

RESEARCH ARTICLE

# When one race is not enough: A relay model explains multisensory response times

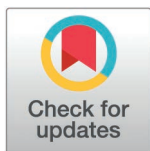
Kalvin Roberts<sup>\*</sup>, Thomas U. Otto<sup>\*</sup>

School of Psychology and Neuroscience, University of St. Andrews, St. Andrews, United Kingdom

\* [ker9@st-andrews.ac.uk](mailto:ker9@st-andrews.ac.uk) (KR), [to7@st-andrews.ac.uk](mailto:to7@st-andrews.ac.uk) (TUO)

## Abstract

Humans typically respond faster to multisensory signals than to their unisensory components, a phenomenon known as the redundant signal effect (RSE). One of the earliest and most influential accounts, the race model, attributes the RSE to statistical facilitation, which arises from parallel, independent processing across sensory modalities. While this model captures some key features of the RSE, it frequently underestimates the observed speed-up leading to violations of the race model inequality (RMI), a benchmark used to test the model's validity. To reconcile this discrepancy, we introduce the relay model, a minimal extension of the race architecture that incorporates cross-modal initiation. In this model, responses result from two sequential race processes, allowing a signal in one modality to initiate the onset of perceptual decision processing in another. This structure retains statistical facilitation as a core principle while introducing a single free model parameter that partitions unisensory processing into gating and decision stages. Through simulations and fits to foundational empirical datasets, we show that the relay model captures both the magnitude and distributional shape of the RSE, including RMI violations. It also accounts for changes in the RSE under asynchronous stimulus onsets and manipulations of signal intensity, which are critical tests in multisensory research. By extending the classical race model with minimal added complexity, the relay model offers a mechanistically explicit and biologically plausible framework for explaining the dynamics of multisensory decision-making.



## OPEN ACCESS

**Citation:** Roberts K, Otto TU (2026) When one race is not enough: A relay model explains multisensory response times. PLoS Comput Biol 22(5): e1013320. <https://doi.org/10.1371/journal.pcbi.1013320>

**Editor:** Ulrik R. Beierholm, Durham University, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

**Received:** July 10, 2025

**Accepted:** May 18, 2026

**Published:** May 26, 2026

**Copyright:** © 2026 Roberts, Otto. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** All data and code used to produce the results and analyses presented in this manuscript is available on a GitHub repository at <https://github.com/kalvinroberts/RelayModelCode>. We have also used Zenodo to assign a DOI to the repository: [10.5281/zenodo.20381337](https://doi.org/10.5281/zenodo.20381337).

## Author summary

How does the brain combine information from different senses, like hearing and seeing? For decades, researchers have found that people respond faster when signals from multiple senses occur together than when signals come from a single sense alone. Two main ideas have been proposed to explain this effect: one suggests that the senses operate independently, racing to trigger a response, while the other suggests that they cooperate to enhance processing.

**Funding:** This work was supported by the UKRI Biotechnology and Biological Sciences Research Council (BB/T00875X/1 to KR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

We here introduce a new computational account, the relay model, that bridges these ideas. In this model, perceptual decisions unfold in two stages. A first gating stage detects incoming signals and initiates a second decision stage that leads to a behavioural response. Importantly, detection in one sense can initiate decision-making in another. This relay-like mechanism allows both competition and cooperation between the senses. By capturing key findings from classic experiments, the model offers a new perspective on how the brain combines multisensory information to guide behaviour.

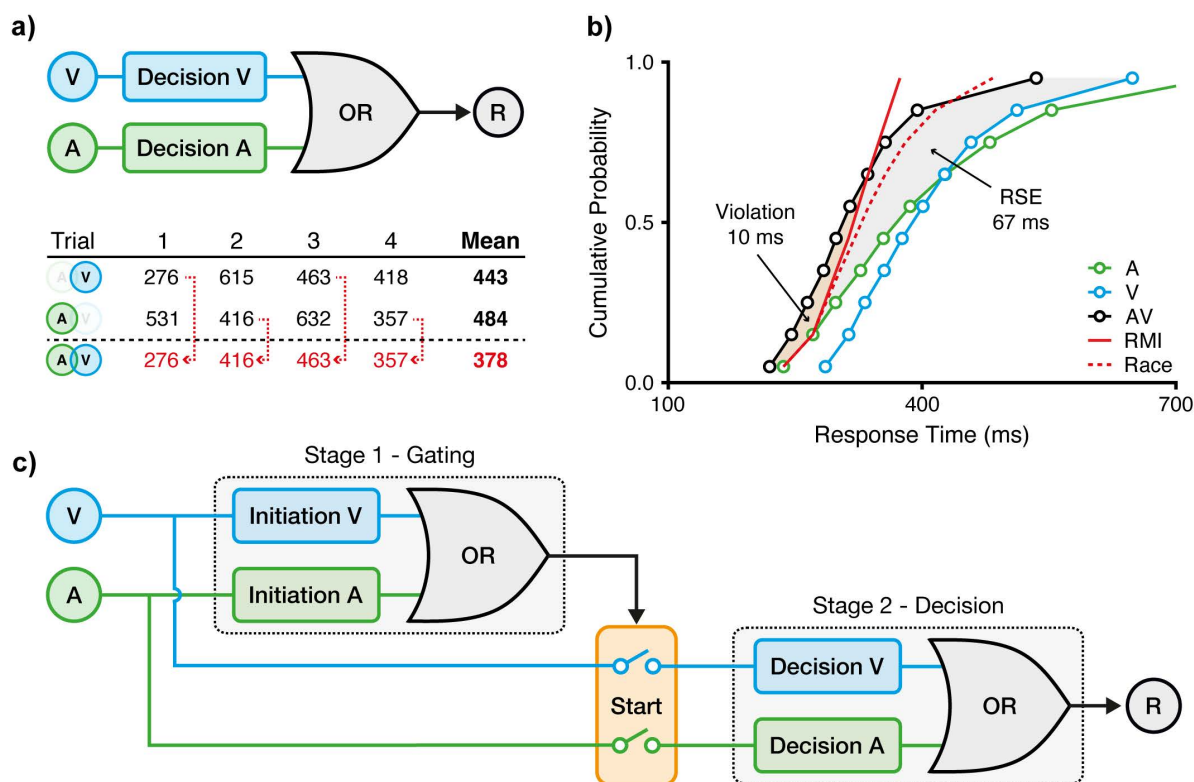
## Introduction

Perceptual decisions are rarely based on information from a single sense. Instead, the brain continuously uses inputs from multiple modalities to guide rapid and adaptive behaviour. This multisensory integration enhances accuracy and accelerates response times (RTs). For instance, if a person both sees an oncoming bicycle and hears its bell, they are likely to react more quickly than if only one of those signals were available. Understanding the computational mechanisms underlying multisensory integration is essential for building models that capture how the brain makes fast and reliable decisions in dynamic environments.

One of the most widely used paradigms to study the benefits of multisensory processing is the redundant signals paradigm [1–4]. In this task, participants are instructed to respond with the same action to different target signals, such as an auditory signal, a visual signal, or both. The critical feature is that signals in the combined condition are redundant, either alone is sufficient for a correct response. Thus, the task effectively implements a logical OR operation [5–7]. The key finding is that responses to redundant signals are, on average, faster than responses to single signals, a phenomenon known as the redundant signals effect (RSE). The effect has been replicated across a wide range of multisensory pairings (e.g., [8–14]), within single modalities (e.g., [15–20]), across populations (e.g., [21–25]), and under varying task instructions (e.g., [26–30]). This robustness makes the RSE a valuable tool for probing the computational principles of multisensory decision-making.

To understand the mechanisms behind this speed-up, it is useful to first consider how unisensory decisions are modelled. Computational and neurophysiological studies suggest that a key component is the accumulation of sensory evidence over time [31–33]. Specialised neurons modulate their firing in response to stimuli, but because of intrinsic noise, evidence must be integrated over time to improve reliability. Once a decision threshold is reached, a behavioural response is triggered. This process gives rise to typical RT distributions, such as the positively skewed inverse Gaussian (IG) [34,35], and has proven effective in capturing unisensory decision dynamics [36]. The central question, then, is: how can these models be extended to multisensory decisions?

A milestone answer was offered by Raab [5], who proposed the classical race model (Fig 1a). Here, two independent evidence accumulation units operate in



**Fig 1. Race models in multisensory decision-making.** **a)** Raab's race model. Auditory (A) and visual (V) signals are processed by two parallel decision units, coupled via a logical OR gate to trigger a response (R). The model predicts faster responses for redundant audiovisual (AV) signals through statistical facilitation, as illustrated by exemplary response times (RTs). On each AV trial, a response is determined by the faster of the two units. Averaging RTs across trials reveals a speed-up of responses for AV compared to unisensory signals. **b)** Miller's race model inequality (RMI) test. The cumulative RT distribution with redundant AV signals is located to the left of the unisensory RT distributions (A, V), reflecting the redundant signals effect (RSE, grey area). The observed RSE exceeds the prediction of Raab's race model, including significant violation of the RMI (orange area), indicating a processing interaction between audition and vision. Data replotted from [4]. **c)** Relay model. To reconcile Raab's race model with RMI violations, we propose a two-stage model comprising two sequential race processes. As an audiovisual processing interaction, the first gating stage provides a start signal for the second decision stage. The model allows for cross-modal initiation; for instance, a fast gating in vision can initiate decision process in both modalities. The second decision stage operates analogously to Raab's race model to trigger a response.

<https://doi.org/10.1371/journal.pcbi.1013320.g001>

parallel, each detecting one of the sensory signals. These units are linked by a logical OR gate, meaning that a response is triggered as soon as one or the other unit reaches its decision threshold. When both signals are presented, statistical facilitation occurs: if one process is slower on any given trial, the other may trigger the response first, thereby reducing average RTs (see example trials in Fig 1a). With audio-visual signals (AV), the mechanism of the race model is captured by selecting the faster of two unisensory RTs (for A and V, respectively):

$$RT_{AV} = \min(RT_A, RT_V) \quad (1)$$

Consequently, the cumulative distribution function (CDF) for RTs in the redundant condition follows probability summation, where the chance of responding by time  $t$  increases by having two independent chances to respond:

$$F_{AV}(t) = F_A(t) + F_V(t) - F_A(t) \times F_V(t) \quad (2)$$

Here,  $F_{AV}(t)$  is the cumulative probability of responding at time  $t$  to both signals, and  $F_A(t)$  and  $F_V(t)$  are the unisensory RT distributions. Thus, race models provide parameter-free predictions of the RSE based on unisensory data, offering concrete and testable predictions about the expected magnitude of the RSE and the conditions under which it should be most pronounced.

Building on probability summation in the race model, two key principles to predict the expected speed-up emerge [37]. First, the *principle of congruent effectiveness* predicts that the RSE is largest when unisensory RT distributions overlap. This overlap can be optimised by introducing a temporal delay between signals, known as stimulus onset asynchrony (SOA), which can offset inherent RT differences. For instance, if auditory responses are on average 30 ms faster, delaying the auditory relative to the visual signal by 30 ms aligns the two unisensory RT distributions and maximises the RSE (e.g., [2,37,38]). Second, the *variability rule* states that the RSE is driven by the variability of RTs in unisensory conditions, which can be modulated for example through signal intensity. As weaker signals tend to produce greater RT variability, the RSE is expected, and has been shown, to increase under such conditions (e.g., [37,39]). Together, these findings show that the race model captures key patterns of multisensory facilitation, reinforcing its value as a simple yet powerful explanatory framework.

Despite their predictive power, race models have clear limitations. As a second milestone, Miller [4] demonstrated that the RSE often exceeds what can be explained by probability summation alone. Allowing for a maximally negative correlation between unisensory processes [40], the probability of responding to redundant signals cannot exceed the sum of the individual unisensory probabilities:

$$F_{AV}(t) \leq F_A(t) + F_V(t) \quad (3)$$

This so-called race model inequality (RMI) provides a testable boundary for race model predictions (Fig 1b). Empirical RMI violations demonstrate that the observed RSE cannot be attributed to the race model alone, implying some interaction between sensory channels. Given the consistent reports of RMI violations, many researchers have moved away from race accounts (with some notable exceptions, [11,41]) and toward *single-threshold* accounts. In these models, sensory evidence from two (or more) modalities may be pooled in a single decision process to reach a shared threshold [e.g., 4, 42]. These pooling models are commonly referred to as coactivation models, although the term has not always been used consistently in the literature [43].

