (C) The violation of the RMI does not automatically support a specific theory.

As a consequence of (A), the assumptions of standard data cleaning procedures are not automatically met. There are experimental paradigms in which standard data cleaning procedures yield improved estimates and valid inference. In redundant signals tasks, matters are more complicated. Biases induced by errors and outliers cannot simply be removed by removing the errors and outliers from the analysis. Even in standard experiments comparing the “mean correct RT” between two experimental conditions, exclusion of fast responses, omitted responses, and outliers is often based on common-sense arguments and ad hoc considerations of statistical power (e.g., Ratcliff, 1993). When the null hypothesis states that “behavior” is equal under two experimental conditions, this statement is easily seen to translate to the hypothesis of equal “mean correct RTs.” In contrast, because of the asymmetry of the RMI and the probabilities involved, conventional error removal and other ad hoc data cleaning procedures may cause bias (Eriksen, 1988; Miller, 2004). Especially in experiments with simple reaction time tasks, it is advisable to include a condition with catch trials. Catch trials reduce the tendency for fast guesses and allow researchers to perform a kill-the-twin correction by applying Inequality 8. For go/no-go tasks or choice response tasks, corresponding kill-the-twin schemes can be developed that basically result in Inequality 10 (e.g., Blurton et al., 2014; Gondan et al., 2010). In general, errors, omissions and outliers are an integral part of RT experiments.

Problems of multiple testing arise as a consequence of (B) because the RMI is predicted to hold for all  $t$ , thus, violations might occur at any  $t$ . Although it is obvious that testing at multiple time points leads to an accumulation of Type I errors (see also Kiesel et al., 2007), many of the reviewed studies did not account for multiple testing. Other studies used standard procedures for Type I error control (i.e., the Bonferroni procedure), which assume the worst case of negatively correlated significance tests. For the test of the RMI, however, the tests at multiple time points or quantiles are positively correlated. If the test is significant at time point  $t$ , it is probably also significant in close proximity to  $t$ . Therefore, we recommend using the permutation test for the RMI (Gondan, 2010). A permutation test that also includes the kill-the-twin correction is provided in the [Supplementary material](#) to this article.

When the RMI fails, which mechanism can account for the observed results? A number of coactivation models have been proposed that can be broadly classified into context-invariant models that assume different mechanisms of integration (e.g., superposition models) and the so-called interactive race models that drop the assumption of context-invariance. The superposition models proposed by Diederich (1995), Schwarz (1989, 1994) and by Miller and Ulrich (2003, Section 6) are all based on additive superposition of activity in the two channels. In line with this, models of neural integration sites predict that integration of complex information is optimal if the

signals arising from different sources are additively integrated (Denève, Latham, & Pouget, 2001). This is in remarkable contrast to the conventional definition of multisensory brain regions that underlines the special role of “superadditive” responses (e.g., Wallace, Meredith, & Stein, 1992; see Stevenson et al., 2014, for a discussion). Given that a number of phenomena can be explained with additive superposition, it seems worth considering them in future investigations of behavior and neural firing (e.g., Stanford, Quessy, & Stein, 2005).

**Summary** Testing the RMI should follow the following recommendations: (A) The RMI should be tested at multiple time points. A permutation test should be used to control the significance level in these multiple tests. (B) Data cleaning should be avoided. Outliers and omitted responses should not be eliminated. Erroneous responses should not be eliminated from the data either. The kill-the-twin correction should be applied instead; in other words, Inequality 8 (simple response tasks) or 10 (go/no-go or choice response tasks) should be tested. (C) When the race model fails, coactivation models should be considered.

**Acknowledgements** We would like to thank Jeff Miller and an anonymous reviewer for helpful their comments and suggestions.

