Cognitive Psychology 164 (2026) 101803

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Integration of redundant signals in dynamic multisensory contexts: the principles and computational mechanisms

Shiqi Tan [a,b], Yuhui Cheng [c], Jieru Chen [a,b], Xiangyong Yuan [a,b,*], Yi Jiang [a,b]

[a] State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing 100101, China
[b] Department of Psychology, University of Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, China
[c] School of Psychology, Nanjing Normal University, Nanjing 210097, China

ARTICLE INFO

Keywords:
Multisensory integration
Perceptual decision making
Redundant signals effect
Race model
Evidence accumulation

ABSTRACT

One benefit of multisensory integration is the acceleration of responses in a perceptual decision task, known as the redundant signals effect (RSE). Based on probability summation, two basic principles have been proposed to explain such benefit: the "principle of congruent effectiveness", which suggests maximal RSE when unisensory performances are comparable, and the "variability rule", which proposes that RSE increases when unisensory performances are more variable. Yet, it remains unclear whether these principles extend to dynamic multisensory contexts and how they manifest in the model architectures proposed to explain RSE. To address these questions, we evaluated RSE in a multisensory context featuring transient onsets and offsets using a change detection task. Our results showed that both principles predicted the rank of the RSE in this dynamic setting, with the principle of congruent effectiveness emerging as the dominant one. Model comparison identified two best-performing model architectures that implement distinct crossmodal interactions (enhanced evidence accumulation rates or lowered decision criterion) in the perceptual process when two unimodal signals race towards their decision criterion. Moreover, the key predictors derived from the principles effectively modulate the crossmodal interactions in the two models. Together, these findings demonstrate that the principles underlying RSE generalize to dynamic multisensory contexts and operate via distinct computational mechanisms, shedding light on how perceptual systems flexibly integrate multisensory cues to optimize decision-making.

1. Introduction

In our daily experiences, we often make perceptual decisions based on multisensory signals, such as the sights we see, the sounds we hear, and the textures we touch (Alais et al., 2010; Bizley et al., 2016; Drugowitsch et al., 2014; Franzen et al., 2020; Ghazanfar & Schroeder, 2006; Raposo et al., 2012). Imagine you are walking down a street on a densely foggy morning. Either a vague outline of the clothing or a fleeting snippet of the voice contributes to your decision that someone is in your path. But when both visual and auditory cues are detected, you may reach the decision more quickly, and turn left or right to avoid a collision. In such a context where decision can be made based on a single signal alone, combining these signals together (redundant signals) greatly facilitates and accelerates the

* Corresponding author at: State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing 100101, China.
E-mail address: yuanxy@psych.ac.cn (X. Yuan).

https://doi.org/10.1016/j.cogpsych.2026.101803
Received 30 July 2025; Received in revised form 11 April 2026; Accepted 20 April 2026
Available online 28 April 2026

decision-making process (Gondan & Minakata, 2016; Otto et al., 2013; Otto & Mamassian, 2016; Stevenson et al., 2014), ensuring significant information are utilized and prompt reactions are made in a shorter time. This is well-known as the redundant signals effect (RSE).

1.1. The principle of redundant signals effect

In classical RSE experiments, observers are required to complete a speeded reaction task by pressing the button as quickly as possible when they perceive the stimulus regardless of the modality it comes from. Usually, the reaction time (RT) to detect the multisensory stimuli is significantly shorter than to detect its corresponding unisensory components (Chua et al., 2022; Hershenson, 1962; Innes & Otto, 2019; Otto et al., 2013; Otto & Mamassian, 2012; Plass & Brang, 2021; Raab, 1962; Roberts et al., 2024; Todd, 1912).

Early electrophysiological studies on cats' superior colliculus established three canonical principles of multisensory integration: signals produce the largest neural enhancements when they are spatially aligned ("spatial rule"), temporally coincident ("temporal rule"), or weak in strength so that combination yields proportionally larger gains ("inverse effectiveness") (Meredith & Stein, 1983; Stein & Meredith, 1993; Stein & Stanford, 2008). Since then, these principles have been repeatedly tested on their ability to predict behavioral benefits of multisensory integration. However, their direct applicability to RSE—the speedup of reaction time for multisensory versus unisensory stimuli—is limited. Empirically, RSE magnitude is largely insensitive to precise spatial alignment (Murray et al., 2005), the temporal relation that maximizes facilitation is variable and often occurs with delayed auditory onsets rather than strict simultaneity (Hershenson, 1962; Miller, 1986), and evidence for inverse effectiveness is mixed across paradigms (Chandrasekaran et al., 2011; Senkowski et al., 2011). Thus, these rules developed from neurophysiology do not straightforwardly predict behavioral speeding. Researchers have been seeking for other behavior-sensitive principles that explain the redundancy gains in RSE.

In the framework of probability summation, the RSE can be straightforwardly explained by a race between two statistical independent decision processes completed in each modality. On each trial when redundant signals are simultaneously presented, the RT is determined by the faster decision process. Starting from the assumption of statistical independence, Raab demonstrated that the RT facilitation increases as the distance between the RTs to the two redundant signals shortened (Raab, 1962). This finding is in concordant with the results collected on the same year that the RT facilitation reached maximum when the stimulus onset asynchrony (SOA) was approximately equal to the difference in RTs to the visual and auditory stimuli (Hershenson, 1962). Since then, it has been broadly accepted that a larger overlap between the RT distributions of redundant signals would increase the RSE.

However, in the original paper of Raab, both the simulation and modelling efforts took little account of the variability of RT distributions, assuming equal variance to simplify the algebra (ref to their Eq. (4)). This contrasts with the original equation (ref to their Eq. (1)), which is distribution free (Raab, 1962). More recently, Otto et al. (2013) derived two principles from the probability summation framework adopted by Raab (1962). One principle, the principle of congruent effectiveness, explicitly quantifies that the degree of overlap between the two unisensory RT distributions as the similarity in their mean and variability. The more similar of them, the larger the RSE. The other principle, the variability rule, additionally states that when the variability of the two unisensory RT distributions are larger, the RSE turns larger.

Along with the simulation, they conducted two empirical experiments to assess the two derived principles (Otto et al., 2013). In the first experiment, the SOA between the unisensory signals was varied, as did in (Hershenson, 1962) and other studies (Blurton et al., 2014; Chandrasekaran et al., 2019; Harrar et al., 2017; Miller, 1986). As predicted by the principle of congruent effectiveness, the maximal RSE occurred at the SOA that counteracted the 33-ms differences in RTs to unisensory signals. In the second experiment, each unisensory signal was presented at one of three levels of signal strength, allowing the RT being measured at nine combinations in the redundant conditions. The observed RSEs were ordered structurally according to the two principles. For example, sound with weak strength and motion with medium strength were most similar in their median and variance of RTs (median difference 12 ms, variance difference 4 ms) and with relatively larger variances (123 and 119 ms), their combination therefore yielded the greatest RT facilitation (67 ms). In contrast, when sound and motion both with strong strength had comparable distances in median and variance of RTs (16 and 6 ms) but the variances themselves were much smaller (67 and 73 ms), the RSE fell down to 39 ms. However, if there were huge distances between median and variance of RTs to sound and motion, the RSE almost vanished (strong sound and weak motion, ref to their Fig. 5). Moreover, if the two principles contradicted with each other, for instance, distances in median and variance of RTs to unisensory signals enlarged while at the same time their variances grew larger, the RSEs stayed rather constant (data pointed by arrows in their Fig. 6).

It is obvious that the key dimensions underlying the two RSE principles are the ultimate relationship between unisensory RT central tendency and variability. That said, if any other principles, such as the previously mentioned three generic principles (spatial, temporal and inverse effectiveness), modulate the RSE, they do so by altering the overlap between unisensory RT distributions (shaping their mean and variability), in accordance with the two principles first proposed by Raab (1962) and later supplemented by Otto et al. (2013). Two interesting questions then arise: Can the two principles be generalized to explain the RSEs in other more complicated multisensory context? In predicting the RSE, are they of the same weight or does one principle dominate the other?

1.2. Computational models interpreting the redundant signals effect

In the probability summation framework of RSE, processing in each modality competes to arrive at a decision module. The decision module computes a logical OR operation: The faster processing that enters it drives a response in the redundant conditions. The model

is therefore metaphorically called a race model. It has two core assumptions, statistical independence and context invariance. The first assumed that in redundant conditions, the processing in the one modality has little or no influence on the processing in the other modality; The second assumed that the processing of unisensory signals in redundant conditions is equal to the processing when they are presented alone as single signals. Then, in the race model the cumulative distribution function (CDF) of RTs to the redundant signals can be predicted by the CDF of RTs to the single signals,

$$P_{AV}(t) = P_A(t) + P_V(t) - P_A(t) \times P_V(t) \quad (1)$$

P_{AV} (*t*), P_A (*t*), P_V (*t*) refers to the cumulative distribution functions of RTs in audiovisual, auditory, visual conditions.[1] Here *t* is a temporal cutoff (i.e. "$\leq t$"); For example, P_{AV} (*t*) expresses the probability that the RT in an AV trial occurs at or before time *t*. Derived from Eq. (1), the upper bound of RSE in the race model, also known as the Miller's bound or the race model inequality, is as followed:

$$P_{AV}(t) \leq P_A(t) + P_V(t) \quad (2)$$

However, the race model inequality is frequently violated by observations, falsifying either the model's two assumptions, or the model architecture (A thorough review about the race model inequality test, ref to (Gondan & Minakata, 2016)). Simply speaking, there must occur some interactions between the two modalities. Two candidates explain which interaction causes the violation, each corresponding to a class of computational implementation (Fig. 1, also see (Colonius & Diederich, 2006; Gondan & Minakata, 2016; Miller, 2016; Mordkoff & Yantis, 1991; Yang et al., 2018)).

Context-variant race models. The first class drops the two assumptions, statistical independence and context invariance of the race model (Innes & Otto, 2019; Mordkoff & Yantis, 1991; Otto et al., 2013; Otto & Mamassian, 2012; Plass & Brang, 2021; Yang et al., 2018). A graphic illustration of these context-variant race models is in Fig. 1a. In one deterministic model released by (Otto & Mamassian, 2012), the RT is transformed to its reciprocal that can be loosely interpreted as the evidence accumulation rate by adopting a linear evidence-accumulating framework in decision-making (the Linear Approach to Threshold with Ergodic Rate model, ref to (Noorani & Carpenter, 2016; Reddi et al., 2003), and Eq. (5)). The race model then estimates the evidence accumulation rate in the redundant conditions as the maximal evidence accumulation rate in the two single evidence accumulation processes. The authors made two critical modifications to this context-invariant race model. First, they included a negative correlation between two unisensory signals to capture the history effect especially the modal switching cost (Gondan et al., 2004; Gondan et al., 2007; Shaw et al., 2020; Spence et al., 2001). Second, they included additional noise to increase the variability of evidence accumulation rates for each unisensory signal in the redundant conditions to improve model performance. Clearly, the two modifications abandoned the assumptions of the original context-invariant race model.[2] The model is then a context-variant one and termed the multisensory noise (MN) model in our study (Fig. 1b).

Based on the MN model, Plass and Brang (2021) recently proposed another two context-variant race models by introducing other types of crossmodal interaction. In their models, the correlation between unisensory signals is retained but the variability of evidence accumulation rate is held constant. One model allows the decision criterion for each unisensory signal to lower down in the redundant conditions, while the other allows the evidence accumulation rate to increase (Fig. 1c and d). The former model is named the crossmodal pre-potentiation (CP) model, and the latter the crossmodal coactivation (CC) model. Since both the shift of criterion and the increase of evidence accumulation rate (or an enhancement of neural activities) have been found in studies of multisensory interaction (Kayser et al., 2008; Lakatos et al., 2007; Lippert et al., 2007; Mercier & Cappe, 2020; Regenbogen et al., 2018; Schirillo, 2011), the two models are behaviorally and neurophysiologically compatible. The two types of multisensory interaction are potentially the consequence of phase resetting between sensory cortices (Lakatos et al., 2007; Mégevand et al., 2020; Mercier et al., 2015; Thorne et al., 2011). Plass & Brang (2021) collected a large dataset that contained RTs to visual, auditory, tactile, and bimodal stimuli, and evaluated the performance of the three context-variant race models (MN, CP, and CC). The results demonstrated that the CP model with a positive correlation and decreased decision criterion outperform the MN and CC models.

[1] The derivation of Eq. (1) under the two assumptions of race model can be found in (Colonius & Diederich, 2006; Gondan & Minakata, 2016; Otto & Mamassian, 2016). Here we briefly described it, as well as the derivation of race model inequality (Eq. 2) from Eq. (1). First, race model assumes $P_{AV}(t) = P\left[min(D_{A\ in\ AV}, D_{V\ in\ AV}) \leq t\right] = P_{A\ in\ AV}(t) \cup P_{V\ in\ AV}(t)$, where D is the decision time and t also represent the cutoff. If statistical independence is assumed, then $P_{AV}(t) = P_{AinAV}(t) + P_{VinAV}(t) - P_{AinAV}(t) \times P_{VinAV}(t)$. If context invariance is additionally assumed, we obtain Eq. (1), $P_{AV}(t) = P_A(t) + P_V(t) - P_A(t) \times P_V(t)$. Because $P(t)$ is a non-negative value, $P_{AV}(t)$ is always smaller than $P_A(t) + P_V(t)$, indicating a negative correlation between D_{AinAV} and D_{VinAV}. Only when the correlation reaches maximal, $P_{AV}(t) = P_A(t) + P_V(t)$. Thus, Eq. (2) denotes statistical dependence between the two unisensory processes within a bimodal trial.

[2] There is no doubt that the inclusion of additional noise (parameter σ) violates the context invariance assumption of the original race model (Eq. (1). However, the negative correlation between unisensory trials due to history effects (parameter ρ) in Otto's model cannot be simply reduced to a violation of statistical independence. Thanks to an anonymous reviewer for bring this critical point to our attention. History effects reliably cause a negative correlation between unisensory processing in the two unimodal conditions, and may simultaneously influence the two unisensory processing in bimodal trials in the same direction. However, the negative correlation arising from history effects (inter-trial dependence) is not necessarily equal to the negative correlation between the two unisensory processing in bimodal conditions (within-trial dependence). Conceptually, statistical independence refers to within-trial independence in probability theory. Only when the two unisensory processing in bimodal trials exhibits a negative correlation that exactly matches the one between unimodal trials, can we safely claim that statistical independence is violated. Otherwise, context invariance is also violated under history effects. Thus, inter-trial dependence due to history effects must be strictly distinguished from statistical (within-trial) dependence. Notably, not only the parameter ρ in Otto's model, but also the other parameters in CC and CA models may simultaneously violate the two race model assumptions.