Importantly, the RMI, and classical race models in general, implicitly assume a single-stage architecture: two isolated, parallel accumulation processes leading directly to the same response (for comprehensive discussions of the underlying assumptions, see [7,44,45]). This assumption is analytically convenient but biologically simplistic. The pathway from sensory transduction to motor execution is unlikely to consist of a single step. Instead, information likely traverses multiple sequential processing stages [46,47]. At a purely agnostic level, these stages represent discrete processing epochs, regardless of their specific cognitive or neural nature. In principle, there could be many such stages, encompassing sensory encoding, feature extraction, and motor preparation [48]. The race model's limitation to a single processing stage is particularly relevant in a multisensory context, where interactions between modalities can occur at multiple levels [49–51]. As soon as the decision process includes sequential stages with the potential for crossmodal interactions, the foundational independence assumptions of the RMI framework no longer hold [52]. This raises a critical question: how can multisensory decision-making models be extended to capture biologically plausible multi-stage dynamics?

In unisensory decision-making, such models are increasingly prevalent. One particularly compelling candidate for the cognitive nature of these distinct stages posits an initial “gating” or selection process followed by a subsequent “decision” process involving evidence accumulation. For example, Carpenter et al. [53] proposed a two-stage architecture in which early detection units trigger a later accumulation unit (see also [47,54–57]). In these models, the initial gating stage determines when evidence accumulation begins, preventing the system from integrating noise before relevant sensory input

is available [47,58]. Without such gating, accumulation could start before stimulus onset, leading to random responses, increased RT variability, and unreliable confidence estimates. Biologically, gating may arise through sensory onset signals, internal go cues, or inhibitory dynamics that suppress pre-stimulus noise [47,58]. In modelling terms, this is often implemented as a gate function or a reset of initial conditions. Separating gating from evidence accumulation therefore provides a highly plausible mechanism that supports both computational stability and biological realism.

In multistage frameworks, the total RT is expressed as the sum of processing times associated with these individual stages (e.g.,  $RT_A = T_{A1} + T_{A2}$  for an auditory stimulus;  $RT_V = T_{V1} + T_{V2}$  for a visual stimulus). This additive logic is also reflected in standard models of perceptual decision making, such as the drift diffusion model [48], where the total RT is partitioned into a core decision process, which incorporates evidence accumulation, and a residual non-decision time, which subsumes early sensory encoding and late motor execution. While such multi-stage frameworks are well established in unisensory contexts, their extension to multisensory decision-making remains comparatively underexplored (but see [42,59]).

In this study, we develop and test an extension of the race model to address this gap in research on multisensory decision making. The model consists of two sequential stages: a first gating stage that detects incoming signals which initiates a second decision stage that leads to a behavioural response (Fig 1c). Critically, each stage processes auditory and visual signals in parallel, as in the standard race model (Equation 1). As a result, the total multisensory RT is given by the sum of two race processes:

$$RT_{AV} = \min(T_{A1}, T_{V1}) + \min(T_{A2}, T_{V2}) \quad (4)$$

At the first gating stage, auditory and visual processes race to initiate the decision stage. The modality that wins this race effectively “passes the baton” to the next stage, which motivates the model’s metaphoric name: the *relay model*. Crucially, initiation at the first gating stage is *non-competitive*. The winner of this first race (e.g., a fast visual signal,  $T_{V1}$ ) does not suppress the other modality but instead initiates the second stage of parallel decision processes for both modalities ( $T_{A2}$  and  $T_{V2}$ ). Because these processes are assumed to be statistically independent, the outcome of the first stage does not deterministically dictate the final response. Due to trial-by-trial variability in evidence accumulation, the initially slower modality can still win the second stage.

A key property of the relay architecture is that it naturally produces RT distributions that can exceed the predictions of standard race models. Because the decision process unfolds across race processes at sequential stages, the effective RT reflects both the faster gating signal and the outcome of a subsequent decision race. Crucially, this structure introduces cross-modal dependencies, which violate the independence assumptions underlying the RMI. As a result, we show below that RMI violations, previously unexplained by the race model, emerge naturally from the relay architecture. In doing so, the model provides a simple and biologically plausible alternative to pooling or coactivation accounts. Beyond reproducing the magnitude of the RSE, we demonstrate that the relay model inherently captures how decision dynamics change with temporal asynchrony or manipulations of signal strength. Thus, by reconciling the milestone contributions by Raab [5] and Miller [4], the model offers a computationally grounded framework that reflects the staged nature of perceptual decisions in multisensory contexts.

## Results

### The relay model explains multisensory processing dynamics

To evaluate the plausibility of the relay model, we tested whether its predictions align with empirical RTs reported in Miller’s benchmark dataset (data estimated from Figure 1 in [4]; see Methods). As a key advancement in understanding multisensory processing, the model assumes that perceptual decision processes comprise latent sequential processing stages. For example, an auditory RT is modelled as the sum of two stages (e.g.,  $RT_A = T_{A1} + T_{A2}$ ).

To constrain the model, we fitted IG distributions to the overall RTs in unisensory auditory and visual conditions (Table 1; see Methods). This distribution is characterised by two parameters: the mean ( $\mu$ ) and the shape parameter ( $\lambda$ ). We estimated these parameters by minimising the root mean squared error (RMSE) between empirical and model CDFs. These unisensory parameters were held fixed to constrain the modelling of the multisensory condition.

To partition the overall RT into stage-specific latent components, we introduce the RT share parameter (Fig 2a). It expresses the contribution of the first stage as a proportion of the overall RT. For instance, if  $RT_A$  is on average 400 ms and the first stage ( $T_{A1}$ ) lasts 80 ms, then  $T_{A1}$  contributes 20% of the total RT. The variability of RTs across trials is partitioned analogously.

We formalised this partitioning approach using the IG distributions fitted to the unisensory conditions. Specifically, we exploit a property of the IG distribution that allows for mathematically consistent partitioning, whereby the overall RT can be split into sequential stages using the RT share parameter, while each latent stage remains IG-distributed. This property preserves analytical tractability, allowing us to model the latent processing stages without resorting to numerical simulations. For simplicity, we used a single RT share parameter for both modalities; however, the model can be extended to support distinct RT shares per modality, which would reflect differences in sensory latencies. The RT share thus expresses the relative contribution of the first of two IG-modelled processing stages, thereby shaping how the system transitions across sensory modalities and stages.

To predict RTs in multisensory conditions, the relay model assumes a non-competitive relay architecture (Fig 1c; Eq 4). During the first gating stage, unisensory processes (e.g.,  $T_{A1}$ ,  $T_{V1}$ ) race to completion. The winner then initiates the second decision stage, triggering a race between the second stage processes (e.g.,  $T_{A2}$ ,  $T_{V2}$ ). The winner of this second race determines the behavioural response. Crucially, this architecture enables cross-modal initiation: for example, a visual stage 1 winner can trigger an auditory stage 2 process. This feature distinguishes the relay model from the classical race model.

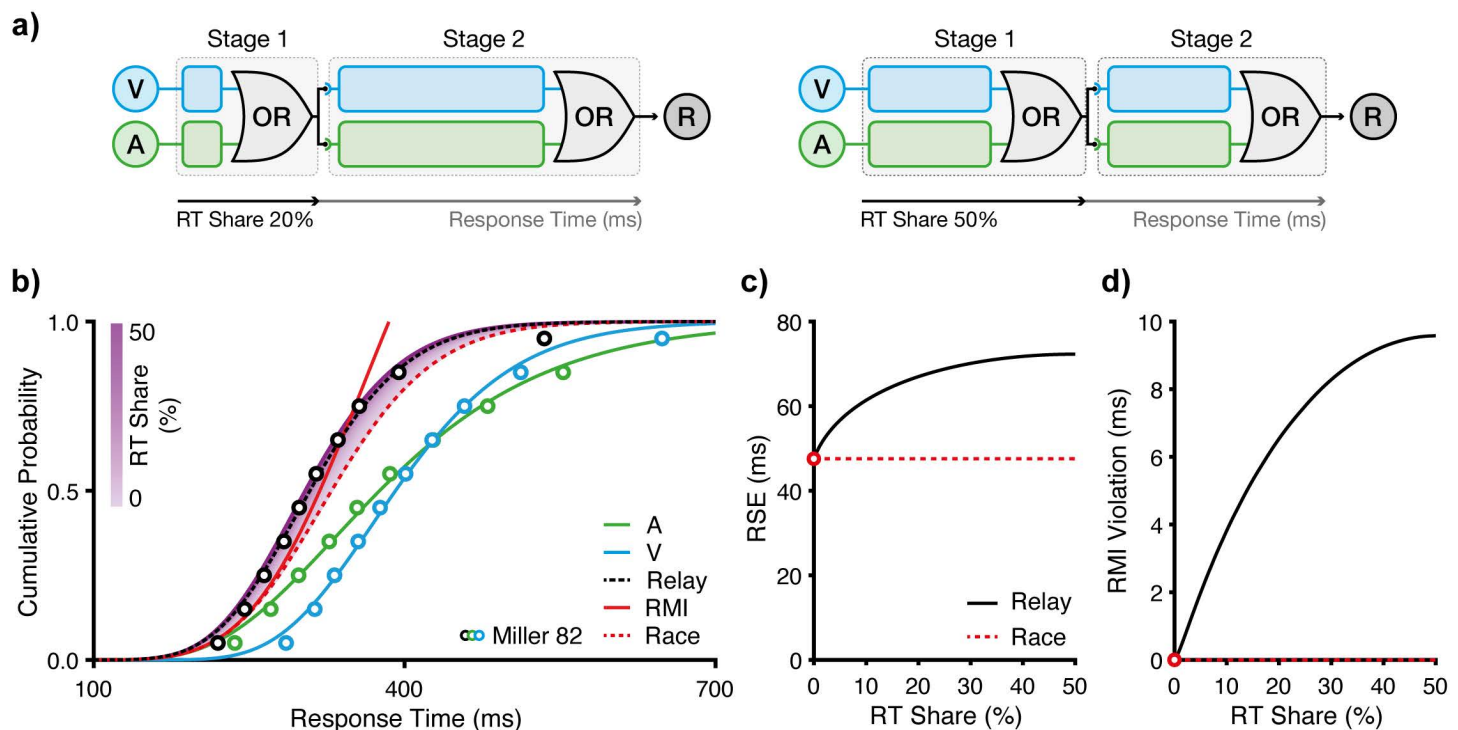
To examine how the RT share affects model behaviour, we simulated the relay model across RT share values from 0% to 50% (Fig 2b). As the two stages add up to 100%, this range encompasses the full dynamic range (values greater than 50% are symmetric). At 0% RT share, the relay model reduces to the classical race model (Fig 1a; Eq 2), which predicts an RSE of 48 ms. In contrast, the relay model predicts an amplification of the RSE up to 72 ms ( $1.5 \times$  the race model) at 50% RT share (Fig 2c). The race model also predicts no violations of the RMI, but with an RT share of 50%, the

**Table 1. Data summary and best-fitting unisensory distribution parameters.**

| Study              | Experiment | N  | Trials (total) | Modality | $\mu$ (ms) | $\lambda$ (ms) |
|--------------------|------------|----|----------------|----------|------------|----------------|
| Miller (1982)      | Exp. 1     | 74 | 1480           | A        | 399        | 3272           |
|                    |            |    |                | V        | 402        | 7309           |
| Miller (1986)      | B.D.       | 1  | 400            | A        | 231        | 3931           |
|                    |            |    |                | V        | 348        | 4979           |
| Otto et al. (2013) | SOA        | 10 | 1600           | A        | 376 (23)   | 8619 (738)     |
|                    |            |    |                | V        | 404 (24)   | 14225 (2138)   |
| Otto et al. (2013) | Intensity  | 20 | 2000           | $A_s$    | 412 (11)   | 13925 (1064)   |
|                    |            |    |                | $A_m$    | 457 (11)   | 11711 (949)    |
|                    |            |    |                | $A_w$    | 508 (11)   | 7888 (595)     |
|                    |            |    |                | $V_s$    | 431 (10)   | 15486 (1450)   |
|                    |            |    |                | $V_m$    | 515 (13)   | 9534 (630)     |

*Note:* Miller (1982) data were fitted to group-average RTs. Otto et al. (2013) data were fitted per participant; values show cross-participant mean (SEM).  $\mu$  and  $\lambda$  are the mean and shape of the IG distribution.

<https://doi.org/10.1371/journal.pcbi.1013320.t001>



**Fig 2. Predictions of the relay model.** **a)** RT share. In the relay model, the RT share denotes the proportion (%) of the overall RT attributed to the first stage: 20% (left) and 50% (right). While these schematic diagrams illustrate this partitioning for a single trial, at the distribution level, the RT share parameter specifies how the entire RT distribution (both its mean and variability) is divided into two sequential stages for each modality. The predicted multisensory RT distribution is then formed by summing the outcomes of the independent races at each stage. Schematic diagrams correspond to the model architecture introduced in Fig 1c. **b)** Modelling based on data from Miller [4]. The relay model is constrained by the RT distribution for auditory (A) and visual (V) signals (Table 1). The shaded region depicts relay model predictions for RT shares ranging from 0% to 50%. The empirical RT distribution for redundant audiovisual (AV) signals is best fit with an RT share of 22.6%. Circles represent the same data shown in Fig 1b. **c)** RSE and **d)** RMI violations as a function of RT share. The relay model predicts RSEs that exceed those of Raab's race model, including systematic RMI violations. Predictions decline symmetrically for RT shares above 50% (e.g., RT shares of 20% and 80% yield identical predictions).

