

The role of context in experiments and models of multisensory decision making

Yue Liu¹ & Thomas U. Otto²

School of Psychology and Neuroscience, University of St. Andrews

Affiliation:

School of Psychology & Neuroscience

St. Mary's Quad, South Street

St. Andrews

KY16 9JP

United Kingdom

Correspondence:

¹ yl37@st-andrews.ac.uk

² to7@st-andrews.ac.uk

1 **Abstract**

2 The availability of signals from multiple senses is often beneficial for perceptual decisions. To
3 study such benefits, models of multisensory decision-making are typically fed with the
4 behavioral performance as measured separately with unisensory component signals. Critically,
5 by doing so, the approach implicitly makes the so-called context invariance assumption, which
6 states that processing of a signal is independent of the experimental context in which it is
7 embedded. However, context invariance is not necessarily true and is difficult to test directly.
8 Here, we aim to assess context invariance indirectly in two testable scenarios. First, to consider
9 the role of stimulus context, we compared unisensory performance in trials that either included
10 a task-irrelevant signal in another modality, or not (unisensory vs. multisensory signal trials). We
11 found that performance was faster but less sensitive in trials that contained a task-irrelevant
12 signal. Hence, stimulus context invariance was violated. Second, to consider the role of
13 instruction context, we compared unisensory performance when participants were asked to
14 detect targets from either one or two modalities (unisensory vs. multisensory instructions). We
15 found that performance was deteriorated in multi- compared to unisensory instructions, which
16 was largely due to modality switch costs in multisensory instructions. Hence, instruction context
17 invariance did not hold either. As performance was variant in both scenarios, context invariance
18 cannot generally assumed to be true. We conclude that models of multisensory decision making
19 should critically consider potential violations of the context invariance assumption as a
20 potentially confounding factor.

21 **Keywords**

22 Context invariance; audio-visual reaction time; perceptual decision making; redundant signal
23 effect; race model; coactivation

24 1. Introduction

25 The availability of multisensory signals is often beneficial when making perceptual decisions
26 (Alais, Newell, & Mamassian, 2010; Chandrasekaran, 2017; Stein & Stanford, 2008; Ursino,
27 Cuppini, & Magosso, 2014). To study benefits in the speed of responses, research has extensively
28 used the redundant signals paradigm with its three basic conditions (Ashby & Townsend, 1986;
29 Diederich & Colonius, 2004a; Giray & Ulrich, 1993; Hershenson, 1962; Kinchla, 1974; Miller, 1982,
30 1986; Todd, 1912; see dark grey box in **Figure 1a**). In two single signal conditions, for example,
31 either an auditory or a visual target is presented (*A-* and *V*-trials¹). In the third redundant signals
32 condition, the two targets are presented together (*AV*-trials). To probe multisensory decisions,
33 the instruction is to respond in all three conditions with the same motor act, and to withhold a
34 response when no target signal is presented (which we refer to as the *AUV*-instruction). The two
35 targets are “redundant” in *AV*-trials because detection of either target is sufficient for a correct
36 response. In fact, an analysis of the task demands shows that the *AUV*-instruction couples two
37 detection tasks by a logic disjunction (Otto & Mamassian, 2017). The long-standing and highly
38 replicated finding is the *redundant signals effect* (*RSE*): responses to redundant signals are faster
39 than responses to unisensory signals (for recent examples, see Crosse, Foxe, & Molholm, 2019;
40 Innes & Otto, 2019; Lunn, Sjoblom, Ward, Soto-Faraco, & Forster, 2019; Murray et al., 2018; Van
41 der Stoep, Van der Stigchel, Van Engelen, Biesbroek, & Nijboer, 2019).

42 --- *Figure 1 about here* ---

43 Two competing model architectures were considered most frequently to account for
44 the *RSE*: race models and pooling models. As championed by Raab (1962), race models assume

¹ Non-standard Abbreviations: **A-/V-trials** stand for trials presenting only auditory or only visual signals. **AV-trials** present both auditory and visual signals. **A-/V- instructions** ask participants to detect either only auditory or only visual signals. The **AUV-instruction** asks to detect any auditory or visual signal. **RSE** is the redundant signals effect.

45 that the two targets are processed by parallel decision units and that a response on a given trial
46 is triggered by the faster of the two units (the term “race model” reflects the analogy to a horse
47 race). If corresponding decision times follow random distributions that overlap, race models
48 predict faster and less variable reaction times (RTs) due to statistical facilitation. The mechanism
49 is that a slower decision in one unit can be compensated for by a faster in the other. Alternatively,
50 pooling models assume that sensory evidence for both targets is combined to satisfy the
51 threshold of a single decision unit (Miller, 1982). As two targets provide more sensory evidence
52 than one, pooling models predict a speed-up of RTs by summing the evidence. The mechanism
53 is somehow similar to the speed-up of RTs as expected when, for example, the volume of a
54 sound or the contrast of a visual stimulus is increased. At this point, it is critical to highlight that
55 we use the term pooling model to avoid ambiguity (Otto & Mamassian, 2017). To describe
56 models with a single decision unit, the alternative term coactivation model has been used rather
57 consistently (e.g., Blurton, Greenlee, & Gondan, 2014; Colonius & Diederich, 2017; Diederich,
58 1994; Mordkoff & Yantis, 1991; Schwarz & Ischebeck, 1994; Townsend & Wenger, 2004; Ulrich
59 & Giray, 1986; Yang, Altieri, & Little, 2018; Zehetleitner, Krummenacher, & Müller, 2009; for a
60 recent review, see Colonius & Diederich, 2018). However, Miller (2016), who coined the term in
61 1982, clarified that the term should not be used to describe basic model architecture.