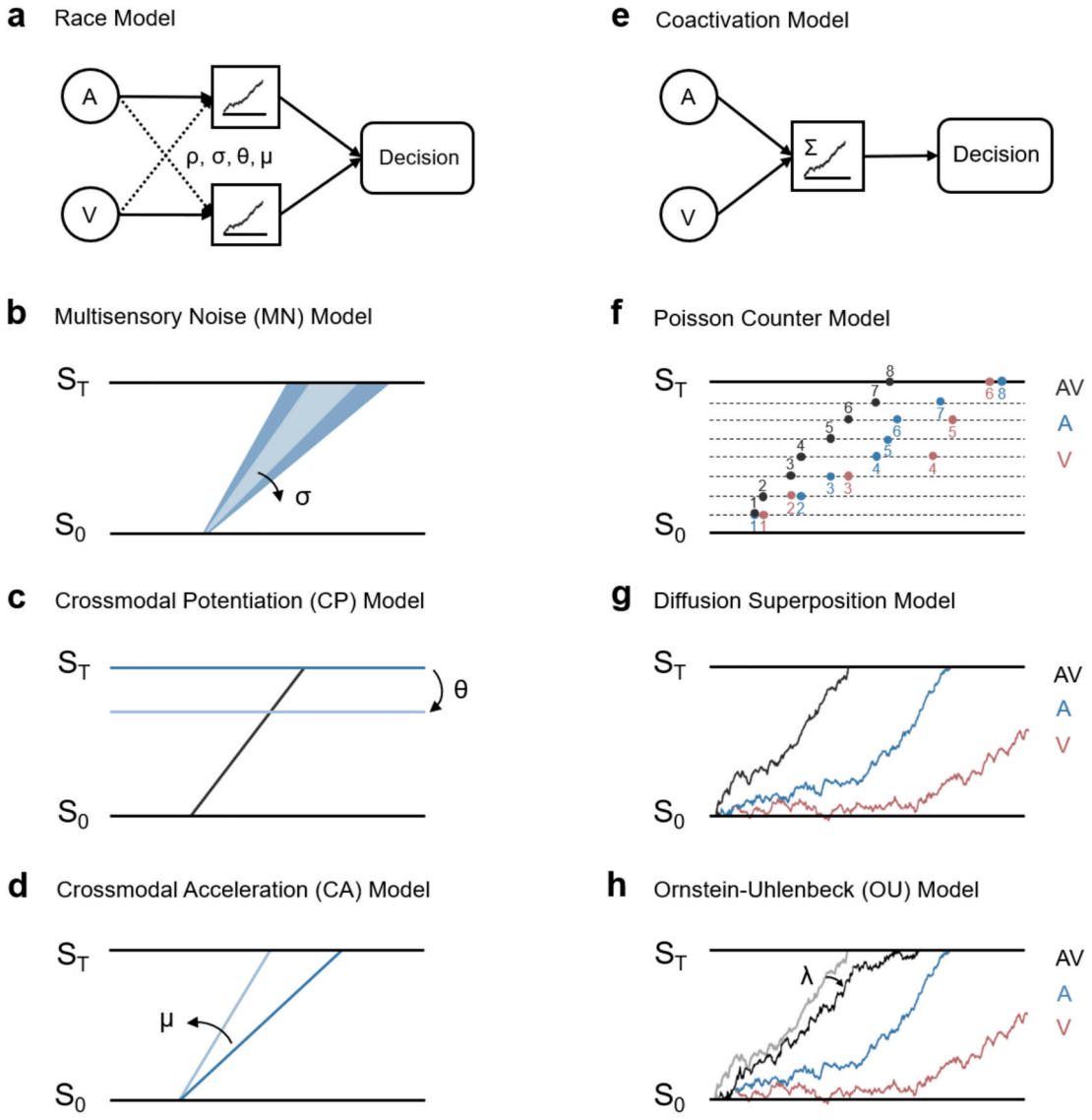

Fig. 1. Two main model architectures for the RSE. (a)(e) The core tenet of context-variant race model class and coactivation model class. Evidence in the two modalities are accumulated in a race or pooled together before reaching the decision threshold. (b)–(d) Three specific context-variant race models. In the multisensory noise (MN) model, the variability of evidence accumulation rate (σ) increases. In the crossmodal potentiation (CP) model, the decision threshold (θ) lowers down, In the crossmodal acceleration (CA) model, the evidence accumulation rate (μ) increases. (f)–(h) Three specific coactivation models. (f) Possion Counter model, the pooled evidences are discrete, following the Possion distribution. (g) Diffusion Superposition model, the pooled evidences are accumulated in a standard Wiener process with constant drift (a perfect integrator). (h) Ornstein-Uhlenbeck (OU) model, the pooled evidences are accumulated in a leaky diffusion process (state-dependent drift, leakage parameter λ). Panels (g) and (h) both represent continuous diffusion accumulators; the OU model differs from the Wiener model by the leak term that causes mean-reversion of the accumulation. S_0 represents the initial state while S_T represents the state at the decision threshold.

Noteworthy, although the CC model in Plass & Brang (2021) can be considered as a type of coactivation model (Miller, 2016), the present study renamed it the crossmodal acceleration (CA) model henceforth to distinguish with other types of coactivation models that have been used for decades to explain the violation of the race model inequality (reviewed in the next section). It is also worth to mention that in other literature, the context-variant race models are named the parallel interactive race models (Mordkoff & Yantis, 1991; Yang et al., 2018). For example, in the unequal-weighted parallel model, the interaction between two unisensory evidence accumulation processes are assumed asymmetric (Yang et al., 2018), in contrast to the three race models reviewed above.

Coactivation models. The other class of models completely abandon the framework of race model in explaining the violation of race model inequality. It instead assumes that the RSE is not caused by a race between two independent or interactive unisensory processing but their summation. Specifically, the processing in each modality is pooled into a common channel before arriving at the decision module. When the pooled information exceeds the criterion of the decision module, a response is triggered (Fig. 1e). The

coactivation models are also consistent with current neurophysiological and neuroimaging evidence: Multisensory neurons and patches activated by inputs from more than one modality are widely distributed in the superior colliculus, primary sensory cortices, and association cortices (e.g., the superior temporal sulcus) (Alais et al., 2010; Beauchamp, Argall, et al., 2004; Beauchamp, Lee, et al., 2004; Ghazanfar et al., 2005; Lemus et al., 2010; Stein & Stanford, 2008; von Saldern & Noppeney, 2013; Werner & Noppeney, 2011).

So far, coactivation models have developed a great variety of examples since Miller (1982), for instance, the Poisson counter model (Diederich, 1995; Schwarz, 1989), the parallel grains model (Miller & Ulrich, 2003), and the diffusion superposition model (Blurton et al., 2014; Chandrasekaran et al., 2019; Colonius & Diederich, 2020; Diederich, 1995; Diederich & Busemeyer, 2003; Schwarz, 1994). Shared by common assumptions, these models only differ in their mathematical description of the pooled evidence accumulation process. Both the Poisson counter and the parallel grains model assume that discrete impulses (or parallel grains) are generated in each modality. A multisensory channel pools these impulses together and submits the count to a decision module which triggers a response once the count reaches a criterion (Fig. 1f). By contrast, the diffusion superposition model assumes that the evidence accumulation in each modality follows a Wiener process. Evidence particle continuously drifts over time with a certain variance. Response is triggered when the particle reaches the absorbing boundary (decision criterion). The evidence accumulation process in multisensory conditions is the superposition of the two unisensory processes (Fig. 1g). A particular extension of the diffusion superposition model introduces a leak in the accumulation (Diederich, 1995; Diederich & Busemeyer, 2003), where evidence from the remote past leaks the most (i.e., an Ornstein-Uhlenbeck process, Fig. 1h). Notably, the Wiener process is recovered as the zero-leak limit of the OUP. Despite their distinct computational approaches, all the coactivation models also demonstrate excellent fit to the mean and variance of RT data collected in the redundant conditions.

One crucial issue is which class of models is superior at accounting for the RSE. Using generalized race model inequality tests, such as the survivor interaction contrast and workload capacity, researchers can distinguish between the underlying model architectures using the same set of RSE data (Blunden et al., 2022; Townsend & Ashby, 1983; Townsend & Nozawa, 1995; Yang et al., 2018). However, the two classes of models have rarely been parametrically evaluated and compared (but see (Blunden et al., 2022; Egan et al., 2025; Johansson & Ulrich, 2025)). And it remained unknown whether and how the established RSE principles manifest in the underlying computational mechanisms of RSE, or specifically, are the principles associated with the model-assumed processes?

1.3. The present study

There are four objectives of the present study. Firstly, whether the principles of RSE can be generalized to a dynamic multisensory context. All previously reported RSEs are constrained in a multisensory onset context (Innes & Otto, 2019; Mercier & Cappe, 2020; Otto et al., 2013; Otto & Mamassian, 2012; Plass & Brang, 2021). Considering real-world perceptual decisions often involve temporally dynamic signals that fluctuate over time, the present study consisting of three experiments created a dynamic multisensory context to mimic these naturalistic sensory fluctuations and to introduce novel dimensions for probing the generalization of RSE principles. In this context, not only onset but also offset changes of unisensory signals were included, and their cross-combination formed either congruent or incongruent multisensory changes (e.g., a flash onset paired with either a sound onset or offset). Participants completed a change detection task irrespective of modality and congruency.

Previous studies have indicated that transient onset and offset changes in sound intensity equally enhance visual detection and search for brightness changes irrespective of whether change types are congruent or not between modalities (Andersen & Mamassian, 2008; Van der Burg et al., 2010). A prior study demonstrated that auditory tone offset paired with visual stimulus onset produced a RT facilitation comparable to that with tone onset, confirming intersensory facilitation effects across both stimulus onset and offset context (Bernstein & Eason, 1970). Therefore, we expected the RSE occurs in the multisensory onset-offset context, and focused on revealing whether the observed RSE rank in this context conforms to the principles of RSE.

By pooling the observations of the three experiments together, the second objective of the present study was to evaluate which principle among the two most contributes to the observed RSE rank. Is there a dominant principle or a key predictor in the principle that explains RSE to a greater extent? The third objective was to compare the performance of the two classes of computational models of RSE, and identify the optimal one using a consistent evaluation metric. Lastly, our fourth objective was to reveal whether and how the (dominant) principle and predictor manifest in the optimal models. Specifically, do some processes in the optimal models vary as a function of the (dominant) predictors in the principles? A systematic evaluation of the principles and computations underlying the RSE in multisensory processing not only consolidates our current understanding of this effect but also shed light on potentially shared computational architectures across diverse multisensory integration phenomenon.

2. Experiments

2.1. Methods

Participants. A total of 52 participants was enrolled, with 20 in Experiment 1 (8 men, age: M = 25.3 years, SD = 3.2), 16 in Experiment 2 (8 men, age: M = 23.0, SD = 1.0), and 16 in Experiment 3 (10 men, age: M = 23.4, SD = 1.6). The sample size was determined based on previous studies on RSE (Innes & Otto, 2019; Otto et al., 2013; Otto & Mamassian, 2012). All but two participants were naive to the purpose of the experiment (Tan and Yuan, two of the authors, participating in Experiment 1). The participants had normal or corrected-to-normal vision and normal hearing, and gave written informed consent prior to the experiment. The current study was conducted in accordance with the Declaration of Helsinki and was approved by the institutional review board of the Institute of Psychology, Chinese Academy of Sciences (H21062).

Apparatus and stimuli. Testing was conducted in a dark, sound-attenuated room. Participants seated their heads on a chin rest at a viewing distance of 60 cm. Throughout the experiments, we used radial checkerboards as the visual stimuli, and white noises as the auditory stimuli (Fig. 2a). Each radial checkerboard consisted of 4 layers with a periodicity of 13, whose initial phase underwent random variations. The radial checkerboards were centrally presented against a grey background (~ 41 cd/m^2) on a LCD screen refreshed at 60 Hz. The visual angle of the radial checkerboard was kept quasi-equal throughout the experiments (2.98°×2.98° in Experiment 1, 2.61°×2.61° in Experiments 2 and 3). Its Michelson contrast in Experiment 1 was ~0.56, with the luminance at its darkest and brightest phase equalling ~19.71 cd/m^2 and ~69.45 cd/m^2, respectively. Its Michelson contrast in the other two experiments had two levels. The high and low contrasts were set to 0.57 (darkest: 19.71 cd/m^2 and brightest: 71.59 cd/m^2) and 0.15