<https://doi.org/10.1371/journal.pcbi.1013320.g002>

relay model predicts RMI violations in the order of 10 ms (Fig 2d). Thus, the RT share, as a single interpretable parameter that implements the relay architecture, enables the model to capture the key effects observed in human multisensory behaviour.

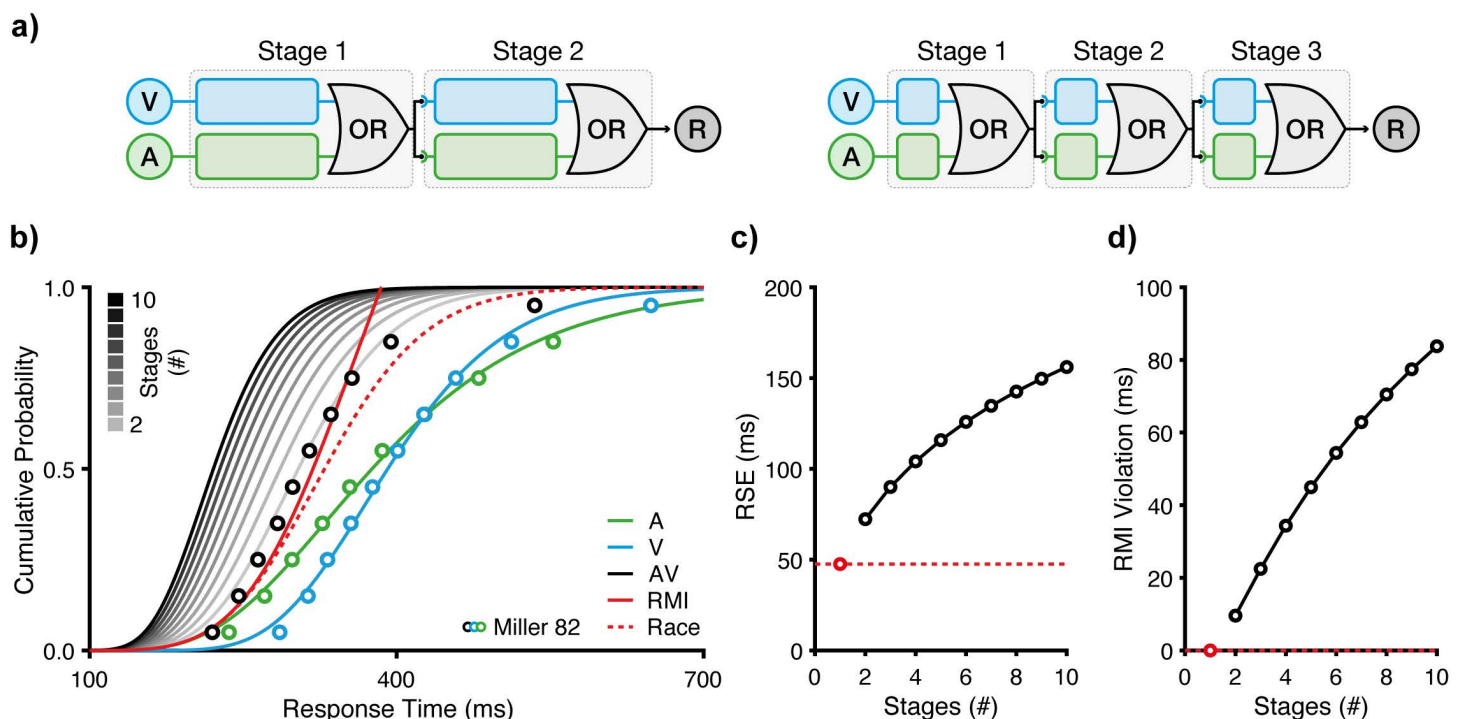
To evaluate the model's fit to empirical data, we estimated the RT share by minimising the RMSE between empirical and predicted multisensory CDFs (Fig 1b). For parsimony, we again constrained auditory and visual RT shares to be equal. The best-fitting RT share was 22.6% (RMSE = 0.023), corresponding to first-stage components of 89.9 ms (SD = 31.4 ms) for audition, and 90.6 ms (SD = 21.3 ms) for vision. To compare the model fit to the empirical data, we computed RSE and RMI violations at the ten quantiles (as opposed to over the whole distribution as in Fig 2c, d). The best-fitting model predicted an RSE of 68 ms (empirical: 67 ms) and RMI violations of 8 ms (empirical: 10 ms). The slight attenuation of the predicted violation is most likely attributable to the constraints of the unisensory fits; specifically, the IG distribution does not perfectly capture the extreme fast tail of the empirical RT distributions, where RMI violations are most prominent (see fastest quantiles in Fig 2b). Yet, overall, the model captures the distributional shape of the AV condition with high accuracy.

To further validate the relay model, we compared its performance against a “shifted” race model, an alternative ad-hoc formulation that can also be used to explain RMI violations with a single free parameter. In this model, the classical race model prediction is simply shifted in time by a free parameter,  $\delta$ , effectively translating the entire distribution to faster RTs. Conceptually, the shift may indicate a non-specific alerting effect, where the presence of multisensory signals elevates baseline arousal and accelerates RTs. The shifted race can account for RMI violations in Miller’s data [4] with a shift of 19.2ms. However, it provides a poorer fit to the dataset (RMSE=0.032) compared to the relay model (RMSE=0.023), which thus shows better explanatory power.

In summary, with only one additional degree of freedom beyond the fits obtained from unisensory conditions, the relay model successfully reproduces a hallmark signature of multisensory integration: larger RSEs than predicted by the race model, including RMI violations. These findings demonstrate that a simple relay architecture offers a parsimonious and powerful framework for modelling the dynamics of multisensory processing.

### Predictions of multi-stage relay models

While our initial analyses focused on a two-stage relay architecture, the model is not inherently restricted to two processing stages. In principle, multisensory decision-making could involve three or more stages, with each stage comprising a race between modalities (Fig 3a). This raises a key question: how many relay stages account for empirical multisensory RTs best?



**Fig 3. Multi-stage relay models.** a) Number of stages. Relay models with two and three sequential stages are shown. Each stage passes a start signal to the next, with the overall RT evenly distributed across stages. The two-stage schematic diagram corresponds to the architecture introduced in Fig 1c. b) Modelling based on data from Miller [4]. Greyscale lines depict predicted cumulative RT distributions for increasing numbers of stages, from two up to ten. Circles represent the same data shown in Fig 1b. c) RSE and d) RMI violations as a function of the number of relay stages. Relay models with three or more stages predict RSEs and RMI violations that far exceed empirical observations. Predictions from Raab’s race model (one stage) are indicated by red circles for reference.

<https://doi.org/10.1371/journal.pcbi.1013320.g003>

To address this, we simulated relay architectures with one to ten stages, using the same unisensory IG fits to constrain each model (Fig 3b). In every simulation, the total RT was evenly divided across  $N$  stages, such that each stage was allocated  $1/N$  of the total RT. Accordingly, the RT share was held fixed at this proportion (i.e., not optimised). This design enabled us to assess how increasing the stage count affects predicted RSEs and RMI violations, while holding model complexity and parameterisation constant.

Simulation results show that both predicted RSEs and RMI violations increase with stage count, but with diminishing returns (Fig 3c,d). The most substantial gain occurs between the race model and the two-stage relay model. Adding more stages yielded progressively smaller gains. Crucially, higher stage counts led to increasingly exaggerated predictions. For example, the ten-stage model predicted an RSE of 156 ms and RMI violations of 84 ms, which far exceed empirically observed values.

To evaluate these architectures quantitatively, we compared predicted to empirical multisensory CDFs (Fig 3b). Because the RT share was fixed (rather than optimised) and all models shared identical unisensory parameters, this provided a fair, complexity-controlled comparison. The two-stage relay model emerged as the best configuration, yielding the lowest RMSE (RMSE = 0.030), more than twice as accurate as the one-stage race model (RMSE = 0.085). Adding further stages degraded model performance (e.g., three-stage RMSE = 0.102; ten-stage RMSE = 0.399), consistent with their overestimation of RSE and RMI effects.

Crucially, the assumption of evenly distributed stages is not a strict requirement of the relay architecture. In a relay with more than two stages, the RT shares can be asymmetric. Indeed, perceptual decision-making can be thought of as asymmetric stages of sensory encoding, evidence accumulation, and motor execution, where the vast majority of RT is dedicated to the accumulation of evidence [32,48,60]. For instance, an alternative three-stage relay configuration could consist of a brief initial stage (10%), a longer middle stage (80%), and a brief final stage (10%). This 10-80-10 distribution yields predictions much closer to the two-stage model (RMSE = 0.039). Therefore, while a two-stage relay is sufficient to capture the current data, there are likely situations where asymmetric, multi-stage relays can be modelled successfully.

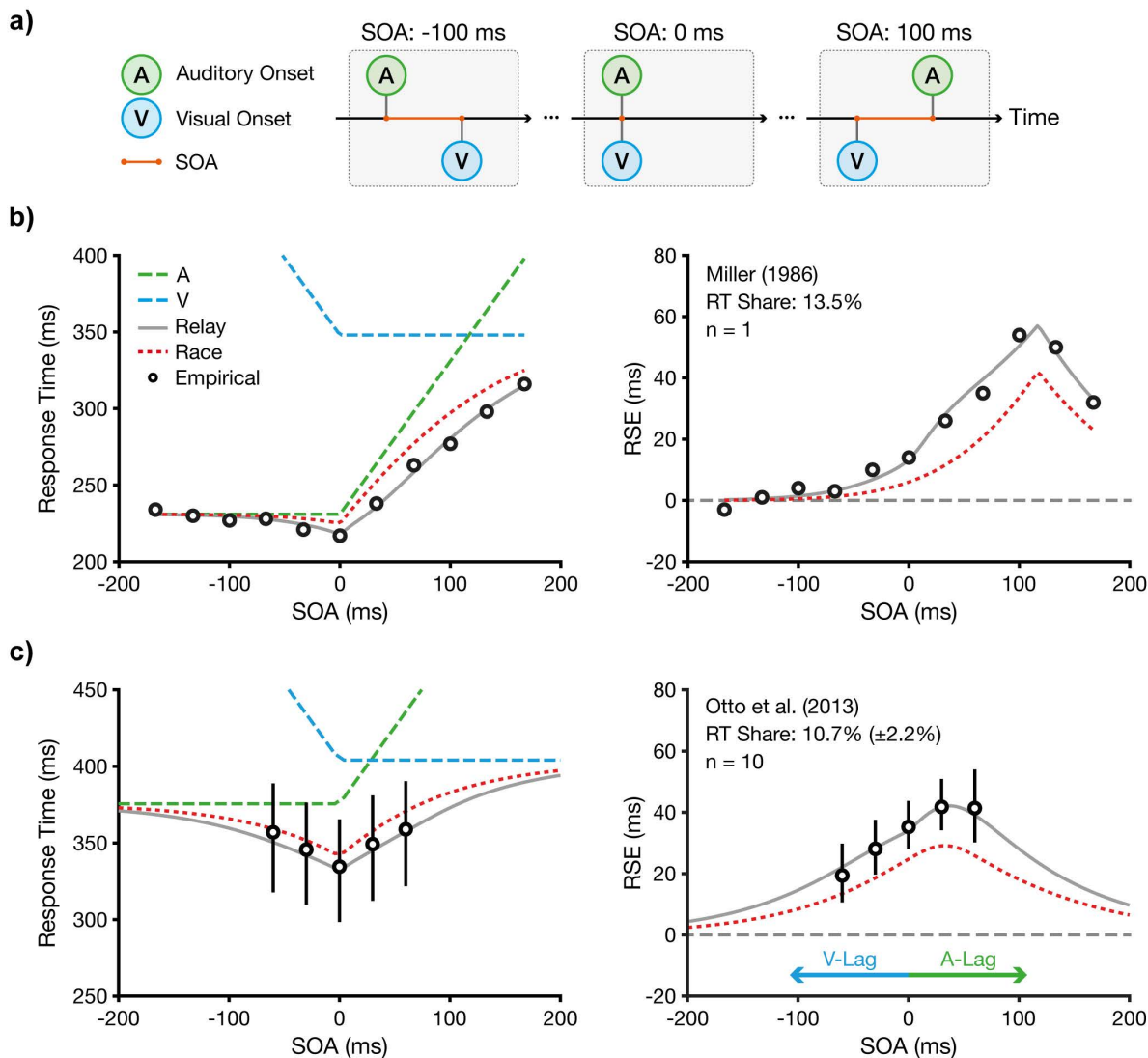
In summary, while the relay architecture can, in principle, support an arbitrary number of stages, the two-stage configuration strikes the optimal balance. It captures observed multisensory enhancements without inflating predictions beyond empirical observations, though unequal multi-stage models remain a viable architecture.