## References

- Bahadur, R. R. (1966). A note on quantiles in large samples. *Annals of Mathematical Statistics*, 37, 577–580.
- Blair, R. C., & Karninski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, 30, 518–524.
- Blurton, S. P., Greenlee, M. W., & Gondan, M. (2014). Multisensory processing of redundant information in go/no-go and choice responses. *Attention, Perception, & Psychophysics*, 76, 1212–1233.
- Chandrasekaran, C., Lemus, L., Trubanova, A., Gondan, M., & Ghazanfar, A. A. (2011). Monkeys and humans share a common computation for face/voice integration. *PLoS Computational Biology*, 7, e1002165.
- Colonus, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, 16, 1000–1009.
- Colonus, H., & Diederich, A. (2006). The race model inequality: Interpreting a geometric measure of the amount of violation. *Psychological Review*, 113, 148–154.
- Colonus, H., & Townsend, J. T. (1997). Activation-state representation of models for the redundant-signals-effect. In A. A. J. Marley (Ed.), *Choice, decision and measurement* (pp. 245–254). New York, NY: Erlbaum.
- Corballis, M. C. (2002). Hemispheric interactions in simple reaction time. *Neuropsychologia*, 40, 423–434.
- Denève, S., Latham, P., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, 4, 826–831.
- Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, 39, 197–215.
- Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? A reaction time analysis. *Psychological Research*, 49, 23–29.
- Diederich, A., & Colonius, H. (1991). A further test of the superposition model for the redundant-signals effect in bimodal detection. *Perception & Psychophysics*, 50, 83–86.
- Dzhafarov, E. N. (2003). Selective influence through conditional independence. *Psychometrika*, 68, 7–25.
- Dzhafarov, E. N., Schweickert, R., & Sung, K. (2004). Mental architectures with selectively influenced but stochastically independent components. *Journal of Mathematical Psychology*, 48, 51–64.
- Eriksen, C. E. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception & Psychophysics*, 44, 191–193.
- Feintuch, U., & Cohen, A. (2002). Visual attention and coactivation of response decisions for features from different dimensions. *Psychological Science*, 13, 361–369.
- Foumier, L. R., & Eriksen, C. W. (1990). Coactivation in the perception of redundant targets. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 538–550.
- Giray, M., & Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 1278–1291.
- Gondan, M. (2009). Testing the race inequality in redundant stimuli with variable onset asynchrony. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 257–259.
- Gondan, M. (2010). A permutation test for the race model inequality. *Behavior Research Methods*, 42, 23–28.
- Gondan, M., & Blurton, S. P. (2013). Generalizations of the race model inequality. *Multisensory Research*, 26, 95–122.
- Gondan, M., Götz, C., & Greenlee, M. W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Attention, Perception, & Psychophysics*, 72, 1692–1709.
- Gondan, M., & Heckel, A. (2008). Testing the race inequality: A simple correction procedure for fast guesses. *Journal of Mathematical Psychology*, 52, 320–323.
- Gondan, M., Lange, K., Rösler, F., & Röder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review*, 11, 307–313.
- Gondan, M., Niederhaus, B., Rösler, F., & Röder, B. (2005). Multisensory processing in the redundant-target effect: An event-related potential study. *Perception & Psychophysics*, 67, 713–726.
- Grice, G. R., & Canham, L. (1990). Redundancy phenomena are affected by response requirements. *Perception & Psychophysics*, 48, 209–213.
- Grice, G. R., Canham, L., & Boroughs, J. M. (1984a). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics*, 35, 451–463.
- Grice, G. R., Canham, L., & Gwynne, J. W. (1984b). Absence of a redundant-signals effect in a reaction time task with divided attention. *Perception & Psychophysics*, 36, 565–570.
- Grice, G. R., & Reed, J. M. (1992). What makes targets redundant? *Perception & Psychophysics*, 51, 437–442.
- Mozolic, J. L., Hugenschmidt, C. E., Pfeiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, 184(1), 39–52.
- Hyndman, R. J., & Fan, Y. (1996). Sample quantiles in statistical packages. *American Statistician*, 50, 361–365.
- Kiesel, A., Miller, J., & Ulrich, R. (2007). Systematic biases and Type I error accumulation in tests of the race model inequality. *Behavioral Research Methods*, 39, 539–551.
- Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, 63, 289–293.
- Luce, R. D. (1986). *Response times: Their role in inferring mental organization*. New York, NY: Oxford University Press.