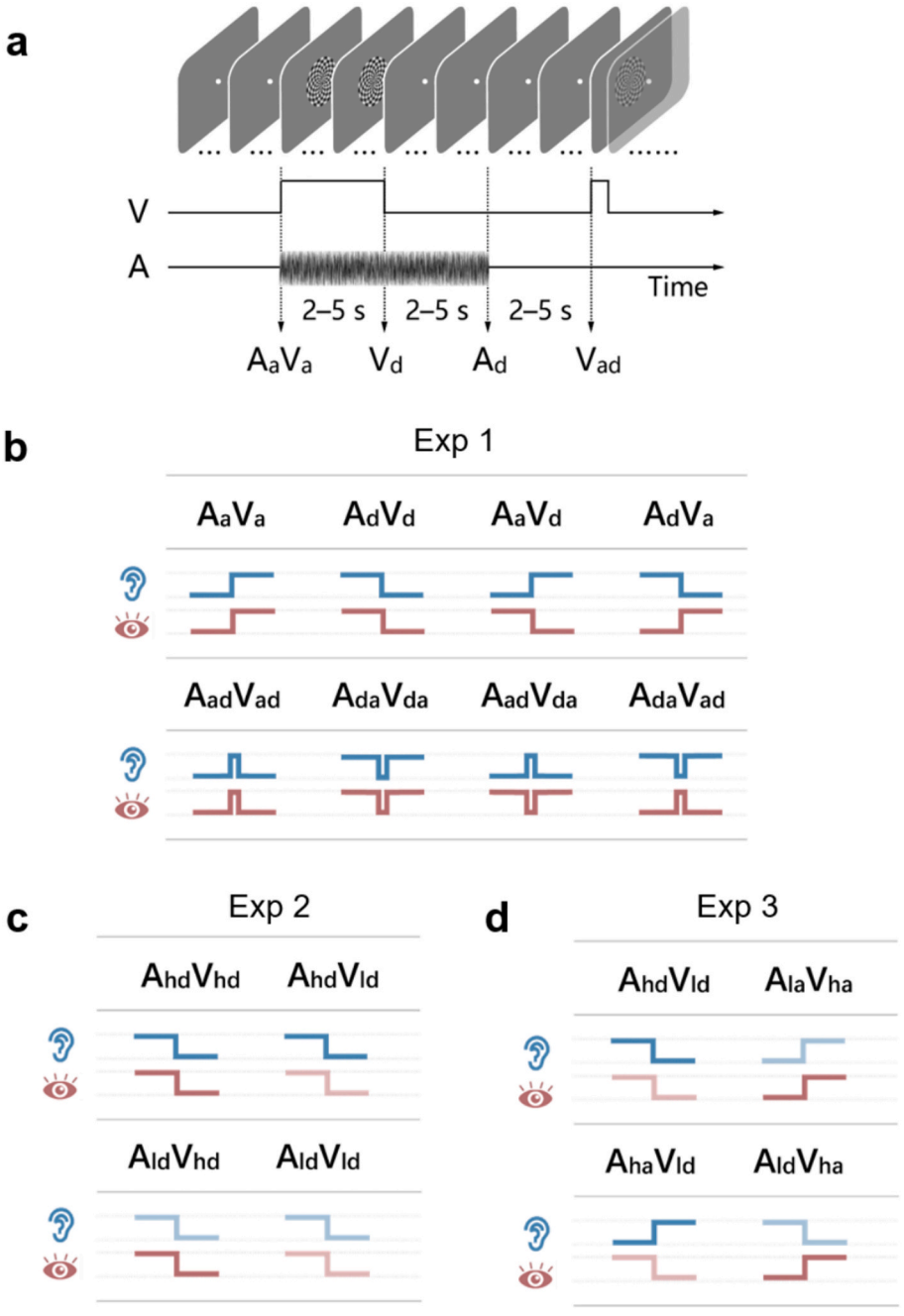

Fig. 2. The design and procedure of experiments. (a) Procedure of Experiment 1. Transient changes in auditory, visual, and audiovisual modalities occurred in a continuous stream. (b)–(d) Types of multisensory changes in each experiment. The blue and red lines represent the auditory and visual modality, respectively. The line brightness represents stimulus contrast/intensity. Brighter lines indicate higher visual contrast or auditory intensity.

(darkest: 35.11 cd/m^2 and brightest: 47.07 cd/m^2) in Experiment 2, and 0.73 (darkest: 13.58 cd/m^2 and brightest: 86.09 cd/m^2) and 0.15 (darkest: 35.11 cd/m^2 and brightest: 47.07 cd/m^2) in Experiment 3.

The white noises sampled at 44100 Hz were presented binaurally through headphones. In Experiment 1, the white noise was presented at an approximate level of 48 dB(A). Experiment 2 included white noises played at two distinct levels, categorized as high (46.5 dB(A)) and low (<30 dB(A), as sounds below 30 dB(A) cannot be accurately measured due to environment noises). Similarly, the white noise was manipulated to be of either high or low intensity (49 and 36 dB(A)) in Experiment 3. A small white fixation dot (0.36° × 0.36° in Experiment 1, 0.31° × 0.31° in Experiments 2 and 3) was displayed at the center of the screen throughout the whole experiments. All stimuli were generated by MATLAB (The MathWorks, Natick, MA) together with Psychtoolbox (Brainard & Vision, 1997; Pelli, 1997).

Design and procedure. To evaluate the two previously established principles, the study manipulated the unisensory RT distribution on purpose following Otto et al. (2013). Experiment 1 changed the closeness between unisensory RTs as well as their variability by manipulating the types of changes while leaving their strength constant. Experiments 2 and 3 did so by manipulating the strength of the auditory and visual stimuli.

In Experiment 1, visual or auditory stimuli could abruptly appear (onset), disappear (offset), suddenly appear-then-disappear (onset pulse) in 100 ms or disappear-then-appear (offset pulse) in 100 ms. Cross combinations of these distinctive unisensory changes resulted in eight types of multisensory changes (A_aV_a, A_dV_d, A_aV_d, A_dV_a, $A_{ad}V_{ad}$, $A_{da}V_{da}$, $A_{ad}V_{da}$, $A_{da}V_{ad}$, where *subscripts a* and *d* are short for *appear* and *disappear*, similarly *subscripts ad* and *da* are short for *appear-then-disappear* and *disappear-then-appear*; Fig. 2b). The eight types of multisensory changes with their corresponding unisensory changes were presented in pseudo-random order in a continuous stimulus stream (Fig. 2a). For instance, after visual and auditory stimuli concurrently appeared (A_aV_a), six candidates occurred in equal probability: the offset of the visual stimulus (V_d), of the auditory stimulus (A_d), or of both (A_dV_d); the offset pulse of

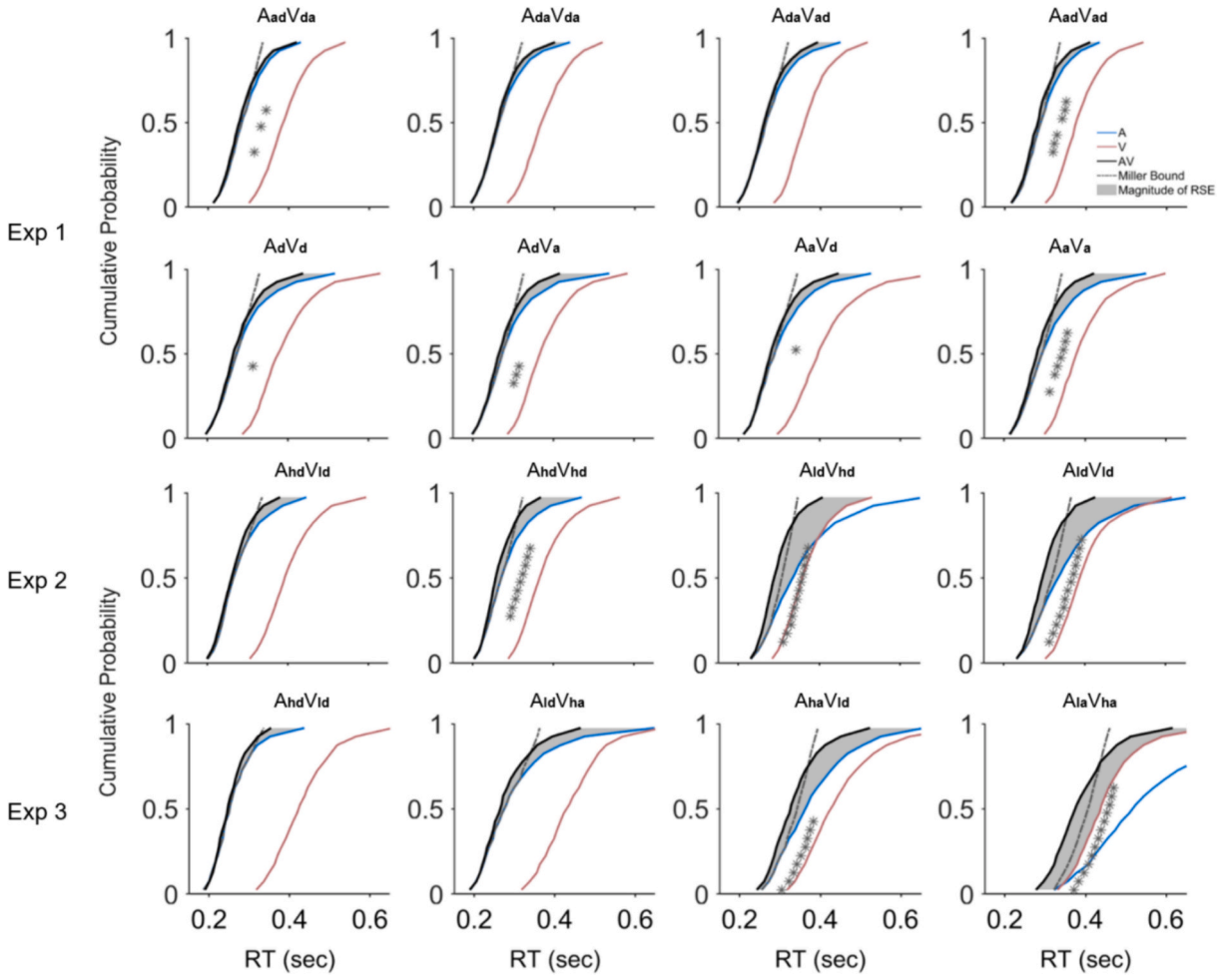

Fig. 3. Empirical RT CDFs for each type of changes in each experiments. The RT CDF for auditory (blue), visual (red) and audiovisual (black) modalities, along with the Miller's bound (grey dashed line) plotted for each type of multisensory changes. The grey area enclosed by the AV CDFs and fastest unisensory RT CDFs (the Grice's bound) represents the magnitude of RSE. *$p < 0.05$, which indicates that cumulative RT at the marked quantile in the redundant condition was significantly shorter than that predicted by the Miller's bound.

the visual stimulus (V_{da}), of the auditory stimulus (A_{da}), or of both ($A_{da}V_{da}$). There was a random interval of 2–5 secs between consecutive two changes. Participants were instructed to maintain on the fixation dot and detect the changes occurring in either visual, auditory or both modalities by pressing a button as quickly as possible. In total, there were 2400 trials divided into 20 blocks, with 100 trials for each change. Participants received 10 blocks per day, and finished the experiment in 2 separate days.

Experiment 2 tested the RSE when stimuli abruptly disappeared (stimulus offset). The contrast of visual stimuli and the intensity of auditory stimuli were set at two different levels, resulting in four combinations of multisensory changes ($A_{hd}V_{hd}$, $A_{hd}V_{ld}$, $A_{ld}V_{hd}$, $A_{ld}V_{ld}$, where the additional *subscripts d* is short for *disappear, h* and *l* denote *high* and *low* contrast/intensity) and eight corresponding unisensory changes (Fig. 2c). Usually, RTs are slowed with a larger variability for stimuli with low contrast/intensity compared to those with high contrast/intensity. For example, combining high-contrast visual stimuli and low-intensity auditory stimuli in the $A_{ld}V_{hd}$ condition would narrow their RT discrepancy but increase the variability compared with combining high-contrast visual stimuli and high-intensity auditory stimuli in the $A_{hd}V_{hd}$ condition. At the beginning of each trial, the audiovisual stimuli were immediately presented, and lasted for a duration of 1–3 secs before one or two of them sudden disappeared. Participants were required to detect the stimulus disappearance as quickly as possible. For unisensory changes, the stimulus in the other modality would also be cleared after participants responded to the target modality. The next trial start after a 1-sec blank. The experiment consisted of 1080 trials, with 90 trials for each change. All trials were randomized and organized in 10 blocks.

Experiment 3 included both onset and offset stimulus changes and set both the contrast of visual stimuli and the intensity of auditory stimuli at two levels. Four combinations of multisensory changes were created by cross combinations of distinctive unisensory changes ($A_{hd}V_{ld}$, $A_{la}V_{ha}$, $A_{ha}V_{ld}$, $A_{ld}V_{ha}$, where the additional *subscripts h* and *l* denote *high* and *low* contrast/intensity and subscripts *a*

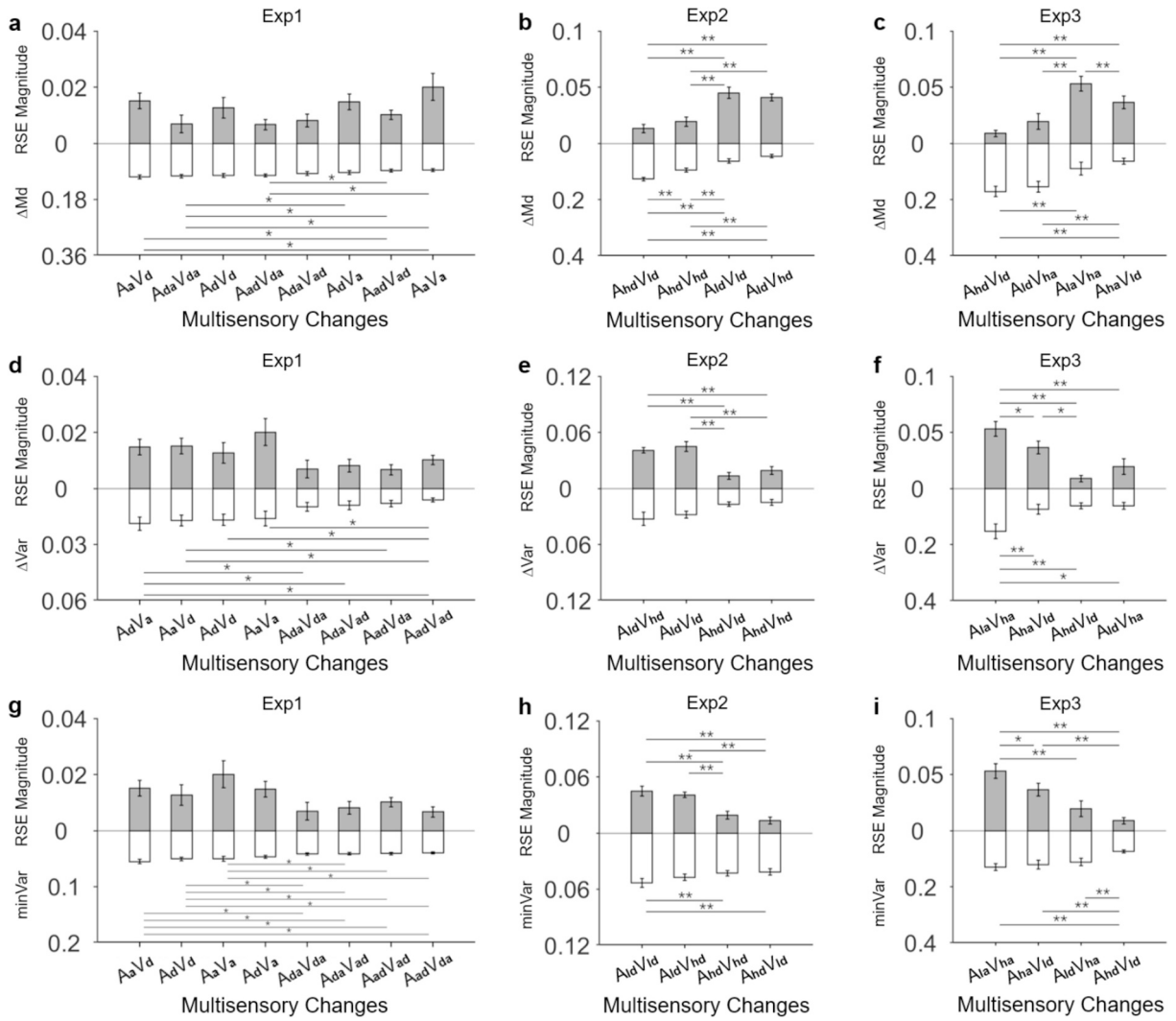

Fig. 4. Relationship between ΔMd, ΔVar, minVar and RSE magnitude in all experiments. The covariation between RSE magnitude and ΔMd (a)–(c), between RSE magnitude and ΔVar (d)–(f), between RSE magnitude and minVar (g)–(i) in the three experiments. The types of multisensory changes were sorted in a descending order according to each predictor in the panels. Error bars indicate the standard error. * $p < 0.05$, ** $p < 0.01$.

and *d* denote *appear* and *disappear;* Fig. 2d). In Experiment 3, the onset and offset changes with different contrast/intensity were smoothly transformed to construct a continuous stimulus stream. For the offset changes, the stimulus gradually ramped up to its full contrast/intensity over 1.5 secs, and then suddenly disappeared after maintaining on for 0.5–1.5 secs, followed by a 1.5-sec blank. Conversely, for the onset changes, the stimulus abruptly appeared after a blank of 2–3 secs, and then gradually faded over 1.5 secs. Participants were told and practiced to only respond to those abrupt changes irrespective of modalities. There were totally 1080 trials, organized in 10 blocks, each of 90 trials.