## Temporal dynamics of multisensory decisions

Having established that the two-stage relay model captures both RSE and RMI violations under standard conditions, we next examined its ability to capture multisensory behaviour under conditions of temporal asynchrony. Specifically, we simulated SOA conditions, in which one sensory signal precedes the other (Fig 4a). As a hallmark manipulation in multisensory research, SOA manipulations provide a stringent test of multisensory models by probing how response dynamics adapt when stimuli are no longer presented simultaneously [e.g., 38,61,62].

We used the benchmark dataset from Miller [38], which systematically varied SOAs between auditory and visual signals from -167 ms (auditory lead) to +167 ms (visual lead) in 33 ms increments (Fig 4b). This dataset exhibits two characteristic features. First, because any deviation from synchrony introduces a delay, the fastest multisensory responses occur under synchronous conditions (SOA = 0 ms). Second, the largest RSE is observed when the SOA compensates for the difference in unisensory RTs, a phenomenon introduced earlier as the principle of congruent effectiveness [37]. In Miller's data, visual RTs were ~100 ms slower than auditory RTs. Accordingly, the largest RSE was observed when the auditory signal lagged by ~100 ms. While the classical race model captures these key characteristics, it systematically underestimates RSE magnitudes across SOA conditions.

To accommodate temporal asynchrony in the relay model, we implemented a simple model extension. Since evidence accumulation cannot begin prior to stimulus onset, we incorporated a temporal delay into the processing of the lagging signal, equal to the experimentally defined SOA. Thus, at stage 1, processing for each modality begins with its signal



**Fig 4. Temporal dynamics in multisensory processing.** **a)** Stimulus onset synchrony (SOA). Negative SOAs represent visual lags relative to auditory signal onset (left). An SOA of 0 ms represents synchrony between the auditory and visual onsets (centre). Positive SOAs represent auditory lags relative to visual onset (right). **b)** Modelling based on data from Miller [38] ( $n = 1$ ). Circles show mean RTs with redundant signals (left) and the RSE (right) as a function of SOA. Dashed lines represent unisensory RTs, adjusted for SOA (e.g., the SOA is added to auditory RTs for positive SOAs). Red lines show predictions from Raab's race model and grey lines the relay model, which fitted the empirical RTs best with an RT share of 13.5%. **c)** Identical analysis and modelling based on data from Otto et al. [37] ( $n = 10$ ). The relay model fitted the empirical RTs best with a mean RT share of  $10.7 \pm 2.2\%$  (model fitting was performed at the participant level). Data points show group averages; error bars represent bootstrapped 95% confidence intervals.

<https://doi.org/10.1371/journal.pcbi.1013320.g004>

onset. For example, if the auditory signal is delayed by  $\tau$ , the stage 1 completion time becomes  $T_1 = \min(T_{A1} + \tau, T_{V1})$ . Critically, we retained the relay model's core feature of cross-modal initiation: stage 2 processing is initiated by the first modality to complete stage 1. If stage 2 is initiated before the delayed signal is physically present, processing of that modality is further delayed until its onset. In the model, we account for this by adding a corresponding residual delay at stage 2. This mechanism becomes increasingly important at larger SOAs, where the lagging modality may be partially or

fully excluded from processing. When the SOA exceeds the combined duration of both stages, the model response effectively reduces to the unisensory condition of the leading signal.

We tested whether this model extension could fully reproduce the empirical data. First, we constrained the model by fitting its unisensory parameters (Table 1). With these parameters fixed, we optimised the model's single free parameter, the RT share, by minimising the RMSE between predicted and observed RSEs across SOA levels. The best-fitting RT share was 13.5% (RMSE = 2.5 ms), corresponding to first-stage durations of 31.2 ms (SD = 7.6 ms) for auditory, and 47.0 ms (SD = 12.4 ms) for vision. Overall, the best-fitting model produced an excellent fit to all empirical data points.

As it is difficult to generalise model performance from a single participant, we applied the same analysis to a second dataset from Otto et al. [37], which employed a redundant signal paradigm with 10 participants (Fig 4c). Here, the SOA varied from -60 ms (auditory lead) to +60 ms (visual lead) in 30 ms increments. Consistent with the previous analysis, the fastest multisensory RTs occurred under the synchronous condition. Moreover, the largest RSE was observed when the SOA compensated for the difference in unisensory RTs. Here, visual RTs were on average ~30 ms slower than auditory RTs. Accordingly, the largest RSE occurred for an SOA of 30 ms. Again, the race model predicted these temporal dynamics but failed to capture the magnitude of the empirical RSE.

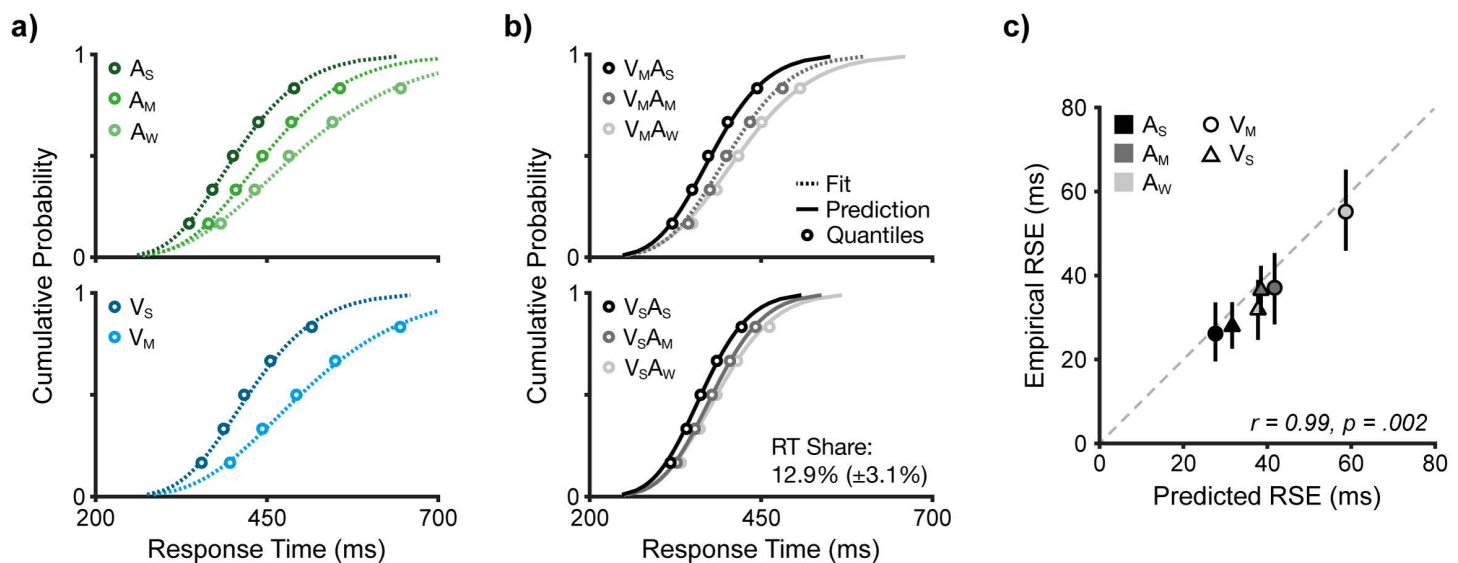
We employed the same model analysis as above, with all fitting performed using RT distributions at the participant level. As a first step, we constrained the model with the IG parameters estimated in unisensory conditions using Quantile Maximum Probability Estimation (QMPE; [63,64]; Table 1). We then fitted the relay model with its single free parameter, RT share, to the RT quantiles across the six multisensory SOA conditions. The mean estimated RT share was  $10.7 \pm 2.2\%$ , with a mean Negative Log-Likelihood (NLL) of 368.1. A sign-flipping permutation test (10,000 iterations) confirmed that this RT share was significantly different from zero ( $p = 0.002$ ). Crucially, with only this single degree of freedom at the multisensory level, the model predicted the observed RSE across all SOA conditions, always falling within the bootstrapped 95% confidence intervals.

Thus, the relay model successfully captured both qualitative and quantitative features of the datasets from Miller [38] and Otto et al. [37]. It predicted the fastest RTs at synchronous onset, identified the SOA that maximised the RSE, and closely matched the RSE magnitudes across all SOA conditions (Fig 4b,c). By incorporating a two-stage architecture that allows for cross-modal initiation, the relay model offers a mechanistically grounded account of how temporal asynchrony shapes multisensory decision-making, thereby seamlessly transitioning between unisensory and multisensory conditions.

## Manipulation of signal strength

To further evaluate the relay model, we analysed data from a redundant signals paradigm with varying signal strength [37]. Such manipulations are established tools known to systematically modulate both the mean and the variability of unisensory RTs [65]. In general, stronger signals yield faster and more consistent RTs, whereas weaker signals produce slower and more variable responses. Critically, race models predict that signal strength manipulations have a similar effect on the RSE as SOA manipulations, because unisensory RTs need to be equally effective to maximise the RSE (principle of congruent effectiveness). In addition, such manipulations are also expected to affect the RSE because, according to race models, it is the variability of unisensory RTs that drives the effect (variability rule). Consequently, it is not surprising that signal strength manipulations also have strong empirical effects on the RSE [37,61,66], making them a stringent test of the model's predictive power.

In the dataset by Otto et al. [37], signal strength was manipulated in both modalities (Fig 5a). We analysed three auditory levels (weak, medium, strong;  $A_w$ ,  $A_m$ ,  $A_s$ ) and two visual levels (medium, strong;  $V_m$ ,  $V_s$ ). The dataset also included a third, weak visual signal, however, performance in this condition was not at ceiling and thus cannot be used in the modelling approach employed here, which assumes ceiling performance (see Methods). The signal strength manipulations produced substantial effects on unisensory RTs, with changes in the mean up to ~100 ms and in the variability up to a factor of two. Using a model fitting strategy at the participant level, we estimated the unisensory parameters using QMPE



**Fig 5. Signal strength manipulation.** a) Unisensory RT distributions. Signal strength systematically modulated unisensory RTs, with strong effects on both mean and variability. Empirical data are represented by quantiles (circles) and IG fits (dashed lines; for best-fitting parameter values, see Table 1). b) Multisensory RT distributions. Signal strength also modulated multisensory RTs. The relay model was fitted in the condition with both modalities at the medium level (dashed line). The best-fitting estimate of the RT share parameter was  $12.9\% \pm 3.1\%$ . Keeping this parameter value fixed, the relay model predicted the RT distributions for the remaining five combinations of signal strength (solid lines). c) RSE across signal strength combinations. A near-perfect correspondence was found between predicted and empirical RSEs. Data taken from [37].

<https://doi.org/10.1371/journal.pcbi.1013320.g005>

[63,64]. For each signal strength condition, we fitted the two unisensory parameters, yielding a total of ten parameters across five RT distributions (Table 1). As in the approaches above, these parameters were held fixed and used to constrain the modelling of the multisensory conditions.

The unisensory conditions were combined using a  $3 \times 2$  factorial design, yielding six multisensory conditions ranging from the weakest ( $A_W V_M$ ) to the strongest combination ( $A_S V_S$ ; Fig 5b). In a first step, we fitted the relay model with a single free parameter (RT share) to the RT quantiles of the multisensory condition with both modalities at the medium level ( $A_M V_M$ ). The mean estimated RT share was  $12.9\% \pm 3.1\%$  (mean NLL = 93.4), which was significantly different from zero (permutation test,  $p < .001$ ). In a second step, we generated model predictions for the remaining five multisensory conditions. Crucially, these predictions were derived with no additional free parameters, relying solely on the best-fitting RT share from the medium condition ( $A_M V_M$ ) and the IG parameters in the underlying unisensory conditions (Table 1). The model predicted all empirical distributions with high fidelity.

To summarise this model analysis, we examined the modulation of the RSE for the different signal strength combinations (Fig 5c). In line with the strong modulation of unisensory RTs, the empirical RSE was significantly modulated by approximately a factor of two, ranging from  $28 \pm 3$  ms ( $A_S V_S$ ) to  $55 \pm 5$  ms ( $A_W V_M$ ; permutation test,  $p < .001$ ). This modulation was accurately predicted by the relay model (as fitted above). A correlation analysis revealed a near-perfect correspondence between the empirical and predicted RSEs ( $r = .99$ , permutation test,  $p = .002$ ; Fig 5c). This result demonstrates that a single, invariant relay mechanism is sufficient to explain the complex pattern of facilitation across varying levels of signal strength without requiring additional parameter adjustments, highlighting its strong explanatory power across a second key experimental manipulation.