- Maris, G., & Maris, E. (2003). Testing the race model inequality: A nonparametric approach. *Journal of Mathematical Psychology*, *47*, 507–514.
- 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.
- Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 160–169.
- Miller, J. (2004). Exaggerated redundancy gain in the split brain: A hemispheric coactivation account. *Cognitive Psychology*, *49*, 118–154.
- Miller, J., & Adam, J. J. (2006). Redundancy gain with static versus moving hands: A test of the hemispheric coactivation model. *Acta Psychologica*, *122*, 1–10.
- Miller, J. (2015). Statistical facilitation and the redundant signals effect: What are race and coactivation Models? *Attention, Perception, & Psychophysics*. doi: 10.3758/s13414-015-1017-z.
- Miller, J., & Lopes, A. (1991). Bias produced by fast guessing in distribution-based tests of race models. *Perception & Psychophysics*, *50*, 584–590.
- Miller, J., & Ulrich, R. (2003). Simple reaction time and statistical facilitation. A parallel grains model. *Cognitive Psychology*, *46*, 101–151.
- Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 520–538.
- Mordkoff, J. T., & Yantis, S. (1993). Dividing attention between color and shape: Evidence of co-activation. *Perception & Psychophysics*, *53*, 357–366.
- Nickerson, R. S. (1973). Intersensory facilitation of reaction time: Energy summation or preparatory enhancement. *Psychological Review*, *80*, 489–509.
- Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, *6*, 155–156.
- Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology*, *22*, 1391–1396.
- Pesarin, F. (2001). *Multivariate permutation tests with applications in biostatistics*. New York, NY: Wiley.
- Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, *83*, 157–171.
- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590.
- Ratcliff, R. (1993). Methods for dealing with reaction-time outliers. *Psychological Bulletin*, *114*, 510–532.
- Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*, 755–759.
- Schröter, H., Ulrich, R., & Miller, J. (2007). Effects of redundant auditory stimuli on reaction time. *Psychonomic Bulletin & Review*, *14*, 39–44.
- Schwarz, W. (1989). A new model to explain the redundant signals effect. *Perception & Psychophysics*, *46*, 498–500.
- Schwarz, W. (1994). Diffusion, superposition, and the redundant targets effect. *Journal of Mathematical Psychology*, *38*, 504–520.
- Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*, 330–336.
- Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *Journal of Neuroscience*, *25*, 6499–6508.
- Stevenson, R. A., Ghose, D., Krueger Fister, J., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., Kurela, L. R., Siemann, J. K., James, T. W., & Wallace (2014). Identifying and quantifying multisensory integration: a tutorial review. *Brain Topography*, *27*, 707–730.
- Todd, J. W. (1912). Reaction to multiple stimuli. *Archives of Psychology*, *25*, 1–65.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York, NY: Cambridge University Press.
- Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–359.
- Townsend, J. T., & Nozawa, G. (1997). Serial exhaustive models can violate the race model inequality: Implications for architecture and capacity. *Psychological Review*, *104*, 595–602.
- Ulrich, R., & Miller, J. (1997). Tests of race models for reaction time in experiments with asynchronous redundant signals. *Journal of Mathematical Psychology*, *41*, 367–381.
- Ulrich, R., Miller, J., & Schröter, H. (2007). Testing the race model inequality: An algorithm and computer programs. *Behavior Research Methods*, *39*, 291–302.
- Van der Stoep, N., Van der Stigchel, S., & Nijboer, T. C. W. (2015). Exogenous spatial attention decreases audiovisual integration. *Attention, Perception, & Psychophysics*, *77*, 464–482.
- Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory inputs in cat cortex. *Experimental Brain Research*, *91*, 484–488.
- Yellott, J. I. (1971). Correction for fast guessing and the speed-accuracy trade-off in choice reaction time. *Journal of Mathematical Psychology*, *8*, 159–199.