Analysis and statistics. Trials with RTs outside the range of 100–2000 ms and those with missed responses were excluded from the dataset of each participant. On average, the percentage of excluded trials was 0.98 ± 0.91%.

Race model inequality test. As reviewed in the *introduction*, the race model inequality test (the Miller's bound, Eq. (2)) is routinely run to evaluate whether the RT facilitation for multisensory stimuli can be predicted by probability summation of the unisensory RT distributions (a statistical facilitation). We performed such test following the approach described in several literature (Gondan & Minakata, 2016; Miller, 1982; Otto, 2019; Otto et al., 2013; Otto & Mamassian, 2016; Plass & Brang, 2021; Stevenson et al., 2014). Specifically, for each participant, we sorted their RTs for each type of changes in ascending order and segmented these RTs into 20 quantiles to obtain the CDF. Then, we compared the observed CDF in multisensory conditions (black solid line in Fig. 3) and the calculated Miller's bound (grey dash line in Fig. 3) at each quantile using a T-max permutation test (Gondan, 2010; Shaw et al., 2020). The tests were conducted using the PERMUTOOLS, an open-source package available at https://github.com/mickcrosse/PERMUTOOLS. If the audiovisual RTs empirically observed are significantly faster than that theoretically derived at one quantile at least, it can be concluded that the RSE in the multisensory condition cannot be explained by statistical facilitation but some form of interaction between modalities.

Quantification of the RSE magnitudes. We quantified the magnitude of the RSE for each type of multisensory changes by the area (Eq. (3)) enclosed by its empirical CDF of RT (black lines in Fig. 3) and the faster out of the two corresponding unisensory changes (Eq. (4), documented as the Grice's bound) (Gondan & Minakata, 2016; Grice et al., 1984; Otto, 2019; Otto & Mamassian, 2016; Stevenson et al., 2014):

$$RSE\,magnitude = \int (P_{AV}(t) - P_{Grice}(t))dt \tag{3}$$

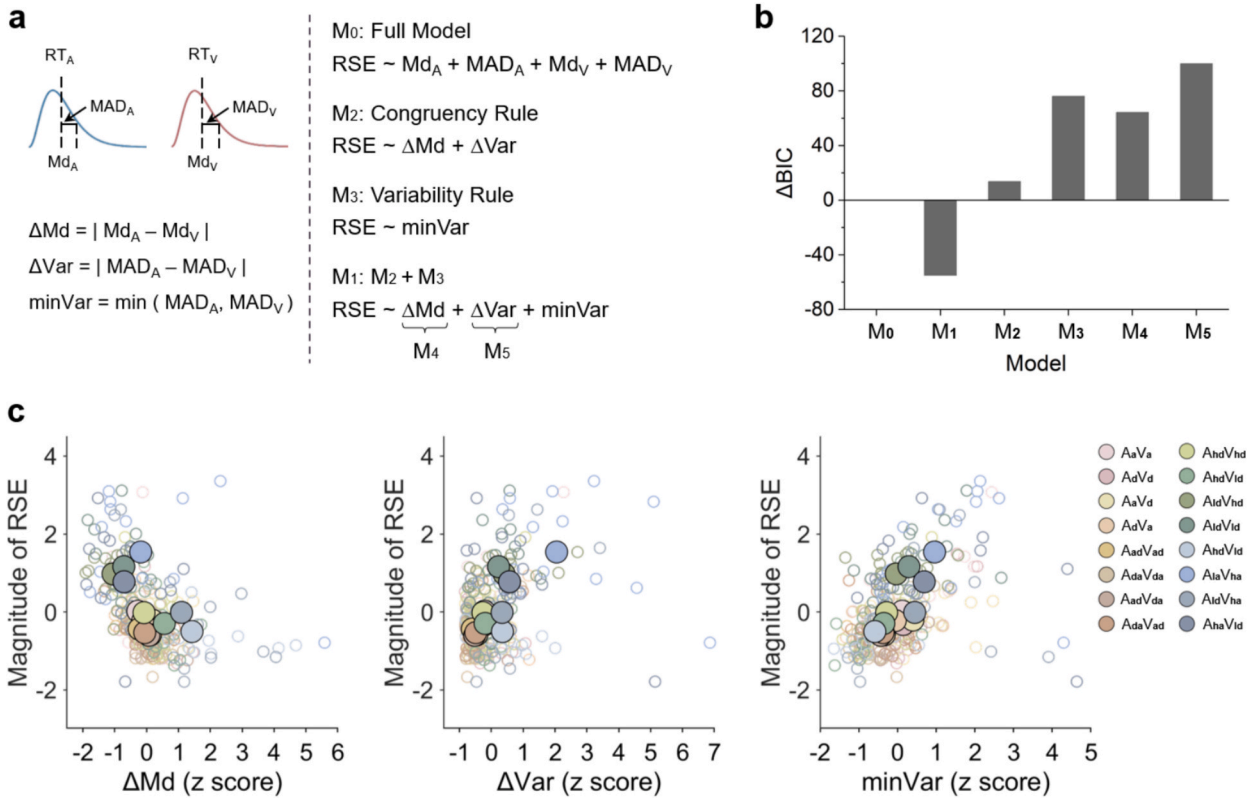

Fig. 5. The results of LMM analysis for all experiments. (a) The overview of the LMMs. The blue and red lines represent RT distribution for auditory and visual stimuli, respectively. Predictor ΔMd, ΔVar, minVar, and their combinations formed six LMMs. (b) ΔBIC of the other five LMMs compared to the M_0 model. The smaller the ΔBIC, the better the model performance. (c) The correlation between RSE magnitude and ΔMd, ΔVar, minVar (z score). The large solid circles represent the mean RSE magnitude for each type of multisensory changes, while the small hollow circles represent the individual RSE magnitude.

$$P_{Grice}(t) = \max[P_A(t), P_V(t)] \quad (4)$$

Quantification of the principle predictors. Three predictors were calculated according to the two derived principles in Otto et al. (2013). The absolute differences of median RT (Md) and median absolute deviation (MAD, another measure of variability) between the two unisensory RT distributions (ΔMd and ΔVar) was utilized to measure their similarity as the principle of congruent effectiveness indicates, while the minimum of their MAD (minVar) was utilized to represent the lower boundary of their variability as the variability rule indicates.

Repeated-measures Analysis of Variances (rmANOVA). To examine whether the RSE magnitude varies across multisensory stimuli as the two principles predict at the group level, we conducted two one-way rmANOVAs for each experiment. The first one tested whether the predictor values significantly differ across multisensory stimuli, while the second tested whether the RSE magnitude significantly differs across multisensory stimuli. Their covariation observed in the two rmANOVAs and corresponding post-hoc analyses were illustrated in Fig. 4. For all ANOVA results, Greenhouse–Geisser corrected p values were reported when the sphericity assumption was violated. All the post-hoc analyses were corrected by the False Discovery Rate (FDR) method, with a significance threshold set at $\alpha = 0.05$. The statistical analysis was performed in MATLAB and SPSS (IBM, Armonk, NY).

Linear mixed model analysis. Linear mixed models (LMMs) were employed to quantitatively assess the contribution of the two principles that account for the rank of the RSE magnitudes across experiments. A total of six LMMs were constructed with separate predictors representing the RSE principles, as listed in Fig. 5a and Table 1. By comparing the model predictability, we could discover which principle and predictor dominates the other. The LMM analysis were performed in MATLAB using the *fitglme* function, with participants as random effect and all the other effects fixed. The predictor variables were normalized before entering into the LMMs.

According to probability summation framework, it is assumed that the RSE magnitude in multisensory cases can be predicted given all the information of the unisensory RTs (Raab, 1962). Hence, the reference model M_0 included all the parameters that characterized the auditory and visual RT distributions (Md_A, MAD_A, Md_V, MAD_V) as predictor variables. Other three predictor variables came from the principle of congruent effectiveness and the variability rule. The M_1 model included ΔMd, ΔVar, and minVar to test the two principles as a whole, while the M_2 and M_3 models focused on the principle of congruent effectiveness (ΔMd and ΔVar) and variability (minVar) rule separately. The M_4 and M_5 models independently assessed the predictability of ΔMd and ΔVar, respectively. The multicollinearity among the three predictors (ΔMd, ΔVar, and minVar) was measured using variance inflation factors (VIF). A VIF value of 1 indicates no multicollinearity and a value below 5 indicates not severe, acceptable multicollinearity. The Bayesian information criterion (BIC) for each model was calculated. A lower BIC score signifies a model with superior performance. The most influential principle and predictor was the one that minimally inflated the BIC of the M_0 model.

3. Results

3.1. Robust RSEs in the dynamic multisensory context irrespective of congruency

Fig. 3 illustrated the CDF of observed RTs in both single and redundant conditions of the three experiments. The race model inequality test showed that the RT CDF from most multisensory changes (the black solid line) significantly surpasses the Miller's bound at one quantile at least (the grey dash line), concluding that the fastened reaction to multisensory stimuli compared to their unisensory components cannot be explained by statistical facilitation but by some form of crossmodal interaction. The results thus demonstrated robust RSEs for most multisensory changes, irrespective of congruent or incongruent.

3.2. The principle of congruent effectiveness and the variability rule in explaining the RSE at the group level

To check whether the RSE rank yielded by distinct types of multisensory changes is concordant with the previously derived two principles (Otto et al., 2013), we examined whether the predictors in the principle of congruent effectiveness and variability rule significantly varied across conditions, and whether the RSE magnitude significantly covaried with the predictors.

The principle of congruent effectiveness is expressed by the similarity in median and variability between the two unisensory RT distributions (ΔMd and ΔVar), while the variability rule is expressed by the lower boundary of the two unisensory RT variability (minVar). For all the three predictors in the two principles, significant effects were found by one-way repeated-measures ANOVAs

Table 1
The results of linear mixed models.

Models	Predictor (β)			BIC	adjusted R^2
	ΔMd	ΔVar	minVar		
M_0				691.81	0.51
M_1	−0.52**	0.26**	0.42**	636.86	0.58
M_2	−0.50**	0.40**		705.48	0.44
M_3			0.48**	767.95	0.29
M_4	−0.52**			756.13	0.35
M_5		0.39**		791.78	0.21

** $p < 0.01$.

across multisensory changes for all experiments (for ΔMd: $Fs > 4.82$, $ps < 0.003$, $\eta_p^2 > 0.20$; for ΔVar: $Fs > 3.50$, $ps < 0.011$, $\eta_p^2 > 0.16$; for minVar: $Fs > 6.41$, $ps < 0.003$, $\eta_p^2 > 0.25$). All significant pairwise comparisons were shown in Fig. 4, the white bars.

The magnitudes of the RSE were quantified as the distance between the multisensory CDF and the Grice's bound (the grey area in Fig. 3), and compared across the types of multisensory changes by the same one-way repeated-measures ANOVA for each experiment, respectively. Significant effects were found for all experiments (Experiment 1: $F(7, 133) = 2.63$, $p = 0.040$, $\eta_p^2 = 0.12$; Experiment 2: $F(3, 45) = 28.90$, $p < 0.001$, $\eta_p^2 = 0.66$; Experiment 3: $F(3, 45) = 15.44$, $ps < 0.001$, $\eta_p^2 = 0.51$). Although not all pairwise comparison reached significance after FDR correction, the rank of RSE magnitudes appeared to be predicted by the rank of two predictors (ΔMd and minVar).

As predicted by the principles of congruent effectiveness, when the central tendency of RTs from the two unisensory conditions was more similar (conditions with smaller ΔMd), the RSE magnitude was larger (Fig. 4a–c). The similarity in unisensory RT variances (ΔVar) is expected to negatively correlate with the RSE magnitude as the principle of congruent effectiveness predicted, but a positive relationship was illustrated (Fig. 4d–f). Regarding the variability rule, an expected positive relationship between minimal unisensory RT variance (minVar) and the RSE magnitude across the three experiments can been seen from Fig. 4g to i. Overall, the results demonstrated the effectiveness of the two principles in predicting the magnitude of RSE in the dynamic multisensory context. Given the weak yet observable correlations among the three predictors across conditions at the group level (especially in Experiments 2 and 3), we further conducted a LMM analysis to assess their contribution and examine which one is dominant in explaining the RSE at the individual level.