## Discussion

The robust speed-up of responses observed with multi- compared to unisensory signals provides a critical window into the computational strategies underlying perceptual decision-making. The RSE not only reflects a prominent performance advantage but also places strong constraints on models that explain how the brain combines information across modalities. As one of the earliest and most influential frameworks, the race model [5] proposes that responses are triggered by the faster unisensory signal to reach its decision threshold, assuming that modalities operate independently and in parallel (Fig 1a). While this model accounts for the RSE through statistical facilitation, it typically underestimates the observed behavioural enhancement. Most notably, it imposes the RMI as an upper bound on the RSE, which is frequently violated empirically (as demonstrated in numerous follow-ups to [4]; Fig 1b). These systematic violations strongly argue against purely parallel architectures and have led many to favour explanations in which multisensory inputs are pooled or integrated into a single decision process (e.g., [67–70]).

To reconcile the conceptual parsimony of the race model with RMI violations, we introduced the relay model, a minimal extension that includes two races at sequential stages (Fig 1c). As a key feature, the relay model allows for cross-modal initiation in addition to statistical facilitation. In its most basic form, the model includes a single free parameter, the RT share, which partitions the overall RT between the processing stages. Our simulations and fits to benchmark data [4] demonstrate that the relay model captures the magnitude of the RSE, including RMI violations, and closely approximates the observed RT distribution (Fig 2). Moreover, we show that a two-stage architecture is sufficient to explain the RSE (Fig 3). We further show that the model captures decision dynamics under manipulations of temporal structure and signal intensity [37,38]. Specifically, the model captures the modulation of the RSE under asynchronous signal presentation (Fig 4) and under manipulations of signal strength (Fig 5). Together, these findings indicate that a two-stage relay configuration is both computationally parsimonious and descriptively powerful, providing an efficient and accurate framework for modelling multisensory decision-making.

Beyond statistical facilitation proposed by the classical race model [5], cross-modal initiation at stage 1 is the main new proposal of the relay model. While the relay model does not necessarily commit to a specific cognitive mechanism underlying the two stages, we propose that modality-specific signals at an initial gating stage race to initiate a second decision stage, which in turn triggers a behavioural response. The concept of gated accumulation is well-established in perceptual decision-making with unisensory signals, where gating selects relevant sensory input and determines the onset of evidence accumulation [e.g., 58,71,72]. Our work implies that the dynamics at the gating stage are not constant but are instead modulated by multisensory stimulation. As our model provides a clear handle on this stage, it may offer a valuable framework for investigating gating mechanisms and decision dynamics more generally.

While our work is currently limited to the modelling of behavioural data, the relay model framework can be linked to physiological evidence. For example, early event-related potential (ERP) components have been shown to predict the initiation of evidence accumulation [73,74], which could correspond to the initiation of the decision stage in the relay model. In multisensory studies, modulation of ERPs has been observed as early as 40–80 ms after stimulus onset [8,75,76], closely matching the first-stage durations predicted by the relay model (30–90 ms across the behavioural datasets analysed here [4,37,38]). Interestingly, some evidence suggests that early multisensory integration effects predominantly enhance neural activity in the non-dominant sensory modality [8]. Given auditory dominance (i.e., faster and more accurate responses), the relay model predicts that the auditory process would win the first race more frequently, allowing cross-modal initiation to benefit visual processing to a greater extent. Conversely, under visual dominance, the opposite pattern would be expected. These ERP markers therefore provide plausible neural correlates for the first model stage, linking the relay architecture to measurable brain dynamics. Future work could test this interpretation more directly through joint modelling of behaviour and neural recordings, potentially disentangling the first and second stages temporally and across sensory modalities.

As a technical note on the mathematical implementation of the relay model, we used a single RT share parameter for both modalities, partitioning processing times under the assumption that all components are IG distributed. Importantly,

the model's ability to account for the data and generate RMI violations does not depend on this specific choice. Because RMI violations are inherently driven by the mechanism of cross-modal initiation, they emerge under a variety of statistical assumptions. For example, a relay architecture with an exponential first stage and a Gaussian second stage also generates RMI violations. This flexibility extends to the model's structural parameters: the assumption of a symmetrical RT share can be easily relaxed. In the presence of temporal asymmetries in peripheral sensory processing, the model can employ separate RT share parameters for each modality. While this would allow for more complex dynamics, our current simulations demonstrate that even a strict, single-parameter baseline is sufficient to capture the main decision dynamics underlying the RSE.

Conceptually, the relay model builds on the classical race model [5], with statistical facilitation as the primary driver of the RSE, but it is not the first framework to propose additional cross-modal interactions. For example, Mordkoff and Yantis [41] introduced an interactive architecture in which crosstalk occurs between parallel processing channels. Similarly, Nickerson [77] attributed facilitatory effects to enhanced response preparedness, whereby a fast auditory signal primes the system to respond more rapidly to a coincident visual stimulus. These frameworks provide important conceptual precedents for extending the classical race model and incorporating separate processing stages in models of multisensory decision-making.

Building on these foundations, the time-window-of-integration (TWIN) model provides an influential contemporary two-stage architecture [28,42]. The model accounts for multisensory facilitation by positing a peripheral sensory processing stage followed by a central pooling stage. While both the TWIN and the relay model share a race process at the initial stage, they differ significantly in their parameterisation and architectures, particularly at the second stage. To capture its complex processing dynamics, the TWIN model utilises a five-parameter architecture, with one free parameter specifically determining the magnitude of the RSE. Given its specific distributional assumptions, model fitting for the TWIN framework is typically performed at the level of mean RTs [78]. In contrast, the relay model maintains the classical race architecture at both stages, explaining the magnitude of the RSE largely through statistical facilitation and cross-modal initiation, both inherent in its structure. Furthermore, the relay model leverages a highly simplified architecture, using the RT share as a single free parameter beyond the fits to unisensory RT distributions, which allows it to accurately predict not only summary statistics but also the full shape of empirical RT distributions.

Furthermore, the relay model provides a mechanistically transparent alternative to other race-based accounts that add cross-modal interaction terms in addition to statistical facilitation as the main mechanism. For example, we included an ad hoc formulation of a shifted race model in the analysis of Miller's benchmark data. While the model performed only slightly worse under synchronous conditions than the relay model, a key distinction emerges when considering SOA data, which requires the model to transition between unisensory and multisensory processing. In this case, the shifted race model fails without introducing additional free parameters. Another interactive race model was proposed by Otto and Mamassian [11], who introduced additive noise between parallel processes. While the addition of noise permits RMI violations, it is insufficient to reproduce the full shape of empirical RT distributions, requiring an additional correlation parameter to capture trial-by-trial dependencies [11]. Furthermore, because these parameters must be explicitly activated during multisensory conditions, the model lacks the intrinsic ability to transition between unisensory and multisensory regimes. In contrast, the relay model seamlessly transitions between regimes based solely on the temporal lag introduced by SOA manipulation: as the SOA increases, the likelihood of cross-modal initiation naturally diminishes, effectively reducing facilitation without requiring parameter adjustments.

The relay model can be easily adapted to other established manipulations of the redundant signals paradigm. For example, it may also be applied to redundant targets within the same sensory modality [15–19,79,80]. As the initiation stage relies on closely related processing pathways, cross-modal initiation is here expected to produce smaller facilitation effects. This prediction is consistent with the typically reduced RMI violations observed in such experiments. The relay model can also be naturally extended to experiments employing three redundant signals [9,20,61,81,82]. The typical

finding is that that RSE for bimodal combinations is very similar to the experiments described in this manuscript; however, the benefit from adding a third signal is relatively smaller (both in terms of RSE and RMI violations). The relay model can be straightforwardly extended to three parallel processes at each stage. As the incremental benefit from statistical facilitation is reduced for additional racers, it directly predicts the observed empirical findings. The relay model therefore generalises naturally across task variants, providing new avenues for related research.

Beyond the redundant signals paradigm, the relay architecture also extends to focused attention tasks. In such tasks, participants are asked to respond to a signal in one modality, which in some trials is paired with a distractor signal in another modality. Although the task requires inhibiting the non-target modality, RTs are paradoxically faster when a distractor is presented alongside the target. The relay model can explain this finding by assuming that the initial gating stage is automatic and non-specific, such that both target and non-target signals are processed in parallel, racing to initiate the second stage and thereby enabling cross-modal initiation. The second stage, however, includes only a single processing pathway for the target signal, in accordance with the task demands. This mechanism produces a facilitation effect without generating erroneous responses to the distractor, aligning with foundational attention theories [e.g., 83] and computational frameworks like the Dual-Stage Two-Phase model [57].

One limitation of the current relay model is that it does not account for trial history effects, which have been shown to influence RTs in RSE experiments [11,39,84–89]. Unisensory responses are particularly sensitive to modality switches, often exhibiting RT costs following a switch that can rival the magnitude of the RSE itself. In contrast, multisensory responses tend to be less affected by the recent trial history (e.g., [84,85]), leading to an asymmetry that can modulate the RSE and exaggerate RMI violations. As such, any comprehensive model of multisensory integration should consider these sequential effects. In unisensory decision-making, recent models have begun to incorporate such dependencies. For instance, Caie et al. [72] introduced a gated accumulation model in which trial history influences the gating dynamics of sensory evidence accumulation. In future work, we propose extending the relay model in a similar direction by embedding history-dependent mechanisms into its initial stage. This extension would enable the model to capture trial history effects in unisensory conditions and generate testable predictions for multisensory responses without requiring additional parameter fitting.

In conclusion, the relay model provides a principled and computationally grounded framework for understanding the temporal dynamics of multisensory decision-making. By simply extending the classic race model to incorporate two races at sequential processing stages, it accounts for the full magnitude of the RSE, including RMI violations, while remaining mechanistically transparent. Its modular design allows for integration with broader decision-theoretic models, including those that incorporate gating, evidence accumulation, and trial history effects. As such, the relay model provides a new tool for linking behavioural findings to underlying neural and cognitive processes in multisensory integration.

## Methods

### Ethics Statement

The analysis carried out in this study involves model fitting of non-contentious anonymous secondary human data and did not involve any new participant data collection (see Data). Ethical approval for data reuse was given by the University Teaching and Research Ethics Committee (UTREC) at the University of St Andrews (0740 - PS-0740-852-2025).

### Data

The data presented in Fig 1b, along with the modelling results in Figs 2 and 3, are based on Experiment 1 from Miller [4]. In this experiment, 74 participants completed a redundant signals task involving three stimulus conditions: a visual stimulus (an asterisk at the centre of the screen), an auditory stimulus (a 780-Hz tone), or both stimuli simultaneously. Participants were instructed to press a key as quickly as possible upon detecting any stimulus. Each

participant completed two blocks of 40 trials, with each block containing 10 trials per stimulus condition as well as 10 catch trials (no stimulus). For analysis, RTs for each condition within each block were rank-ordered and then averaged across blocks and participants [4,90]. This yielded three cumulative group RT distributions (auditory, visual, and audiovisual), each based on 1480 trials. As raw data are not available from the original publication, we extracted group-level data using graphical digitisation (Figure 1 in [4]). The figure was digitised, and the pixel-to-millisecond scaling was calibrated using a known 100 ms interval. RT values were then read directly from the scaled x-axis coordinates of the plotted points.

The data underlying Fig 4b are taken from Miller's follow-up study [38]. Two participants completed the same redundant signals task as described above. Crucially, this study introduced temporal asynchronies between the auditory and visual stimuli. Audiovisual trials were divided into 11 conditions: a synchronous condition, five with the visual stimulus leading (SOAs from -33 to -167 ms), and five with the auditory stimulus leading (SOAs from 33 to 167 ms), all in 33 ms increments. Each participant completed 40 blocks of trials. Each block contained 170 trials, including 10 trials per stimulus conditions (13 conditions: auditory, visual, and 11 audio-visual SOAs) and 40 catch trials (no stimulus). We used data only from participant B.D., whose responses showed significant RSEs in 7 of the 11 SOA conditions. By contrast, participant K.Y.'s data showed significant RSEs in only three conditions, a finding that Miller described as "somewhat atypical in view of the consistency with which RSEs are obtained" [38]. Summary statistics were reported in the original publication for each condition, which included the mean, standard error, and median of RTs, as well as the magnitude of the RSE (Table 1 in [38]). The analysed data comprised 400 trials per condition.

The data underlying Fig 4c are based on Experiment 1 from Otto et al. [37]. Ten participants performed a redundant signals task involving a visual stimulus (random dot motion), an auditory stimulus (a 294-Hz tone), or both stimuli. For redundant conditions, the experiment featured five SOA conditions ranging in increments of 30 ms from -60 ms to +60 ms. The experiment was conducted in four blocks per participant. Following the original protocol, we removed outliers and retained 40 valid trials per block and condition. To combine data across the four blocks, we used Vincent averaging [90]. While Otto et al. [37] originally analysed group-level data, we analysed the data at the participant level. This approach provides a stricter test of the relay model, as it requires the model to capture the RT distributions of individual participants rather than a smoothed aggregate. Overall, the dataset comprised 1600 trials per condition.

The data underlying Fig 5 are based on Experiment 2 from Otto et al. [37]. Twenty participants performed a similar redundant signals task as described above. Here, the experiment manipulated auditory and visual signal strength in a 3x3 design. Auditory signals were weak ( $A_w$ ; 49.5 dB), medium ( $A_m$ ; 50.5 dB), and strong ( $A_s$ ; 53 dB). Visual signals varied in motion coherence at levels weak ( $V_w$ ; 20%), medium ( $V_m$ ; 30%), and strong ( $V_s$ ; 50%). For redundant conditions, the experiment tested all nine multisensory combinations (e.g.,  $A_w V_s$  or  $A_s V_m$ ). Standard tests of the race model as well as our modelling approach require near-ceiling performance, implying that signals are detected on every trial. As the weak visual condition ( $V_w$ ) exhibited a high miss rate (>20%), we did not include conditions involving this level in our analysis. Data collection occurred over two sessions on separate days. As the dataset did not permit linking participant data across days, we followed the original protocol and treated each block as an independent sample. For each block, we removed outliers and retained the last 50 valid trials in each condition. Overall, the dataset comprised 2000 trials per condition. Data underlying all analyses can be accessed from [91].

## Unisensory RT distributions

Race models allow for parameter-free predictions of the RSE directly from the observed unisensory RT distributions. Basic predictions, such as those from Raab's race model (probability summation, Eq 2) or the RMI (Eq 3), can be computed from empirical data without distributional assumptions (e.g., [92,93]). However, generating more complex predictions, such as those required by the relay model, requires assuming a particular RT distributional form.

For all simulations, we modelled unisensory RTs using the IG distribution (also known as Wald distribution; [94]). The IG distribution is well-suited for modelling non-negative, positively skewed data such as RTs [34,35]. Consequently, we assumed that unisensory RTs are IG-distributed with mean  $\mu$  and shape parameter  $\lambda$ :

$$RT \sim IG(\mu, \lambda) \quad (5)$$

For the group-level RT distributions extracted from [4], we estimated the IG parameters for each unisensory condition separately:  $(\mu_A, \lambda_A)$  for the auditory condition, and  $(\mu_V, \lambda_V)$  for the visual condition. Parameters were optimised using MATLAB's *fmincon* function, minimising the RMSE between the empirical and the simulated distributions at the 10 available quantiles.

For the SOA data extracted from [38], we estimated  $(\mu_A, \lambda_A)$  and  $(\mu_V, \lambda_V)$ , using the method of moments approach [95]. Here, the fitting procedure minimised the RMSE between the observed and predicted descriptive statistics (mean and standard deviation). Although [38] reported standard errors, we converted these to standard deviations (using  $n=400$ ) to directly fit the standard deviation of the simulated distribution. The analytic mean was directly computed from  $\mu$  and the standard deviation from  $\sqrt{\mu^3/\lambda}$ .

For both the SOA and the signal strength datasets from Otto et al. [37], unisensory parameter estimation was performed at the individual participant level using Quantile Maximum Probability Estimation (QMPE; [63,64]). For each condition, we computed five quantiles from the raw RT distributions, defining six bins. Parameters were optimised by minimising the negative log-likelihood (NLL) of the observed bin counts using MATLAB's *fmincon* function. For the SOA dataset [37], IG parameters  $(\mu, \lambda)$  were fitted independently for auditory and visual conditions, resulting in four unisensory parameters per participant. For the signal strength dataset [37], IG parameters were fitted independently for each intensity level (five conditions:  $A_S, A_M, A_W, V_S$ , and  $V_M$ ), resulting in ten unisensory parameters per participant.

Table 1 reports the best-fitting IG parameters for all unisensory conditions. These parameters were held fixed and used to constrain all subsequent modelling of multisensory conditions, including race model predictions and relay model simulations.

### Stage-based partitioning of unisensory RTs

The relay model assumes that unisensory RTs are composed of two (or more) sequential processing stages (Fig 1c). Accordingly, the total RT can be expressed as

$$RT = \sum_{i=1}^n T_i \quad (6)$$

where  $i$  is the index of  $n$  stages.

To estimate the contribution of these (latent) stages, we implemented an approach that partitions the overall RT into stage-specific processing times. This approach relies on a key property of inverse Gaussian (IG) distributions: under specific conditions, the sum of IG-distributed random variables is itself IG-distributed. Specifically, if

$$T_i \sim IG(\mu w_i, \lambda w_i^2) \quad \text{with } w_i > 0, \quad (7)$$

describes the processing time at (latent) stages, then the total unisensory RT remains IG-distributed with mean  $\mu$  and shape parameter  $\lambda$  (Eq 5). This property arises because the IG parameters are scaled with the weights  $w_i$ , such that the convolution preserves the original distributional form. While typically used as a summation rule, we exploit this property in reverse: it allows us to decompose a single, empirically fitted IG distribution into (latent) constituent stages. As a result, the unisensory

RT distributions, which are overall described by IG distributions (Table 1), can equivalently be expressed as the sum of two or more (latent) IG-distributed processing stages, with the mean and variance partitioned via the weights  $w_i$ .

For model fitting and simulation purposes in the multisensory context, we imposed two constraints on the partitioning approach. First, we used normalised weights  $w_i$  that sum to one:

$$\sum_{i=1}^n w_i = 1 \quad (8)$$

Thus, weights  $w_i$  can be interpreted as the proportion of the total RT contributed by a specific stage (in terms of both mean and variance), which we refer to as the RT share (Fig 2a). For example,  $w_1 = 0.2$  indicates that stage 1 accounts for 20% of the total RT. Stage 2 would then account for the remaining 80%.

Second, for simplicity, we assumed that the same RT share configuration applies to both audition and vision. For example, if the RT share at the first auditory stage is 20%, we assumed the same RT share for the corresponding visual stage. This simplification facilitates model fitting and interpretation but can be relaxed in future experiments aimed at isolating modality-specific stage dynamics.

### Relay model

The basic relay model consists of two sequential race units (Fig 1c). Using the partitioning approach established above, for stage  $i$ , let the auditory and visual processing times be distributed as:

$$T_{Ai} \sim IG(\mu_A w_i, \lambda_A w_i^2), \quad T_{Vi} \sim IG(\mu_V w_i, \lambda_V w_i^2) \quad (9)$$

where  $w_i$  represents the RT share of stage  $i$ . Because this partitioning preserves the IG form, the corresponding probability density functions (PDFs) and cumulative distribution functions (CDFs) can be expressed analytically. We denote the PDFs as  $f_{Ai}(t)$  and  $f_{Vi}(t)$ , and the CDFs as  $F_{Ai}(t)$  and  $F_{Vi}(t)$ , respectively.

In the multisensory condition, stage  $i$  is completed as soon as either modality finishes processing. The duration of stage  $i$  is therefore given by:

$$T_{AVi} = \min(T_{Ai}, T_{Vi}) \quad (10)$$

The corresponding CDF, denoted  $F_{AVi}(t)$ , can be computed analogously to Raab's race model (Eq 2), assuming statistical independence between auditory and visual processes within each stage. It can be analogously expressed as

$$F_{AVi}(t) = 1 - S_{Ai}(t) \times S_{Vi}(t) \quad (11)$$

where  $S_{Ai}(t) = 1 - F_{Ai}(t)$  denotes the survival function. The corresponding PDF, denoted  $f_{AVi}(t)$ , is given by:

$$f_{AVi}(t) = f_{Ai}(t)S_{Vi}(t) + f_{Vi}(t)S_{Ai}(t) \quad (12)$$

As the relay model assumes cross-modal initiation, the total multisensory  $RT_{AV}$  is given by the sum of the durations of the sequential stages (Eq 4). For analytical tractability, we assume independence between stages, such that the durations of successive stages are treated as independent random variables. To compute the CDF of  $RT_{AV}$ , we convolved the CDF of the first stage ( $F_{AV1}$ ) with the PDF of the second stage ( $f_{AV2}$ ):

$$F_{AV}(t) = \int_0^{\infty} F_{AV1}(t-x)f_{AV2}(x)dx \quad (13)$$

For simulations shown in Fig 3, we extended the relay model to more than two stages. The overall CDF for a multi-stage relay model was computed iteratively: starting with the CDF of the first stage, we successively convolved it with the PDFs of the additional stages. Each iteration follows the same computation as in Eq 13.

To provide a simple alternative to the relay model, we also fitted a shifted race model, a variant of the race model in which the predicted RT is shifted by a constant  $\delta$ , acting as a single free parameter. The corresponding CDF is given by:

$$F_{AV}(t|\delta) = 1 - [S_A(t + \delta) \times S_V(t + \delta)] \quad (14)$$

where  $F_{AV}(t|\delta)$  denotes the probability of a multisensory response by time  $t$  given a shift  $\delta$ , and  $S_m(t + \delta)$  is the survivor function of modality  $m$  evaluated at the shifted time.

To account for SOA manipulations [37,38], we extended the two-stage relay model to incorporate onset lags. We defined the lag,  $\tau$ , as the time difference between the onsets of the leading and the lagging signal. This lag is not a free model parameter, instead it is directly determined by the experimentally defined SOA condition. Adding  $\tau$  can affect both relay model stages ( $T_{AV1}$ ,  $T_{AV2}$ ). At stage 1,  $\tau$  introduces a temporal disadvantage for the lagging modality. For example, if the visual signal is delayed by  $\tau$ , the duration of stage 1 is given by:

$$T_{AV1} = \min(T_{A1}, T_{V1} + \tau) \quad (15)$$

As described above, the first stage initiates the second stage at time  $T_{AV1}$ . Intuitively, if stage 2 begins before the delayed signal is presented ( $T_{AV1} < \tau$ ), that modality enters the race with a residual temporal disadvantage (equal to  $\tau - T_{AV1}$ ). Continuing with the example of a delayed visual signal, the duration of stage 2 is therefore given by:

$$T_{AV2} = \min(T_{A2}, T_{V2} + \max(0, \tau - T_{AV1})) \quad (16)$$

To describe the effect of temporal delays between signals, we divided the model predictions into two scenarios depending on the time  $u$  at which stage 2 is initiated. In the first, early initiation scenario, stage 2 begins before the lagging signal has been presented ( $u < \tau$ ). In this case, stage 1 is driven entirely by the leading signal. For example, if the visual signal lags by  $\tau$ , the contribution of the early initiation scenario to the overall model CDF is:

$$\Phi_{Early}(t) = \int_0^{\min(t, \tau)} \underbrace{f_{A1}(u)}_{\text{Stage 1}} \underbrace{[1 - S_{A2}(t - u)S_{V2}(t - \tau)]}_{\text{Stage 2}} du \quad (17)$$

Here,  $f_{A1}(u)$  denotes the probability density of the stage 1 processing time of the leading (auditory) modality. The stage 2 term follows Eq 11 but is adapted to reflect that visual processing can only begin after the lag  $\tau$  has elapsed. In other words, stage 2 begins immediately for the leading modality (after time  $u$ ), whereas the lagging modality can only start after its signal onset (at time  $\tau$ ), resulting in a residual temporal disadvantage.

In the second, late initiation scenario, stage 2 begins after the lagging signal has been presented ( $u \geq \tau$ ). In this case, stage 1 is driven by a race between both modalities. Continuing with the example where the visual signal lags by  $\tau$ , the contribution of the late initiation scenario to the overall model CDF is:

$$\Phi_{Late}(t) = \int_{\tau}^t \underbrace{g(u)}_{\text{Stage 1}} \underbrace{[1 - S_{A2}(t - u)S_{V2}(t - u)]}_{\text{Stage 2}} du \quad (18)$$

Here,  $g(u)$  denotes the probability density of the stage 1 completion time under a race between both modalities, which can be computed following [Eq 12](#), adapted to account for the lag  $\tau$  in the visual modality:

$$g(u) = f_{A1}(u)S_{V1}(u - \tau) + f_{V1}(u - \tau)S_{A1}(u) \quad (19)$$

Regarding the stage 2 term, because both signals are present at time  $u$ , processing starts immediately for both modalities, and [Eq 11](#) applies directly. The total predicted CDF is then given by the sum of the two components:

$$F_{AV}(t) = \Phi_{Early}(t) + \Phi_{Late}(t) \quad (20)$$

## RSE and RMI violations

To inspect the key features of the predictions derived from the relay model, we measured the RSE as well as violations of the RMI ([Figs 2](#) and [3](#)). To quantify these effects, we used geometric measures on the level of CDFs analogous to approaches described elsewhere [\[93,96\]](#).