3.3. *The principle of congruent effectiveness and the variability rule in explaining the RSE at the individual level*

To assess the contribution of the previously derived two principles (Otto et al., 2013) and its separate predictors in explaining the RSE rank observed in the dynamic multisensory context, six LMMs were constructed using data from individual observation. Each LMM represents a separate principle or predictor. Table 1 and Fig. 5a showed all the predictors in the six LMMs. To summarize, model M_0 included all the parameters from the two unisensory RT distributions (Md_A, MAD_A, Md_V, MAD_V), serving as a reference model. The principle of congruent effectiveness expressed by ΔMd and ΔVar is submitted to M_2 model, while the variability rule expressed by minVar is submitted to M_3 model. The M_1 model included the ΔMd, ΔVar and minVar to test the predictability of the two principles together, whereas the M_4 and M_5 each contained ΔMd or ΔVar to elucidate their independent predictability.

The VIF value ranged from 1.22 to 1.48, indicating no significant multicollinearity among the three predictors in our LMMs. Table 1 and Fig. 5b–c display the results. First, the M_1 was the best performed models with the lowest BICs, even better than M_0 that included the original two unisensory RT distributions. It demonstrates that the the principle of congruent effectiveness and the variability rule,

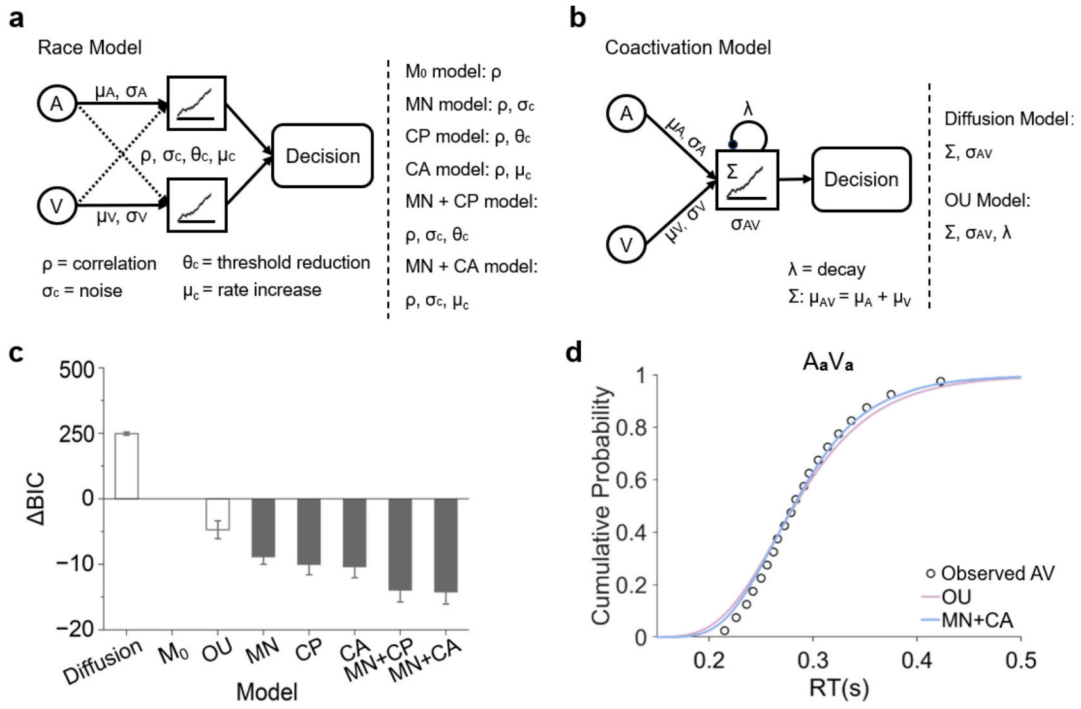

Fig. 6. Schematic diagrams and modelling results. (a) Schematic diagram of the six context-variant race models and their parameters. (b) Schematic diagram of the two coactivation modes and their parameters. (c) ΔBIC of all eight models compared to M_0 model. Context-variant race models and coactivation models are colored grey and white, respectively. (d) The predictive curves of the best models in the context-variant race model class (MN + CA model) and in the coactivation model class (OU model), and the observed CDFs, averaged across participants in the A_aV_a condition.

derived from the unsensory RT distributions, provide excellent predictive power for the magnitude of RSE. Second, the variability rule plays a less important role in predicting the RSE, evidenced by a relatively large BIC for M_3, and a relatively small deterioration when minVar was excluded from M_1 ($BIC_{M2}-BIC_{M1} = 68.62$). By contrast, the principle of congruent effectiveness owns a higher predictive power, as it performed better than the variability rule ($BIC_{M2}-BIC_{M3} = -62.47$) and removing it severely reduced the model performance ($BIC_{M3}-BIC_{M1} = 131.09$). The LMM analysis further revealed that ΔMd weights more than ΔVar in the principle of congruent effectiveness ($BIC_{M4}-BIC_{M5} = -35.65$).

To conclude, the principle of congruent effectiveness dominates the variability rule in explaining the RSE rank in our dynamic multisensory context; among the three independent predictors in the principle, the closeness between the central tendency of unisensory RTs (ΔMd) dominates the other two. Noteworthy, the prediction of ΔVar at the individual level also contradicts with the prediction by the principle of congruent effectiveness, the same to the findings at the group level.

4. Computational modelling

Having confirmed the generalization and dominance of RSE principles in the dynamic multisensory context, we examined whether and how the principle reflects in the computational models proposed to explain the RSE. As reviewed in the *Introduction*, there are two main classes of computational models, the context-variant race model and the coactivation model, each developing different mechanisms to explain the RSE (Colonius & Diederich, 2020; Diederich, 1995; Diederich & Busemeyer, 2003; Otto & Mamassian, 2012; Plass & Brang, 2021). Here using data from all the three experiments, we constructed six representative context-variant race models and two representative co-activation models to evaluate their performance, and explored whether and how the principles are linked to the optimal model.

4.1. The procedure of model fitting

The six context-variant race models we constructed each assumes distinct types of interaction between unisensory changes (Fig. 6a, (Otto & Mamassian, 2012; Plass & Brang, 2021)). First of all, in the framework of evidence accumulation, the evidence accumulation process in the single condition can be simply approximated using the Linear Approach to Threshold with Ergodic Rate (LATER) model (Noorani & Carpenter, 2016; Reddi et al., 2003):

$$RT_i \sim \frac{1}{N\left(\mu_i, \sigma_i^2\right)}, \text{ where } i = A \text{ or } V \tag{5}$$

With initial evidence state 0 and the decision threshold 1, the RT can be loosely interpreted as a reciprocal of the evidence accumulation rate (or drift rate) that follows a normal distribution with mean μ and variance σ^2. For all subclass of race models, the RT in redundant condition is consistently determined as the minimum of the decision time in the single conditions, which corresponds to the reciprocal of the maximum of the two drift rates in the single conditions:

$$RT_{AV} \sim \frac{1}{\max[N(\mu_A, \sigma_A^2), N(\mu_V, \sigma_V^2)]} \tag{6}$$

where the maximum of two normal distributions on the denominator is analytically solved following Eq. (1) and (2) in (Nadarajah & Kotz, 2008). Readers can also refer to Eq. (S1) and (S2) in Otto & Mamassian (2012)'s supplementary information, and Eq. (10) in Yang et al. (2018).

The six context-variant race models are built on Eq. (5) with the parameters (μ_A, μ_V, σ_A, and σ_V) fixed for deriving multisensory RT distributions through solving Eqs. ((6)–(11)) in the same manner. Of the six models, the reference model M_0 assumes a correlation between two normal distributions of unisensory evidence accumulation rates (parameter ρ in Eqs. (S1) and (S2) in Otto & Mamassian (2012)). It is the only free parameter that awaits optimization.

The MN model includes another free parameter additional noise (σ_c) into the M_0 model (Otto & Mamassian, 2012), which allows the variability of the evidence accumulation rates for unisensory components to vary in the redundant condition:

$$RT_{AV} \sim \frac{1}{\max\left[N\left(\mu_A, (\sigma_A + \sigma_c)^2\right), N\left(\mu_V, (\sigma_V + \sigma_c)^2\right)\right]} \tag{7}$$

By contrast, the CP model adds an alternative free parameter (θ_c) into the M_0 model to represent a lowered decision criterion, while the CA model adds an alternative free parameter (μ_c) into the M_0 model to represent an increased evidence accumulation rate:

$$RT_{AV} \sim \frac{1 - \theta_c}{\max[N(\mu_A, \sigma_A^2), N(\mu_V, \sigma_V^2)]} \tag{8}$$

$$RT_{AV} \sim \frac{1}{\max[N(\mu_A + \mu_c, \sigma_A^2), N(\mu_V + \mu_c, \sigma_V^2)]} \tag{9}$$

The remaining two models, MN + CP and MN + CA, each simulates two types of interaction, with the multisensory RT distribution can be estimated as:

$$RT_{AV} \sim \frac{1-\theta_c}{\max\left[N\left(\mu_A,(\sigma_A+\sigma_c)^2\right),N\left(\mu_V,(\sigma_V+\sigma_c)^2\right)\right]} \tag{10}$$

$$RT_{AV} \sim \frac{1}{\max\left[N\left(\mu_A+\mu_c,(\sigma_A+\sigma_c)^2\right),N\left(\mu_V+\mu_c,(\sigma_V+\sigma_c)^2\right)\right]} \tag{11}$$

For MN + CP model, free parameters are ρ, σ_c and θ_c; For MN + CA model, free parameters are ρ, σ_c and μ_c.

The two coactivation models we constructed are the diffusion superposition model and the Ornstein-Uhlenbeck (OU) model (Fig. 6b, (Colonius & Diederich, 2020; Diederich, 1995; Diederich & Busemeyer, 2003)). The diffusion superposition model assumes a straightforward sum of the evidence accumulation across unisensory modalities (Eqs. (12) and (13)), while the OU model incorporates a leaky integration of the sum (leak parameter λ, in Eq. (14)):

$$\mu_{AV} = \mu_A + \mu_V \tag{12}$$

$$dX(t) = \mu_{AV} + \sigma_{AV} dW(t) \tag{13}$$

$$dX(t) = \mu_{AV} - \lambda X(t) + \sigma_{AV} dW(t) \tag{14}$$

where μ_{AV} is the evidence-accumulation drift rate for the multisensory accumulator, which is the sum of unisensory drifts μ_A and μ_V (Eq. (12)). The σ_{AV} is the diffusion coefficient controlling trial-to-trial variability of the accumulation process. And $W(t)$ indicates the standard Wiener process. The same to the race models, the decision or absorbing boundary is fixed at 1 and the accumulation starts from 0 in the two coactivation models. The evidence accumulation process for single modalities can also be simulated by Eq. (13) but with only one drift rate left (μ_A, *or* μ_V) and with the diffusion coefficients replaced with σ_A, *or* σ_V. With the above parameter set, both the unisensory and multisensory RT distributions generated by the diffusion superposition model (Eq. (13)) have analytic solution, equalling to a Wald distribution or an inverse Gaussian distribution (Chandrasekaran et al., 2019):

$$f(t|\mu_{AV},\sigma_{AV}^2) = \frac{1}{\sqrt{2\pi\sigma_{AV}^2 t^3}} \times \exp\left(-\frac{(1-\mu_{AV}t)^2}{2\sigma_{AV}^2 t}\right) \tag{15}$$

$$F(t|\mu_{AV},\sigma_{AV}^2) = \phi\left(\mu_{AV}t|\sigma_{AV}^2 t\right) + \exp\left(\frac{2\mu_{AV}}{\sigma_{AV}^2}\right)\phi\left(-\mu_{AV}t|\sigma_{AV}^2 t\right) \tag{16}$$

here *t* denotes the first-passage time of the evidence particle reaching the absorbing boundary (i.e., RT). The *f* is the probability density function of RT, while *F* is the cumulative probability function of RT. Φ denotes the cumulative normal distribution, evaluated at μ_{AV}t with variance $\sigma_{AV}^2 t$. Note that we did not allow $\sigma_{AV}^2 = \sigma_A^2 + \sigma_V^2 + 2\rho_{AV}\sigma_A\sigma_V$ in the two coactivation models and did not include a motor variance, which differs from previous studies (Blurton et al., 2014; Chandrasekaran et al., 2019; Schwarz, 1994). The reason for this choice was stated in the *Discussion*.

The multisensory RT distributions generated by the OU model are numerically approximated using Eqs. (39), (40), (59), and (28) in (Smith, 2000). When deriving the RT distribution in the redundant conditions, we fixed two unisensory parameters μ_A and μ_V as did for the race models. Thus, for the diffusion superposition model, only the parameter σ_{AV} is free, and for the OU model, the two free parameters are σ_{AV} and λ.

Specifically, the model fitting procedure had been divided into two steps. First, we fitted the corresponding unisensory RT distributions with the LATER model (Eq. (5)) or the diffusion model (Eq. (13), but entered the unisensory drift and diffusion coefficients) for each type of multisensory changes and each participant, and acquired their best parameters ($\mu_A, \mu_V, \sigma_A, \sigma_V$). Subsequently, we fitted the multisensory RT distributions for the six context-variant race models and the two coactivation models. As described above, we fixed the four unisensory parameters throughout the fitting of multisensory RT distributions for the eight models. This pipeline can identify the model that best predicts the multisensory responses given the unisensory responses, and ensure that the model performance reflects genuine cross-modal interactions rather than re-scaling of unisensory parameters. Multisensory RT distributions predicted by the context-variant race models were generated by Eqs. ((6)–(11)) using Eqs. (S1) and (S2) in Otto & Mamassian (2012) with customized scripts adapted from the RSE-box (Otto, 2019). Multisensory RT distributions predicted by the diffusion superpostion model were generated by Wald distribution, while multisensory RT distributions predicted by the OU model were generated by numerical algorithm in (Smith, 2000) using custom scripts. The link of all the codes used in the modelling can be found in the *Data and Code Availability*.

Given the RTs in the redundant conditions were generated differently in the race models versus coactivation models, and consequently some have analytic solutions while others do not, we opted not to use maximum likelihood estimation to optimize the model performance. Instead, we adopted an alternative approach commonly employed in RT modelling (Johansson & Ulrich, 2025; Ratcliff & Smith, 2004): A statistic G^2 was calculated to minimize the error between the observed and predicted RT distributions:

$$G^2 = 2\sum\nolimits_i Np_i \ln\left(\frac{p_i}{\pi_i}\right) \tag{17}$$

where N is the number of observations grouped into bins, p_i is the proportion of the observations in the ith bin and π_i is the proportion in the bin predicted by the model (Ratcliff & Smith, 2004). In our fits, RTs from each type of multisensory change and each participant in all three experiments were segmented into 20 quantiles for both its empirical and predicted proportions. For simplicity, we set N to 100. We utilized the *fmincon* function in MATLAB to maximize the G^2. Specifically, we provided a plausible range of each parameter, and selected five parameter combinations that generated the highest G^2 scores among several random combinations within the range. We then used these five combinations as initial points to maximize G^2. The parameter combination with the highest G^2 scores was finally defined as the best parameters for the model fitted.