## Model fitting

In addition to model simulations, we fitted the relay model to empirical data from [\[4,37,38\]](#). As outlined above, the relay model was constrained by the unisensory RT distributions estimated for each dataset ([Table 1](#)). To account for the observed multisensory RTs, the model included a single free parameter: the RT share of stage 1, constrained to the interval  $[0, 0.5]$ .

To fit the relay model to the multisensory RT distributions from [\[4\]](#), we compared empirical quantiles to the corresponding quantiles predicted by the relay model CDF ([Eq 13](#)). The RMSE between the empirical and model CDF was minimised by adjusting the RT share parameter. Optimisation was conducted using MATLAB's *fminbnd* function. All fitting procedures, including parameter recovery simulations, can be accessed from [\[91\]](#).

Similarly, we fitted the shifted race model to the same data ([Eq 14](#)). Here, the RMSE between the empirical and model CDF was minimised by adjusting the shift parameter  $\delta$ . The parameter search was constrained to ensure physical plausibility. The upper bound was fixed at 500 ms, while the lower bound was dynamically set for each condition as to ensure that the shifted times remained strictly positive. Optimisation was performed using MATLAB's *fmincon* function.

To fit the relay model to the multisensory RTs from [\[38\]](#), we used the relay model adapted for temporal asynchronies ([Eq 20](#)). Miller [\[38\]](#) defined the RSE as the difference in mean RT between the faster unisensory and the multisensory condition. To generate model predictions, we computed representative RT distributions for each condition and SOA by evaluating the inverse CDF (quantile function) from [Equation 20](#) at 200 evenly spaced probability values. Using these predicted RTs, we calculated the RSE analogously to Miller [\[38\]](#). We estimated the optimal RT share by minimising the RMSE between the empirical and predicted RSEs across the 11 SOA conditions. Optimisation was performed using MATLAB's *fminbnd* function.

The RT share parameter was fit to the multisensory conditions in both the SOA and signal strength datasets from Otto et al. [\[37\]](#) using QMPE [\[63,64\]](#). For the SOA dataset, we fitted the RT share parameter by minimising the joint NLL summed across all six SOA conditions (-60, -30, 0, 30, and 60 ms). This approach ensures that a single multisensory RT share accounts for the entire range of temporal asynchronies. For the signal strength dataset, we applied a stricter test of model generalisability. To test the model's predictive validity, we fitted the RT share parameter only to the congruent medium-intensity condition ( $A_M V_M$ ). This fitted RT share was then used to predict RTs for remaining five signal strength combinations (e.g.,  $A_S V_S$  or  $A_W V_M$ ) without further optimisation, effectively treating these conditions as out-of-sample tests.

## Author contributions

**Conceptualization:** Calvin Roberts, Thomas U. Otto.

**Data curation:** Calvin Roberts.

**Formal analysis:** Calvin Roberts.

**Funding acquisition:** Calvin Roberts.

**Project administration:** Thomas U. Otto.

**Software:** Calvin Roberts.

**Supervision:** Thomas U. Otto.

**Visualization:** Calvin Roberts.

**Writing – original draft:** Calvin Roberts.

**Writing – review & editing:** Calvin Roberts, Thomas U. Otto.

## References

1. Todd JW. Reaction to multiple stimuli. Science Press. 1912.
2. Hershenson M. Reaction time as a measure of intersensory facilitation. J Exp Psychol. 1962;63:289–93. <https://doi.org/10.1037/h0039516> PMID: 13906889
3. Kinchla RA. Detecting target elements in multielement arrays: A confusability model. Perception & Psychophysics. 1974;15(1):149–58. <https://doi.org/10.3758/bf03205843>
4. Miller J. Divided attention: Evidence for coactivation with redundant signals. Cogn Psychol. 1982;14(2):247–79. [https://doi.org/10.1016/0010-0285\(82\)90010-x](https://doi.org/10.1016/0010-0285(82)90010-x) PMID: 7083803
5. Raab DH. Division of psychology: Statistical facilitation of simple reaction times\*. Trans NY Acad Sci. 1962;24(5 Series II):574–90. <https://doi.org/10.1111/j.2164-0947.1962.tb01433.x>
6. Townsend JT, Wenger MJ. A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. Psychol Rev. 2004;111(4):1003–35. <https://doi.org/10.1037/0033-295X.111.4.1003> PMID: 15482071
7. Otto TU, Mamassian P. Multisensory decisions: The test of a race model, its logic, and power. Multisensory Research. 2017;30(1):1. <https://doi.org/10.1163/22134808-00002541>
8. Giard MH, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. J Cogn Neurosci. 1999;11(5):473–90. <https://doi.org/10.1162/089892999563544> PMID: 10511637
9. Hecht D, Reiner M, Karni A. Enhancement of response times to bi- and tri-modal sensory stimuli during active movements. Exp Brain Res. 2008;185(4):655–65. <https://doi.org/10.1007/s00221-007-1191-x> PMID: 17992522
10. Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, et al. Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex. 2005;15(7):963–74. <https://doi.org/10.1093/cercor/bhh197> PMID: 15537674
11. Otto TU, Mamassian P. Noise and correlations in parallel perceptual decision making. Curr Biol. 2012;22(15):1391–6. <https://doi.org/10.1016/j.cub.2012.05.031> PMID: 22771043
12. Wuergler SM, Crocker-Buque A, Meyer GF. Evidence for auditory-visual processing specific to biological motion. Seeing Perceiving. 2012;25(1):15–28. <https://doi.org/10.1163/187847611X620892> PMID: 22353566
13. Höchenberger R, Busch NA, Ohla K. Nonlinear response speedup in bimodal visual-olfactory object identification. Frontiers in Psychology. 2015;6. <https://doi.org/10.3389/fpsyg.2015.01477>
14. Matyjek M, Kita S, Torralba Cuello M, Soto Faraco S. Multisensory integration of speech and gestures in a naturalistic paradigm. Hum Brain Mapp. 2024;45(11):e26797. <https://doi.org/10.1002/hbm.26797> PMID: 39041175
15. Feintuch U, Cohen A. Visual attention and coactivation of response decisions for features from different dimensions. Psychol Sci. 2002;13(4):361–9. <https://doi.org/10.1111/1467-9280.00465> PMID: 12137140
16. Katzner S, Busse L, Treue S. Feature-based attentional integration of color and visual motion. J Vis. 2006;6(3):269–84. <https://doi.org/10.1167/6.3.7> PMID: 16643095
17. Koene AR, Zhaoping L. Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. J Vis. 2007;7(7):6.1–14. <https://doi.org/10.1167/7.7.6> PMID: 17685802

18. Mordkoff JT, Yantis S. Dividing attention between color and shape: Evidence of coactivation. *Percept Psychophys.* 1993;53(4):357–66. <https://doi.org/10.3758/bf03206778> PMID: [8483699](#)
19. Schröter H, Frei LS, Ulrich R, Miller J. The auditory redundant signals effect: An influence of number of stimuli or number of percepts?. *Atten Percept Psychophys.* 2009;71(6):1375–84. <https://doi.org/10.3758/APP.71.6.1375> PMID: [19633352](#)
20. Engmann S, Cousineau D. Triple redundant signals effect in the visual modality. *Univ Psychol.* 2013;12(5). <https://doi.org/10.11144/javeriana.upsy12-5.trse>
21. Brandwein AB, Foxe JJ, Butler JS, Russo NN, Altschuler TS, Gomes H, et al. The development of multisensory integration in high-functioning autism: High-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cereb Cortex.* 2013;23(6):1329–41. <https://doi.org/10.1093/cercor/bhs109> PMID: [22628458](#)
22. Brang D, Williams LE, Ramachandran VS. Grapheme-color synesthetes show enhanced crossmodal processing between auditory and visual modalities. *Cortex.* 2012;48(5):630–7. <https://doi.org/10.1016/j.cortex.2011.06.008> PMID: [21763646](#)
23. Corballis MC. Interhemispheric neural summation in the absence of the corpus callosum. *Brain.* 1998;121 (Pt 9):1795–807. <https://doi.org/10.1093/brain/121.9.1795> PMID: [9762966](#)
24. Harrar V, Tammam J, Pérez-Bellido A, Pitt A, Stein J, Spence C. Multisensory integration and attention in developmental dyslexia. *Curr Biol.* 2014;24(5):531–5. <https://doi.org/10.1016/j.cub.2014.01.029> PMID: [24530067](#)
25. Van der Stoep N, Van der Stigchel S, Van Engelen RC, Biesbroek JM, Nijboer TCW. Impairments in multisensory integration after stroke. *J Cogn Neurosci.* 2019;31(6):885–99. [https://doi.org/10.1162/jocn\\_a\\_01389](https://doi.org/10.1162/jocn_a_01389) PMID: [30883294](#)
26. Blurton SP, Greenlee MW, Gondan M. Multisensory processing of redundant information in go/no-go and choice responses. *Atten Percept Psychophys.* 2014;76(4):1212–33. <https://doi.org/10.3758/s13414-014-0644-0> PMID: [24627209](#)
27. Hughes HC, Reuter-Lorenz PA, Nozawa G, Fendrich R. Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *J Exp Psychol Hum Percept Perform.* 1994;20(1):131–53. <https://doi.org/10.1037//0096-1523.20.1.131> PMID: [8133219](#)
28. Diederich A, Colonius H. Modeling the time course of multisensory interaction in manual and saccadic responses. *The Handbook of Multisensory Processes.* The MIT Press. 2004. 395–408. <https://doi.org/10.7551/mitpress/3422.003.0030>
29. Van der Stoep N, Van der Smagt MJ, Notaro C, Spock Z, Naber M. The additive nature of the human multisensory evoked pupil response. *Sci Rep.* 2021;11(1):707. <https://doi.org/10.1038/s41598-020-80286-1> PMID: [33436889](#)
30. Liu Y, Otto TU. The role of context in experiments and models of multisensory decision making. *Journal of Mathematical Psychology.* 2020;96:102352. <https://doi.org/10.1016/j.jmp.2020.102352>
31. Bogacz R. Optimal decision-making theories: linking neurobiology with behaviour. *Trends Cogn Sci.* 2007;11(3):118–25. <https://doi.org/10.1016/j.tics.2006.12.006> PMID: [17276130](#)
32. Gold JI, Shadlen MN. The neural basis of decision making. *Annual Review of Neuroscience.* 2007;30(1):535–74. <https://doi.org/10.1146/annurev.neuro.29.051605.113038>
33. Ratcliff R, Smith PL, Brown SD, McKoon G. Diffusion decision model: Current issues and history. *Trends Cogn Sci.* 2016;20(4):260–81. <https://doi.org/10.1016/j.tics.2016.01.007> PMID: [26952739](#)
34. Matzke D, Wagenmakers E-J. Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychon Bull Rev.* 2009;16(5):798–817. <https://doi.org/10.3758/PBR.16.5.798> PMID: [19815782](#)
35. Anders R, Alario F-X, Van Maanen L. The shifted Wald distribution for response time data analysis. *Psychol Methods.* 2016;21(3):309–27. <https://doi.org/10.1037/met0000066> PMID: [26867155](#)
36. Evans NJ, Wagenmakers EJ. Evidence accumulation models: Current limitations and future directions. *TQMP.* 2020;16(2):73–90. <https://doi.org/10.20982/tqmp.16.2.p073>
37. Otto TU, Dassy B, Mamassian P. Principles of multisensory behavior. *J Neurosci.* 2013;33(17):7463–74. <https://doi.org/10.1523/JNEUROSCI.4678-12.2013> PMID: [23616552](#)
38. Miller J. Timecourse of coactivation in bimodal divided attention. *Percept Psychophys.* 1986;40(5):331–43. <https://doi.org/10.3758/bf03203025> PMID: [3786102](#)
39. Innes BR, Otto TU. A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. *Sci Rep.* 2019;9(1):2921. <https://doi.org/10.1038/s41598-019-39924-6> PMID: [30814642](#)
40. Colonius H. Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology.* 1990;34(3):253–75. [https://doi.org/10.1016/0022-2496\(90\)90032-5](https://doi.org/10.1016/0022-2496(90)90032-5)
41. Mordkoff JT, Yantis S. An interactive race model of divided attention. *J Exp Psychol Hum Percept Perform.* 1991;17(2):520–38. <https://doi.org/10.1037//0096-1523.17.2.520> PMID: [1830090](#)
42. Colonius H, Arndt P. A two-stage model for visual-auditory interaction in saccadic latencies. *Percept Psychophys.* 2001;63(1):126–47. <https://doi.org/10.3758/bf03200508> PMID: [11304009](#)
43. Miller J. Statistical facilitation and the redundant signals effect: What are race and coactivation models?. *Atten Percept Psychophys.* 2016;78(2):516–9. <https://doi.org/10.3758/s13414-015-1017-z> PMID: [26555650](#)
44. Luce RD. Response times: their role in inferring elementary mental organization. Oxford University Press. 1986.