To determine the most suitable model for interpreting the data, we computed the BIC score for each model, which was transferred from G^2 using the following formula:

$$BIC = G^2 - 2\sum_i Np_i \ln(p_i) + M\ln(N) \tag{18}$$

where p_i and π_i are the same as in Eq. (17) and M represents the number of free parameters in the model. We then averaged the BIC scores across types of multisensory changes and participants to obtain a mean score for each model (Table 2). For further clarification, we calculated ΔBIC, which represents the difference between the BIC scores of other candidate models and the M_0 model (Fig. 6c). We also calculated Bayesian Factor (BF) between any two models to further assess the significance of their differences. A BF score greater than 3 or less than 1/3 indicates a significant difference in fitting performance between them (Table 3).

To examine the variability of the fitness of the eight candidate models, we performed two more analysis. First, we checked whether the best model varies individually (Fig. 7a). We averaged the BICs across types of multisensory changes for each model and each participant, and calculated the BIC weights:

$$\Delta_i(BIC) = BIC_i - BIC_{\min} \tag{19}$$

$$W_i(BIC) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(BIC)\right\}}{\sum_{k=1}^{k}\exp\left\{-\frac{1}{2}\Delta_k(BIC)\right\}} \tag{20}$$

where $BIC_{\min}$ represents the lowest BIC among all models for each participant, and $W_i(BIC)$ represents the probability that Model i is the best fit for each participant given the data and the set of candidate models, and k denotes the total number of models (Wagenmarkers & Farrell, 2004).

Second, we examined whether the best model varies with the types of multisensory changes (Fig. 7b). The analysis procedure is the same as the one described above, except that we averaged BIC across participants instead of types of multisensory changes. The $BIC_{\min}$ now represents the lowest BIC among all models for each type of multisensory changes, and the $W_i(BIC)$ represents the probability that Model i is the best fit for each type of multisensory changes given the data and the set of candidate models.

5. Results

5.1. The RSE can be best accounted for by two context-variant race models

Our model fitting results showed that the context-variant race models performed better than the co-activation models (Fig. 6c, Tables 2 and 3). Among the six context-variant race models, MN + CP and MN + CA were superior to others, while within the two coactivation models, OU was the best. Fig. 6d plotted the fitting curves of the MN + CA model and the OU model for the A_aV_a condition as an example, which graphically illustrates that the MN + CA model provided a better fitness compared to the OU model (BIC_{MN+CA}-$BIC_{OU} = -9.53$, BF = 117.39). We also assessed the variability of the model fitness for each model. The BIC weight was calculated, which represents the probability for one model to be the best fit given the data from all candidate models. It was found that the MN +

Table 2
Best fit parameters for all computational models.

Models	Fixed parameter				Parameters						BIC
	μ_A	σ_A	μ_V	σ_V	ρ	σ_c	θ_c	μ_c	σ_{AV}	λ	
Race Model	3.46	0.72	2.60	0.42							
M_0					0.09						696.17
MN					0.04	0.06					687.28
CP					0.24		0.03				686.12
CA					0.36			0.13			685.76
MN + CP					0.51	−0.06	0.04				682.20
MN + CA					0.55	−0.05		0.14			681.92
Coactivation Model	3.29		2.52								
Diffusion									1.29		945.04
OU									0.33	3.99	691.45

Table 3
The Bayes factors of each pair of computational models.

Models	Bayes factors						
	MN	CP	CA	MN + CP	MN + CA	OU	Diffusion
M_0	0.01 [s]	0.01 [s]	0.01 [s]	$9.25e^{-4s}$	$8.06e^{-4s}$	0.09[s]	$1.10e^{+54s}$
MN		0.56	0.47	0.08[s]	0.07[s]	8.07[s]	$9.41e^{+55s}$
CP			0.84	0.14[s]	0.12[s]	14.40[s]	$1.68e^{+56s}$
CA				0.17[s]	0.15[s]	17.18[s]	$2.00e^{+56s}$
MN + CP					0.87	102.25[s]	$1.19e^{+57s}$
MN + CA						117.39[s]	$1.37e^{+57s}$
OU							$1.17e^{+55s}$

[s] Bayes factor greater than 3 or less than 1/3 indicates a significant difference between the two models.

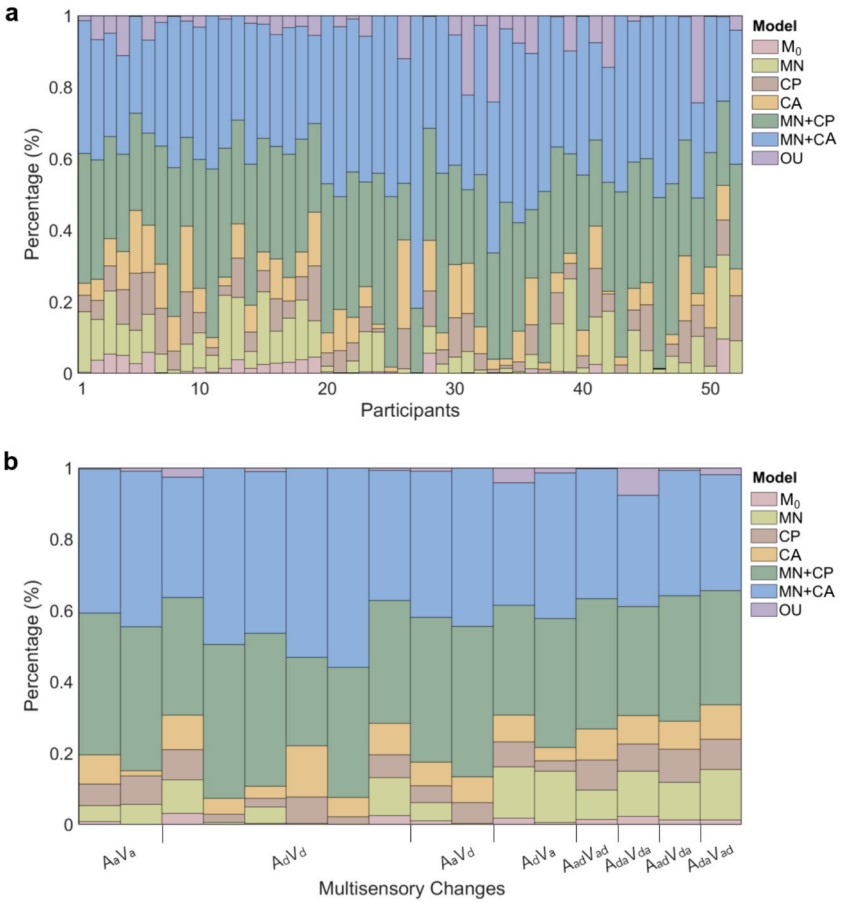

Fig. 7. The BIC weights of candidate computational models. (a) The BIC weight of each candidate model computed and drawn for each participant. (b) The BIC weight of each candidate model computed and drawn for each type of multisensory changes. Note, the diffusion model was not included due to its extremely small BIC weight; some multisensory types were tested multiple times in different experiments, while others were tested once only in one experiment. Each color represents one candidate model. Indicated by the bar area, the MN + CP and MN + CA models are the best two models, not only consistent across participants but across types of multisensory changes.

CP and MN + CA models are the top two models not only for almost every participant (Fig. 7a) but also for every type of multisensory changes (Fig. 7b). The results of BIC weight analysis substantiated the stability of model fitness.

5.2. The predictors in the two principles closely links with the crossmodal interactive processes in the optimal models

To investigate whether and how the principles of RSE manifest in the optimal computational models (MN + CP and MN + CA), we further quantified the relationships between parameters in the models (θ_c, μ_c, and σ_c) and predictors in the principles of RSE (ΔMd, ΔVar, and minVar). How these parameters were modulated by the predictors were assessed across change types using LMM analysis

similar to the aforementioned. A significant coefficient may inform us of the model processes that the predictors in the principles modulates, which may in turn lead to RT facilitation in redundant conditions. As shown in Table 4, we found that for the two best-fitting race models (MN + CP and MN + CA models), the parameter σ_c was positively modulated by ΔVar ($\beta = 0.43/0.42$, $ps < 0.01$), while negatively modulated by minVar ($\beta = -0.36/-0.29$, $ps < 0.01$). The parameter θ_c in the MN + CP model and the parameter μ_c in the MN + CA model was both negatively modulated by ΔMd ($\beta = -0.47/-0.48$, $ps < 0.01$), and positively modulated by minVar ($\beta = 0.31/0.25$, $ps < 0.01$). The relationships between predictors and model parameters in the MN + CA model were illustrated in Fig. 8. Noteworthy, in the two models, parameters μ_c and θ_c were more effectively modulated by the three predictors especially by ΔMd and minVar than parameter σ_c in terms of predictive power (refer to BIC and adjusted R^2 in Table 4). In short, the results indicate that when two signals have more similar central tendencies or larger variability in their RTs, combining them together in a redundant condition may further lower the decision criterion or increase the accumulation rate.

Although coactivation models were not the best fits, we found in the OU model, the leakage of evidence (λ) was negatively modulated by ΔVar ($\beta = -0.40$, $p < 0.01$) and minVar ($\beta = -0.54$, $p < 0.01$), while the diffusion coefficient (σ_{AV}) was positively modulated by ΔMd ($\beta = 0.28$, $p < 0.01$) and minVar ($\beta = 0.52$, $p < 0.01$). The parameter λ was more effectively modulated than σ_{AV} (Table 4, $BIC_{\sigma AV} = 710.38$, $BIC_{\lambda} = 511.06$). A larger λ indicates more leakage and smaller RSE magnitude. Thus, the prediction of ΔVar for λ still contradicts with the principle of congruent effectiveness, a finding consistent with our experimental results. However, the predictor in the variability rule (minVar)—combining together two signals with larger variability in their RTs—may strongly reduce the leakage of integrated evidence, demonstrating that the principle can operate in a totally different model architecture.

6. Discussion

Overall, the present study demonstrates the generalization of established RSE principles to a dynamic multisensory onset-offset context. Greater overlap as well as larger variability of the unisensory RT distributions leads to larger RSE, as stated in the principle of congruent effectiveness and the variability rule (Otto et al., 2013). Crucially, our results further show that the principle of congruent effectiveness plays a dominant role compared to the variability rule in explaining the RSE rank, and the similarity in the central tendency of unisensory RT distributions dominates the other two predictors concerning the variability of unisensory RT distributions. An evaluation of two classes of computational mechanisms reveals that the context-variant race models outperform the coactivation models in explaining the RSE. Among the three predictors, the similarity in the central tendency of unisensory RTs most strongly modulates the assumed crossmodal interaction in the winning computational architectures (e.g., enhancing the evidence accumulation rates, or lowering the decision criterion in the context-variant race models).

6.1. *Possible reasons the principle of congruent effectiveness outweighs the variability rule*

Raab (1962) had concluded that the similarity between RT distributions of single signals determines the RT facilitation in redundant conditions. However, until Otto et al. (2013), the role that the variability in RT distributions of single signals played received less attention than their mean. In their study, the authors confirmed that variability in unisensory RTs is able to modulate the RSE via rigorous mathematical simulations and experimental data. This is also supported by evidence from both the group and the individual levels in the present study. But in general, as our results showed, the principle of congruent effectiveness (ΔMd + ΔVar) overwhelms the variability rule (minVar) in explaining the RSE. Moreover, neither the similarity in the variability of unisensory RTs (ΔVar) nor the variability itself (minVar) were as effective as the similarity in the central tendency of unisensory RTs (ΔMd). In addition, ΔVar positively predicted the RSE magnitude (Table 1), challenging the validity of the principle's claim.

However, manipulations of stimulus strength or type rarely change the variance of RT distributions alone without also shifting their central tendency. Empirically and theoretically, the mean and variance of the reaction-time distribution covary: sequential-sampling models predict systematic co-variation between the across-trial mean (or median) RT and the across-trial dispersion (variance) when core parameters (e.g., drift rate) are altered (Ratcliff, 1978; Wagenmakers & Brown, 2007), and this prediction is consistently supported by descriptive fits of empirical data (e.g., ex-Gaussian), which show that experimental manipulations typically affect both

Table 4
Fit parameters in the best models modulated by RSE predictors.

Models	Predictor (β)				BIC	adjusted R^2
		ΔMd	ΔVar	minVar		
Race Model						
MN + CP	σ_c	0.01	0.43**	−0.36**	786.32	0.19
	θ_c	−0.47**	0.12*	0.31**	745.79	0.33
MN + CA	σ_c	−0.02	0.42**	−0.29**	795.84	0.16
	μ_c	−0.48**	0.10	0.25**	754.50	0.30
Coactivation Model						
OU	σ_{AV}	0.28**	−0.05	0.52**	710.38	0.39
	λ	0.07	−0.40**	−0.54**	511.06	0.76

* $p < 0.05$, ** $p < 0.01$.

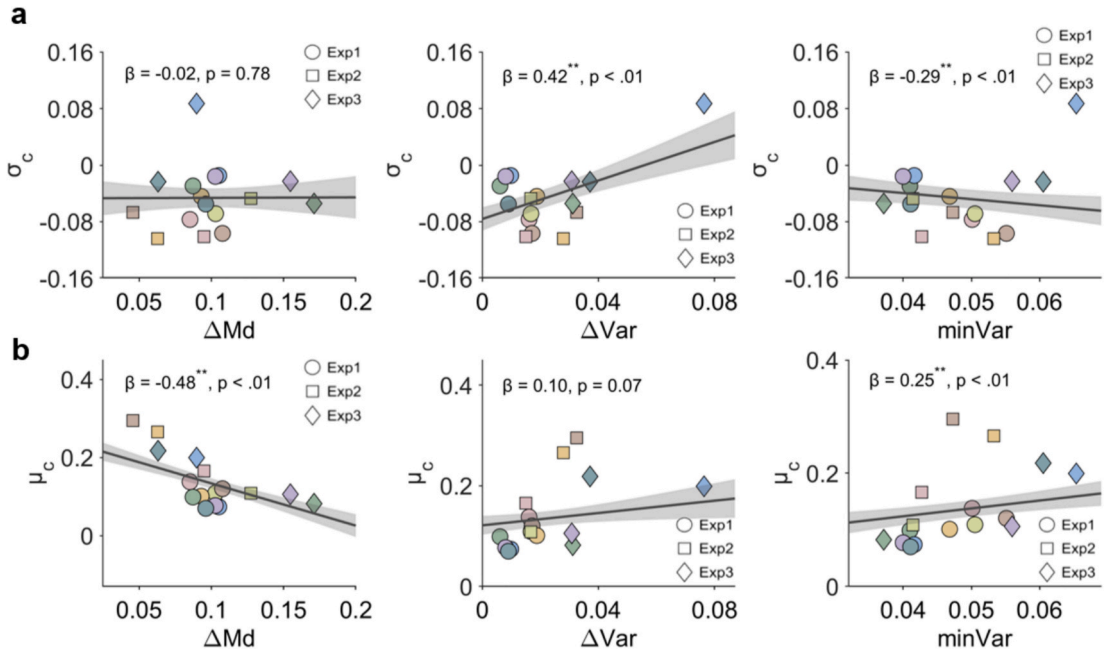

Fig. 8. MN + CA Model Predictions. (a) and (b) A graphic illustration of the prediction of ΔMd, ΔVar, minVar for parameters σ_c, μ_c in the MN + CA model, respectively. Circles represent corresponding data from Experiment 1, squares from Experiment 2, and rhombuses from Experiment 3. * $p < 0.05$, ** $p < 0.01$.

location and shape parameters of RT distributions (Heathcote & Popiel, 1991). This means we cannot independently manipulate the RT variability and experimentally isolate its role in modulating the RSE, as we can with the central tendency of RTs by manipulating the stimulus onset asynchrony (SOA) between two unisensory signals (e.g., Raab (1962) and Experiment 1 in Otto et al. (2013)). Additionally, in these experiments manipulating stimulus strength or type, it is more difficult to anticipate the changes in RT variability (ΔVar and minVar) compared to the changes in its central tendency (ΔMd). This explains why the present study observed significant RSE changes as a function of the RT variability (Fig. 4), whereas in Otto et al. (2013)'s, its effectiveness was only demonstrated by its ability to cancel out the expected RSE based on the principle of congruent effectiveness. By and large, it is the characteristics of the unisenosry RT distributions that ultimately determines the RSE. Which principle is dominant may flexibly depend on these RT distributions. Given that changes in the similarity of RT central tendency are more reliably expected in most contexts, the present study supports a simplification of the two principles into a single key predictor for a more parsimonious explanation.

6.2. *Other constraints on the principles of RSE*

Effects of multisensory integration are robust when crossmodal stimuli appear at the same time, at the same location and are of the same category (Andersen & Mamassian, 2008; Diaconescu et al., 2011; Frassinetti et al., 2002; Roberts et al., 2024). In our study, however, crossmodal stimuli with transient onsets and offsets—whether congruent or incongruent—elicited comparable RSEs. In RSE, signals are redundant as long as each of them is sufficient for a correct response. Therefore, when task is defined by change detection, the timing of the change matters more than its type. Consistently, previous studies also uncovered that stimulus congruency between physical changes did not remarkably influence multisensory decision-making (Andersen & Mamassian, 2008; Bernstein & Eason, 1970; Van der Burg et al., 2010), potentially indicating that onset and offset changes share a fundamental crossmodal communication mechanism, such as phase resetting (Brang et al., 2022; Mégevand et al., 2020; Mercier et al., 2013; Mercier et al., 2015; Thorne et al., 2011). But it cannot be simply inferred that other stimulus congruency, especially the semantic congruency, is unable to modulate the RSE (Diaconescu et al., 2011; Laurienti et al., 2004; Roberts et al., 2024), and reformulate the RSE principles, which awaits further examination in the future.

In addition to stimulus congruency, another famous principle in multisensory research is the inverse effectiveness principle, which states bimodal stimuli with low or mediate intensity elicit more significant integration effect than bimodal stimuli with higher intensity (Meredith & Stein, 1983; Stein & Stanford, 2008; von Saldern & Noppeney, 2013). In RSE, this principle is challenged by the fact that the rank of RSE is not determined by the absolute intensity of crossmodal stimuli but by their relative intensity (their similarity), summarized in the principle of congruent effectiveness. However, considering that unisensory RTs vary as a function of signal intensity, combination of too strong or too weak unisensory stimuli in the redundant condition still imposes influence on the RSE and its principles. For example, when the signal strength of unisensory stimuli continuously increases, the RTs gradually shorten until they reach a minimum which is constrained by perceptual and motor ability. Even unisensory RTs have identical median and variability at this time, the RSE would unavoidably decrease to zero. Conversely, when signal strength of unisensory stimuli continuously decrease,

the RTs gradually lengthen until they are not able to be processed in parallel but in serial when combined together. At this time, the RSE would turn zero too. Therefore, it is obvious that the RSE principles are probably constrained by stimulus intensity. In line with this, other factors such as task difficulty, workload capacity (Blunden et al., 2022; Townsend & Ashby, 1983; Townsend & Nozawa, 1995; Yang et al., 2018) may impose constraints on the RSE and its principles in the same manner.

6.3. Comparison between two classes of computational architectures underlying RSE

The two classes of computational models underlying the RSE rely on distinct implementations of how redundant signals are integrated to facilitate decision-making (Colonius & Diederich, 2020; Diederich, 1995; Diederich & Busemeyer, 2003; Otto & Mamassian, 2012; Plass & Brang, 2021). The present study evaluated the performance of eight representative models and discovered that the RSE data in the dynamic multisensory context is optimally fitted by two context-variant race models (MN + CA/CP models). But it should be noted that with noise (σ_c) as a free parameter, the MN + CP model that fixes the evidence accumulation rate (μ_c) while adjusts the decision criterion (θ_c, Eq. (10) is mathematically equivalent to the MN + CA model that fixes the decision criterion (θ_c) while adjusts the evidence accumulation rate (μ_c, Eq. (11).

The three race models that only free one candidate process (MN, CP and CA) had nearly equal performance, which is in contrast to Plass & Brang (2021) who found the CP model is the best. Since Plass & Brang (2021) tested the RSE in a context consisting of onset signals in auditory, visual, and tactile modalities while we tested the RSE in a context consisting of diverse types of signal changes in the audiory and visual modalities, currently the causes underlying the discrepancy in the model performances remain inconclusive. Nevertheless, the three models performed worse than the two more flexible models (MN + CA/CP models) even with penalties for model complexity. Therefore, for redundant signals in dynamic settings, crossmodal talk potentially increases the evidence accumulation rate or lowers the decision criterion, simultaneously reducing the noise in evidence accumulation (positive μ_c and θ_c, with a negative σ_c in Table 2).

The noise reduction process in the race models indicates the evidence accumulation process has smaller variability and higher reliability. Given the behavioral and neural evidence collectively support reduced rather than enhanced variability in multisensory integration (Kayser et al., 2010; Murray et al., 2019; Plass & Brang, 2021), the noise reduction in the two optimal race models seems more explainable and plausible. This result was also reported in Plass & Brang (2021)'s MN model, however, a noise increase was found in Otto et al. (2013)'s and even in our own MN model. So as to the correlation parameter (ρ) that characterizes the correlation between unisensory RT distributions, our fitting results demonstrated a positive relationship among all the race models, Due to history effect, the correlation is assumed negative. A positive one probably indicates that the cost of modal switching is overwhelmed by crossmodal talk that realigns the neural oscillatory phases to a similar excitability (Lakatos et al., 2007; Mégevand et al., 2020; Mercier et al., 2015; Thorne et al., 2011). The results are consistent with Plass & Brang (2021)'s CA and CP models but contradictory with most models in Otto et al. (2013)'s.

The coactivation models, albeit having provided excellent fit for the mean and variance of RTs in redundant conditions with different SOAs (Blurton et al., 2014; Chandrasekaran et al., 2019; Diederich, 1995; Schwarz, 1989; Schwarz, 1994), did not perform as well as the context-variant race models in fitting the RT distributions in our dynamic onset-offset context. The less competitiveness of the coactivation model is not because the diffusion coefficient was freed in the fitting procedure. If the diffusion coefficient in the redundant conditions was composed by those from the single conditions (e.g., $\sigma_{AV}^2 = \sigma_A^2 + \sigma_V^2 + 2\rho_{AV}\sigma_A\sigma_V$), the model performance would not be better. However, introducing an additional leaky process during evidence accumulation (the OU model) significantly improved the model performance compared to the standard diffusion model, which highlights the importance of model architecture rather than fitting procedure. The leaky process in the OU model is initially proposed to simulate a lost of evidence during accumulation. In the framework of crossmodal interaction, it can be explained as subadditive responses to bimodal stimuli, i.e., the responses to bimodal stimuli are smaller than summed responses to unimodal stimuli (Michail et al., 2022; von Saldern & Noppeney, 2013; Werner & Noppeney, 2011). It can be alternatively explained as a crossmodal inhibition: during evidence accumulation, the unimodal channels suppress each other's evidence accumulation rates so that the pooled evidence in the common channel accumulate more slowly. Otherwise, the RT in the redundant conditions is too short to be observed in the empirical data. Crossmodal inhibition has recently been included in artificial network models to interpret the behavioral facilitation in RSE (Cuppini et al., 2020; Vastano et al., 2022).

In brief, our model comparison results favor the context-variant race models relative to the coactivation models. Consistently, other studies employing similar modelling comparison procedure or other analysis methods (e.g., survivor interaction contrast analysis, and capacity analysis in systems factorial technology) using data from other representative OR-decision tasks (e.g., a Go/No-Go task) also support race models: for context-invariant ones, see (Blunden et al., 2022; Johansson & Ulrich, 2025); for context-variant ones, see (Mordkoff & Yantis, 1991; Yang et al., 2018). Note in these studies, the context-invariant/variant race model is equivalent to the independent/interactive parallel processing models with unlimited capacity and self-terminating stopping rules. Even with converging evidence of modelling, it should be very cautious in adjudicating between the two classes of models. For instance, a newly published study combining RTs and neural indices reported that a subadditive coactivation model outperformed a series of context-variant race models when greater parametric flexibility was allowed (e.g., the decision criteria in single and redundant conditions were free parameters during fitting)(Egan et al., 2025). And as previous neurophysiological and neuroimaging studies have revealed, the neural mechanisms underlying the two classes of models could be flexibly deployed to cope with different task demands (van Atteveldt et al., 2014). A promising direction in the future is to integrate these two classes of computational implementation into a unified architecture.

6.4. The computational models flexibly adapt to dynamic multisensory contexts

We found that the key predictor derived from the RSE principles most effectively modulates the evidence accumulation rate or decision criterion in the two optimal race model architectures, while its influence on the noise reduction is not significant. In other words, when unimodal stimuli are more similar in their central tendency of RTs, their combination increases evidence accumulation rates or decreases decision criteria, leading to a larger RSE. Other predictors, such as the lower boundary of RT variability, also significantly modulates the evidence accumulation rate or decision criterion in these two models. Additionally, the two predictors relating to RT variability influence the noise reduction process in these models. For instance, larger unisensory RT variability is associated with greater noise reduction (σ_c becomes more negative). However, the explanatory power of these predictors is much lower than that of the key predictor. These results demonstrate that the race models can adaptively tune their underlying processes—such as increasing the evidence accumulation rate, lowering the decision criterion, or reducing the noise in evidence accumulation—to optimize performance based on the characteristics of redundant signals.

By contrast, in Otto et al. (2013)'s study, the race model (the MN model) fitted in one redundant condition can perfectly predict the observed RSE in other redundant conditions, indicating that no additional adjustment of the noise was needed to adapt to conditions with varying unisensory RT distributions. However, the same model in our study required adjustment of the noise reduction process to optimize its performance, suggesting that models with flexible underlying processes are more adaptive to diverse contexts. More intriguingly, we notice that the RSE principles, although derived based on probability summation framework, could modulate the processes in the other model architectures have no relationship with the framework. For example, in the OU model, both the leaky and the diffusion process are modulated by predictors from the RSE principles. The association of the RSE principles with the computational architecture of coactivation models further demonstrates the generality and flexibility of these principles.

7. Limitations

First, having been investigated for more than a hundred years, the redundant signals effect was not merely an effect across modality, it has also been observed across features in one modality (Blunden et al., 2022; Fischer & Miller, 2008; Mordkoff & Yantis, 1991; Townsend & Nozawa, 1995). The current report and conclusion about the principles and computational models were based on evaluation in dynamic multisensory onset-offset contexts. Its ubiquity has not been examined across other contexts, including the semantic one. Second, for the sake of simplicity, the present study did not include the motor or non-decision time in the model fitting. Since a fair comparison between the two classes of models should assume a constant non-decision time, including it or not would not affect the outcome. But a more unified model that simultaneously consider all the processes should be considered in the future. Third, the present adopted a model comparison approach to assess the predictability of the two classes of models. However, there is another approach capable to distinguish the underlying model architecture of RSE, such as, the systems factorial technology (Blunden et al., 2022; Townsend & Ashby, 1983; Townsend & Nozawa, 1995; Yang et al., 2018). A more systematic investigation unifying the two approaches may provide more information about the model architecture.

8. Conclusion

The present study confirms the generalization of RSE principles to a dynamic multisensory onset-offset context, where the principle of congruent effectiveness—mainly driven by the similarity in the central tendency of unisensory performances—dominates in predicting RSE, outperforming the variability rule. Moreover, context-variant race models provide a better account of RSE than coactivation models, and the key predictor adaptively modifies the crossmodal interactive processes in the optimal race models. Collectively, our findings provide valuable insights into how we integrate redundant unisensory signals to make quick decisions and lay a solid foundation for future exploration of its neural mechanisms.

CRediT authorship contribution statement

Shiqi Tan: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yuhui Cheng:** Writing – review & editing, Validation, Resources, Investigation, Funding acquisition. **Jieru Chen:** Writing – review & editing, Validation, Software, Investigation. **Xiangyong Yuan:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Yi Jiang:** Writing – review & editing, Supervision, Software, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by grants from the STI2030-Major Projects (2021ZD0204200, 2021ZD0203800), the National Natural

Science Foundation of China (32430043, 31600884, 211061B32402), the Interdisciplinary Innovation Team (JCTD-2021-06), Youth Innovation Promotion Association of the Chinese Academy of Sciences, the Key Research and Development Program of Guangdong, China (2023B0303010004), and the Fundamental Research Funds for the Central Universities.