45. Ashby FG, Townsend JT. Varieties of perceptual independence. *Psychol Rev.* 1986;93(2):154–79. <https://doi.org/10.1037//0033-295x.93.2.154> PMID: [3714926](#)
46. Sternberg S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica.* 1969;30:276–315. [https://doi.org/10.1016/0001-6918\(69\)90055-9](https://doi.org/10.1016/0001-6918(69)90055-9)
47. Purcell BA, Heitz RP, Cohen JY, Schall JD, Logan GD, Palmeri TJ. Neurally constrained modeling of perceptual decision making. *Psychol Rev.* 2010;117(4):1113–43. <https://doi.org/10.1037/a0020311> PMID: [20822291](#)
48. Ratcliff R, McKoon G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.* 2008;20(4):873–922. <https://doi.org/10.1162/neco.2008.12-06-420> PMID: [18085991](#)
49. Noppeney U. Perceptual inference, learning, and attention in a multisensory world. *Annual Review of Neuroscience.* 2021;44(1):449–73. <https://doi.org/10.1146/annurev-neuro-100120-085519>
50. Rohe T, Noppeney U. Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Curr Biol.* 2016;26(4):509–14. <https://doi.org/10.1016/j.cub.2015.12.056> PMID: [26853368](#)
51. Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory?. *Trends in Cognitive Sciences.* 2006;10(6):278–85. <https://doi.org/10.1016/j.tics.2006.04.008>
52. Townsend JT, Nozawa G. Serial exhaustive models can violate the race model inequality: Implications for architecture and capacity. *Psychological Review.* 1997;104(3):595–602. <https://doi.org/10.1037/0033-295X.104.3.595>
53. Carpenter RHS, Reddi BAJ, Anderson AJ. A simple two-stage model predicts response time distributions. *J Physiol.* 2009;587(Pt 16):4051–62. <https://doi.org/10.1113/jphysiol.2009.173955> PMID: [19564395](#)
54. Sun P, Landy MS. A Two-stage process model of sensory discrimination: An alternative to drift-diffusion. *J Neurosci.* 2016;36(44):11259–74. <https://doi.org/10.1523/JNEUROSCI.1367-16.2016> PMID: [27807167](#)
55. Reddi BA. Decision making: The two stages of neuronal judgement. *Curr Biol.* 2001;11(15):R603–6. [https://doi.org/10.1016/s0960-9822\(01\)00363-3](https://doi.org/10.1016/s0960-9822(01)00363-3) PMID: [11516966](#)
56. Rüter J, Marcille N, Sprekeler H, Gerstner W, Herzog MH. Paradoxical evidence integration in rapid decision processes. *PLoS Comput Biol.* 2012;8(2):e1002382. <https://doi.org/10.1371/journal.pcbi.1002382> PMID: [22359494](#)
57. Hübner R, Steinhauser M, Lehle C. A dual-stage two-phase model of selective attention. *Psychol Rev.* 2010;117(3):759–84. <https://doi.org/10.1037/a0019471> PMID: [20658852](#)
58. Schall JD, Purcell BA, Heitz RP, Logan GD, Palmeri TJ. Neural mechanisms of saccade target selection: gated accumulator model of the visual-motor cascade. *Eur J Neurosci.* 2011;33(11):1991–2002. <https://doi.org/10.1111/j.1460-9568.2011.07715.x> PMID: [21645095](#)
59. Murray CA, de Larrea-Mancera ESL, Glicksohn A, Shams L, Seitz AR. Revealing multisensory benefit with diffusion modeling. *Journal of Mathematical Psychology.* 2020;99:102449. <https://doi.org/10.1016/j.jmp.2020.102449>
60. Kelly SP, O'Connell RG. Internal and external influences on the rate of sensory evidence accumulation in the human brain. *J Neurosci.* 2013;33(50):19434–41. <https://doi.org/10.1523/JNEUROSCI.3355-13.2013> PMID: [24336710](#)
61. Diederich A, Colonius H. Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Percept Psychophys.* 2004;66(8):1388–404. <https://doi.org/10.3758/bf03195006> PMID: [15813202](#)
62. Diederich A, Colonius H. Intersensory facilitation in the motor component?. *Psychol Res.* 1987;49(1):23–9. <https://doi.org/10.1007/bf00309199>
63. Heathcote A, Brown S, Cousineau D. QMPE: Estimating lognormal, wald, and weibull RT distributions with a parameter-dependent lower bound. *Behav Res Methods Instrum Comput.* 2004;36(2):277–90. <https://doi.org/10.3758/bf03195574> PMID: [15354694](#)
64. Heathcote A, Brown S, Mewhort DJK. Quantile maximum likelihood estimation of response time distributions. *Psychon Bull Rev.* 2002;9(2):394–401. <https://doi.org/10.3758/bf03196299> PMID: [12120806](#)
65. Wagenmakers E-J, Brown S. On the linear relation between the mean and the standard deviation of a response time distribution. *Psychol Rev.* 2007;114(3):830–41. <https://doi.org/10.1037/0033-295X.114.3.830> PMID: [17638508](#)
66. Senkowski D, Saint-Amour D, Höfle M, Foxe JJ. Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage.* 2011;56(4):2200–8. <https://doi.org/10.1016/j.neuroimage.2011.03.075> PMID: [21497200](#)
67. Schwarz W. Diffusion, superposition, and the redundant-targets effect. *J Math Psychol.* 1994;38(4):504–20. <https://doi.org/10.1006/jmps.1994.1036>
68. Schwarz W. A new model to explain the redundant-signals effect. *Percept Psychophys.* 1989;46(5):498–500. <https://doi.org/10.3758/bf03210867> PMID: [2813037](#)
69. Miller J, Ulrich R. Simple reaction time and statistical facilitation: A parallel grains model. *Cogn Psychol.* 2003;46(2):101–51. [https://doi.org/10.1016/s0010-0285\(02\)00517-0](https://doi.org/10.1016/s0010-0285(02)00517-0) PMID: [12643892](#)
70. Diederich A. Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *J Math Psychol.* 1995;39(2):197–215. <https://doi.org/10.1006/jmps.1995.1020>
71. Purcell BA, Schall JD, Logan GD, Palmeri TJ. From salience to saccades: multiple-alternative gated stochastic accumulator model of visual search. *J Neurosci.* 2012;32(10):3433–46. <https://doi.org/10.1523/JNEUROSCI.4622-11.2012> PMID: [22399766](#)

72. Caie B, Endres D, Khan AZ, Blohm G. Choice anticipation as gated accumulation of sensory predictions. *J Neurophysiol*. 2025;133(4):1159–75. <https://doi.org/10.1152/jn.00041.2024> PMID: 39970908
73. Loughnane GM, Newman DP, Bellgrove MA, Lalor EC, Kelly SP, O'Connell RG. Target selection signals influence perceptual decisions by modulating the onset and rate of evidence accumulation. *Curr Biol*. 2016;26(4):496–502. <https://doi.org/10.1016/j.cub.2015.12.049> PMID: 26853360
74. Nunez MD, Gosai A, Vandekerckhove J, Srinivasan R. The latency of a visual evoked potential tracks the onset of decision making. *Neuroimage*. 2019;197:93–108. <https://doi.org/10.1016/j.neuroimage.2019.04.052> PMID: 31028925
75. Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Brain Res Cogn Brain Res*. 2002;14(1):115–28. [https://doi.org/10.1016/s0926-6410\(02\)00066-6](https://doi.org/10.1016/s0926-6410(02)00066-6) PMID: 12063135
76. Gondan M, Röder B. A new method for detecting interactions between the senses in event-related potentials. *Brain Res*. 2006;1073–1074:389–97. <https://doi.org/10.1016/j.brainres.2005.12.050> PMID: 16427613
77. Nickerson RS. Intersensory facilitation of reaction time: Energy summation or preparation enhancement?. *Psychol Rev*. 1973;80(6):489–509. <https://doi.org/10.1037/h0035437> PMID: 4757060
78. Kandil FI, Diederich A, Colonius H. Parameter recovery for the time-window-of-integration (TWIN) model of multisensory integration in focused attention. *J Vis*. 2014;14(11):14. <https://doi.org/10.1167/14.11.14> PMID: 25253874
79. Krummenacher J, Grubert A, Töllner T, Müller HJ. Saliency-based integration of redundant signals in visual pop-out search: Evidence from behavioral and electrophysiological measures. *J Vis*. 2014;14(3):26. <https://doi.org/10.1167/14.3.26> PMID: 24648196
80. Poom L. Integration of colour, motion, orientation, and spatial frequency in visual search. *Perception*. 2009;38(5):708–18. <https://doi.org/10.1068/p6072> PMID: 19662946
81. Wang WY, Hu L, Valentini E, Xie XB, Cui HY, Hu Y. Dynamic characteristics of multisensory facilitation and inhibition. *Cogn Neurodyn*. 2012;6(5):409–19. <https://doi.org/10.1007/s11571-012-9197-x> PMID: 24082962
82. Pomper U, Brincker J, Harwood J, Prihodko I, Senkowski D. Taking a call is facilitated by the multisensory processing of smartphone vibrations, sounds, and flashes. *PLoS One*. 2014;9(8):e103238. <https://doi.org/10.1371/journal.pone.0103238> PMID: 25116195
83. Neisser U. *Cognitive psychology*. East Norwalk, CT, US: Appleton-Century-Crofts. 1967.
84. Shaw LH, Freedman EG, Crosse MJ, Nicholas E, Chen AM, Braiman MS, et al. Operating in a multisensory context: Assessing the Interplay between multisensory reaction time facilitation and inter-sensory task-switching effects. *Neuroscience*. 2020;436:122–35. <https://doi.org/10.1016/j.neuroscience.2020.04.013> PMID: 32325100
85. Gondan M, Lange K, Rösler F, Röder B. The redundant target effect is affected by modality switch costs. *Psychon Bull Rev*. 2004;11(2):307–13. <https://doi.org/10.3758/bf03196575> PMID: 15260198
86. Gondan M, Vorberg D, Greenlee MW. Modality shift effects mimic multisensory interactions: An event-related potential study. *Exp Brain Res*. 2007;182(2):199–214. <https://doi.org/10.1007/s00221-007-0982-4> PMID: 17562033
87. Plass J, Brang D. Multisensory stimuli shift perceptual priors to facilitate rapid behavior. *Sci Rep*. 2021;11(1):23052. <https://doi.org/10.1038/s41598-021-02566-8> PMID: 34845325
88. Crosse MJ, Foxe JJ, Tarrit K, Freedman EG, Molholm S. Resolution of impaired multisensory processing in autism and the cost of switching sensory modality. *Commun Biol*. 2022;5(1):601. <https://doi.org/10.1038/s42003-022-03519-1> PMID: 35773473
89. Van der Stoep N, Van der Stigchel S, Nijboer TCW. Exogenous spatial attention decreases audiovisual integration. *Atten Percept Psychophys*. 2015;77(2):464–82. <https://doi.org/10.3758/s13414-014-0785-1> PMID: 25341648
90. Ratcliff R. Group reaction time distributions and an analysis of distribution statistics. *Psychol Bull*. 1979;86(3):446–61. <https://doi.org/10.1037//0033-2909.86.3.446> PMID: 451109
91. Roberts K, Otto TU. Relay Model Code and Accompanying Data. 2026. <https://doi.org/10.5281/zenodo.20382132>
92. Ulrich R, Miller J, Schröter H. Testing the race model inequality: An algorithm and computer programs. *Behav Res Methods*. 2007;39(2):291–302. <https://doi.org/10.3758/bf03193160> PMID: 17695357
93. Otto TU. RSE-box: An analysis and modelling package to study response times to multiple signals. *TQMP*. 2019;15(2):112–33. <https://doi.org/10.20982/tqmp.15.2.p112>
94. Chhikara RS, Folks JL. *The inverse Gaussian distribution: theory, methodology, and applications*. New York: Marcel Dekker. 1988.
95. McFadden D. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*. 1989;57(5):995. <https://doi.org/10.2307/1913621>
96. Colonius H, Diederich A. The race model inequality: Interpreting a geometric measure of the amount of violation. *Psychol Rev*. 2006;113(1):148–54. <https://doi.org/10.1037/0033-295X.113.1.148> PMID: 16478305