Data availability

All data and code that support the findings of this study are open access at https://www.scidb.cn/en/s/aeAr2i.

References

Alais, D., Newell, F. N., & Mamassian, P. (2010). *Multisensory Processing in Review: from Physiology to Behaviour*.
Andersen, T. S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision Research, 48*(25), 2537–2544. https://doi.org/10.1016/j.visres.2008.08.018
Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience, 7*(11), 1190–1192. https://www.nature.com/articles/nn1333.
Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41*(5), 809–823.
Bernstein, I. H., & Eason, T. R. (1970). Use of tone offset to facilitate reaction time to light onset. *Psychonomic Science, 20*(4), 209–210. https://doi.org/10.3758/bf03329022
Bizley, J. K., Jones, G. P., & Town, S. M. (2016). Where are multisensory signals combined for perceptual decision-making? *Current Opinion in Neurobiology, 40*, 31–37. https://doi.org/10.1016/j.conb.2016.06.003
Blunden, A. G., Hammond, D. A., Howe, P. D., & Little, D. R. (2022). Characterizing the time course of decision-making in change detection. *Psychological Review, 129*(1), 107.
Blurton, S. P., Greenlee, M. W., & Gondan, M. (2014). Multisensory processing of redundant information in go/no-go and choice responses. *Attention, Perception, & Psychophysics, 76*, 1212–1233.
Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.
Brang, D., Plass, J., Sherman, A., Stacey, W. C., Wasade, V. S., Grabowecky, M., Ahn, E., Towle, V. L., Tao, J. X., Wu, S., Issa, N. P., & Suzuki, S. (2022). Visual cortex responds to sound onset and offset during passive listening [Article]. *Journal of Neurophysiology, 127*(6), 1547–1563. https://doi.org/10.1152/jn.00164.2021
Chandrasekaran, C., Blurton, S. P., & Gondan, M. (2019). Audiovisual detection at different intensities and delays. *Journal of Mathematical Psychology, 91*, 159–175. https://doi.org/10.1016/j.jmp.2019.05.001
Chandrasekaran, C., Lemus, L., Trubanova, A., Gondan, M., & Ghazanfar, A. A. (2011). Monkeys and humans share a common computation for face/voice integration. *PLoS computational biology, 7*(9), Article e1002165. https://doi.org/10.1371/journal.pcbi.1002165&type=printable
Chua, S. F. A., Liu, Y., Harris, J. M., & Otto, T. U. (2022). No selective integration required: A race model explains responses to audiovisual motion-in-depth. *Cognition, 227*, Article 105204. https://doi.org/10.1016/j.cognition.2022.105204
Colonius, H., & Diederich, A. (2006). The race model inequality: Interpreting a geometric measure of the amount of violation. *Psychological Review, 113*(1), 148–154. https://doi.org/10.1037/0033-295x.113.1.148
Colonius, H., & Diederich, A. (2020). Formal models and quantitative measures of multisensory integration: A selective overview. *The European Journal of Neuroscience, 51*(5), 1161–1178. https://doi.org/10.1111/ejn.13813
Cuppini, C., Ursino, M., Magosso, E., Crosse, M. J., Foxe, J. J., & Molholm, S. (2020). Cross-sensory inhibition or unisensory facilitation: A potential neural architecture of modality switch effects. *Journal of Mathematical Psychology, 99*, Article 102438. https://doi.org/10.1016/j.jmp.2020.102438
Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology, 106*(6), 2896–2909. https://doi.org/10.1152/jn.00303.2011
Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology, 39*(2), 197–215.
Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology, 47*(3), 304–322. https://doi.org/10.1016/s0022-2496(03)00003-8
Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *eLife, 3*, Article e03005. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4102720/pdf/elife03005.pdf.
Egan, J. M., Gomez-Ramirez, M., Foxe, J. J., O'Connell, R. G., & Kelly, S. P. (2025). Distinct audio and visual accumulators co-activate motor preparation for multisensory detection. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-025-02280-9
Fischer, R., & Miller, J. (2008). Differential redundancy gain in onset detection versus offset detection. *Perception & Psychophysics, 70*(3), 431–436. https://doi.org/10.3758/pp.70.3.431
Franzen, L., Delis, I., De Sousa, G., Kayser, C., & Philiastides, M. G. (2020). Auditory information enhances post-sensory visual evidence during rapid multisensory decision-making. *Nature Communications, 11*(1), 5440. https://doi.org/10.1038/s41467-020-19306-7
Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research, 147*, 332–343. https://doi.org/10.1007/s00221-002-1262-y
Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience, 25*(20), 5004–5012. https://www.jneurosci.org/content/jneuro/25/20/5004.full.pdf.
Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences, 10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008
Gondan, M. (2010). A permutation test for the race model inequality [Article]. *Behavior Research Methods, 42*(1), 23–28. https://doi.org/10.3758/brm.42.1.23
Gondan, M., Lange, K., Rösler, F., & Röder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review, 11*, 307–313.
Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention Perception & Psychophysics, 78*(3), 723–735. https://doi.org/10.3758/s13414-015-1018-y
Gondan, M., Vorberg, D., & Greenlee, M. W. (2007). Modality shift effects mimic multisensory interactions: An event-related potential study. *Experimental Brain Research, 182*, 199–214.
Grice, G. R., Canham, L., & Gwynne, J. W. (1984). Absence of a redundant-signals effect in a reaction-time task with divided attention. *Perception & Psychophysics, 36*(6), 565–570. https://doi.org/10.3758/bf03207517
Harrar, V., Harris, L. R., & Spence, C. (2017). Multisensory integration is independent of perceived simultaneity. *Experimental Brain Research, 235*(3), 763–775. https://doi.org/10.1007/s00221-016-4822-2
Heathcote, A., & Popiel, S. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological bulletin, 109*(2), 340. https://doi.org/10.1037/0033-2909.109.2.340
Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology, 63*(3), 289.
Innes, B. R., & Otto, T. U. (2019). A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. *Scientific Reports, 9*, 2921. https://doi.org/10.1038/s41598-019-39924-6
Johansson, R. C., & Ulrich, R. (2025). Redundancy gains in absolute judgments of loudness and brightness. *Computational Brain & Behavior*, 1–12.

Kayser, C., Logothetis, N. K., & Panzeri, S. (2010). Visual enhancement of the information representation in auditory cortex. *Current Biology, 20*(1), 19–24.
Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex, 18*(7), 1560–1574.
Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279–292. https://doi.org/10.1016/j.neuron.2006.12.011
Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research, 158*(4), 405–414. https://doi.org/10.1007/s00221-004-1913-2
Lemus, L., Hernández, A., Luna, R., Zainos, A., & Romo, R. (2010). Do sensory cortices process more than one sensory modality during perceptual judgments? *Neuron, 67*(2), 335–348.
Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research, 1173*, 102–109.
Mégevand, P., Mercier, M. R., Groppe, D. M., Golumbic, E. Z., Mesgarani, N., Beauchamp, M. S., Schroeder, C. E., & Mehta, A. D. (2020). Crossmodal phase reset and evoked responses provide complementary mechanisms for the influence of visual speech in auditory cortex. *Journal of Neuroscience, 40*(44), 8530–8542. https://www.jneurosci.org/content/jneuro/40/44/8530.full.pdf.
Mercier, M. R., & Cappe, C. (2020). The interplay between multisensory integration and perceptual decision making. *NeuroImage, 222*, Article 116970. https://doi.org/10.1016/j.neuroimage.2020.116970
Mercier, M. R., Foxe, J. J., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., & Molholm, S. (2013). Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration. *NeuroImage, 79*, 19–29. https://doi.org/10.1016/j.neuroimage.2013.04.060
Mercier, M. R., Molholm, S., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., & Foxe, J. J. (2015). Neuro-oscillatory phase alignment drives speeded multisensory response times: An electro-corticographic investigation. *Journal of Neuroscience, 35*(22), 8546–8557. https://doi.org/10.1523/jneurosci.4527-14.2015
Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389–391. https://doi.org/10.1126/science.6867718
Michail, G., Senkowski, D., Holtkamp, M., Wachter, B., & Keil, J. (2022). Early beta oscillations in multisensory association areas underlie crossmodal performance enhancement. *NeuroImage, 257*, Article 119307. https://doi.org/10.1016/j.neuroimage.2022.119307
Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals [; Research Support, U.S. Gov't, Non-P.H.S.]. *Cognitive Psychology*, *14*(2), 247-279. https://doi.org/10.1016/0010-0285(82)90010-x.
Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics, 40*(5), 331–343.
Miller, J. (2016). Statistical facilitation and the redundant signals effect: What are race and coactivation models? *Attention, Perception, & Psychophysics, 78*, 516–519.
Miller, J., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model [Review]. *Cognitive Psychology, 46*(2), 101–151. https://doi.org/10.1016/s0010-0285(02)00517-0
Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17*(2), 520.
Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2005). Grabbing your ear: Rapid auditory–somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex, 15*(7), 963–974.
Murray, M. M., Thelen, A., Ionta, S., & Wallace, M. T. (2019). Contributions of intraindividual and interindividual differences to multisensory processes. *Journal of Cognitive Neuroscience, 31*(3), 360–376. https://direct.mit.edu/jocn/article-abstract/31/3/360/28962/Contributions-of-Intraindividual-and?redirectedFrom=fulltext.
Nadarajah, S., & Kotz, S. (2008). Exact distribution of the max/min of two Gaussian random variables. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 16*(2), 210–212.
Noorani, I., & Carpenter, R. H. (2016). The LATER model of reaction time and decision. *Neuroscience and Biobehavioral Reviews, 64*, 229–251. https://doi.org/10.1016/j.neubiorev.2016.02.018
Otto, T. U. (2019). *RSE-box: An analysis and modelling package to study response times to multiple signals*. The Quantitative Methods for Psychology.
Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *The Journal of Neuroscience, 33*(17), 7463–7474. https://doi.org/10.1523/JNEUROSCI.4678-12.2013
Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology, 22*(15), 1391–1396. https://doi.org/10.1016/j.cub.2012.05.031
Otto, T. U., & Mamassian, P. (2016). Multisensory decisions: The test of a race model, its logic, and power. *Multisensory Research, 30*(1), 1–24. https://doi.org/10.1163/22134808-00002541
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.
Plass, J., & Brang, D. (2021). Multisensory stimuli shift perceptual priors to facilitate rapid behavior. *Scientific Reports, 11*(1), 23052. https://doi.org/10.1038/s41598-021-02566-8
Raab, D. H. (1962). Statistical facilitation of simple reaction times [Article]. *Transactions of the New York Academy of Sciences, 24*(5), 574. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x
Raposo, D., Sheppard, J. P., Schrater, P. R., & Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *Journal of Neuroscience, 32*(11), 3726–3735. https://www.jneurosci.org/content/jneuro/32/11/3726.full.pdf.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59.
Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*.
Reddi, B. A. J., Asrress, K. N., & Carpenter, R. H. S. (2003). Accuracy, information, and response time in a saccadic decision task [Article]. *Journal of Neurophysiology, 90*(5), 3538–3546. https://doi.org/10.1152/jn.00689.2002
Regenbogen, C., Seubert, J., Johansson, E., Finkelmeyer, A., Andersson, P., & Lundström, J. N. (2018). The intraparietal sulcus governs multisensory integration of audiovisual information based on task difficulty. *Human Brain Mapping, 39*(3), 1313–1326.
Roberts, K., Jentzsch, I., & Otto, T. U. (2024). Semantic congruency modulates the speed-up of multisensory responses. *Scientific Reports, 14*(1). https://doi.org/10.1038/s41598-023-50674-4
Schirillo, J. A. (2011). Cross-modal detection using various temporal and spatial configurations. *Attention, Perception, & Psychophysics, 73*, 237–246.
Schwarz, W. (1989). A new model to explain the redundant-signals effect. *Perception & Psychophysics, 46*(5), 498–500. https://doi.org/10.3758/bf03210867
Schwarz, W. (1994). Diffusion, superposition, and the redundant-targets effect. *Journal of Mathematical Psychology, 38*(4), 504–520.
Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage, 56*(4), 2200–2208.
Shaw, L. H., Freedman, E. G., Crosse, M. J., Nicholas, E., Chen, A. M., Braiman, M. S., Molholm, S., & Foxe, J. J. (2020). Operating in a multisensory context: Assessing the interplay between multisensory reaction time facilitation and inter-sensory task-switching effects. *Neuroscience, 436*, 122–135. https://doi.org/10.1016/j.neuroscience.2020.04.013
Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*.
Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics, 63*(2), 330–336.
Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press.
Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*(5), 406. https://doi.org/10.1038/nrn2377
Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., Kurela, L. R., Siemann, J. K., James, T. W., & Wallace, M. T. (2014). Identifying and quantifying multisensory integration: A tutorial review. *Brain Topography, 27*(6), 707–730. https://doi.org/10.1007/s10548-014-0365-7
Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *Journal of Neuroscience, 31*(10), 3853–3861. https://doi.org/10.1523/jneurosci.6176-10.2011
Todd, J. W. (1912). *Reaction to multiple stimuli*. Science Press.
Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*(4), 321–359.

van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: Flexible use of general operations. *Neuron, 81*(6), 1240–1253. https://doi.org/10.1016/j.neuron.2014.02.044

Van der Burg, E., Cass, J., Olivers, C. N., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS One1, 5*(5), Article e10664. https://doi.org/10.1371/journal.pone.0010664

Vastano, R., Costantini, M., Alexander, W. H., & Widerstrom-Noga, E. (2022). Multisensory integration in humans with spinal cord injury. *Scientific Reports, 12*(1), 22156. https://doi.org/10.1038/s41598-022-26678-x

von Saldern, S., & Noppeney, U. (2013). Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *The Journal of Neuroscience, 33*(20), 8841–8849. https://doi.org/10.1523/JNEUROSCI.3020-12.2013

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review, 114*(3), 830.

Wagenmarkers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*.

Werner, S., & Noppeney, U. (2011). The contributions of transient and sustained response codes to audiovisual integration. *Cerebral Cortex, 21*(4), 920–931. https://doi.org/10.1093/cercor/bhq161

Yang, C.-T., Altieri, N., & Little, D. R. (2018). An examination of parallel versus coactive processing accounts of redundant-target audiovisual signal processing. *Journal of Mathematical Psychology, 82*, 138–158.