62 For multisensory and cognitive research, it is a central theme to understand how sensory
63 evidence is combined and, therefore, to discriminate between the competing model
64 architectures. A frequently applied approach checks if race models can be ruled out, which uses
65 the convenient feature that race models directly predict the RSE on the level of RT distributions.
66 Let $P(T_{AUV} \leq t | AV)$ be the cumulative probability that a response under the AUV-instruction
67 given an AV-trial has been triggered at time t . In a milestone contribution, Miller (1982) derived
68 an upper bound for the RSE to be in accordance with race models:

$$P(T_{AUV} \leq t | AV) \leq P(T_{AUV} \leq t | A) + P(T_{AUV} \leq t | V), \quad \text{Equation 1}$$

69 where $P(T_{AUV} \leq t | A)$ and $P(T_{AUV} \leq t | V)$ are the cumulative probabilities of observing
70 responses in the two single signal conditions. Inequality 1 is also known as Miller's bound, which
71 corresponds to the race model prediction assuming a maximal negative correlation between
72 decision times to auditory and visual signals (Colonius, 1990; a negative correlation implies that
73 observing a fast response to one signal would make it more likely to observe a slow response to
74 the other). The handy feature of Miller's bound is that it relates RTs in redundant and single
75 signal conditions (the left-hand term is given by the empirical CDF in the redundant signals
76 condition, the right-hand term is given by the sum of the empirical CDFs in the single signal
77 conditions; **Figure 1b**, top panel). If Miller's bound is violated at any time t , the mainstream view
78 is that race models cannot explain the RSE and that the alternative class of pooling models needs
79 to be involved instead (for recent reviews of the approach, see Gondan & Minakata, 2016; Otto
80 & Mamassian, 2017). As violations of Miller's bound replicate well with multisensory signals,
81 pooling models have dominated the research area in the last decades.

82 A critical issue with Miller's bound is that the test makes the so-called context invariance
83 assumption (e.g., Ashby & Townsend, 1986; Colonius, 1990; Luce, 1986; Miller, 2016; Otto &
84 Mamassian, 2017). To make exact predictions, Miller's bound should be computed based on the
85 processing times of the two parallel decision units on presentation of redundant signals (**Figure**
86 **1b**, bottom panel). Unfortunately, it is difficult to extract these times empirically as it is difficult
87 to know if a response on presentation of redundant signals was triggered by one or the other
88 decision unit (Luce, 1986). As illustrated in **Figure 1c**, the applied alternative is to assume that
89 the processing times of the parallel decision units remain unchanged in distribution independent
90 of the experimental context, that is, whether a single or a redundant signal is presented:

$$P(T_A \leq t | AV) = P(T_A \leq t | A),$$

Equation 2

$$P(T_V \leq t | AV) = P(T_V \leq t | V).$$

91 By assuming context invariance, empirical RTs as measured in the single signal conditions can be
92 used to compute Miller's bound as stated in Inequality 1. At this point, it is therefore crucial to
93 realize that any study that has used Miller's bound, knowingly or not, has tested race models
94 under the assumption of context invariance.

95 Is it important to consider context invariance? Violations of Miller's bound are typically
96 used as evidence to reject race models as an explanation of the RSE, which seems like a valid
97 deductive argument: If race models explain the RSE, then Miller's bound provides an upper limit.
98 Empirical evidence demonstrates that Miller's bound is violated (denying the consequent).
99 Therefore, race models cannot explain the RSE. The problem with the argument is that the first
100 premise is incomplete. The correct conditional statement reads: If race models explain the RSE
101 and if context invariance holds, then Miller's bound provides an upper limit. As the evidence
102 stays the same, it can only be concluded that the race model and the context invariance are not
103 both true. It is therefore crucial to realize that violations of Miller's bound cannot be used to
104 rule out all race models, that is, a cognitive architecture that assumes two parallel decision units
105 (e.g., Gondan & Minakata, 2016; Otto & Mamassian, 2010, 2017; Yang et al., 2018). A valid
106 alternative is that the context invariance assumption may be wrong.

107 Given that Miller's bound is frequently used in research on multisensory RTs, an
108 important question is whether or not the research field is aware of the context invariance
109 assumption. As mentioned above, the assumption is considered in early theoretical work (e.g.,
110 Ashby & Townsend, 1986; Colonius, 1990; Luce, 1986), but it is only implicit in Miller (1982) and
111 it is not mentioned in a frequently cited methods paper promoting the test (Ulrich, Miller, &
112 Schröter, 2007). Given that Miller (1982) used the terms separate activation and race model

113 synonymously, the add-on “separate” may imply that context invariance is an inherent feature
114 of this model class, but this link is rarely made explicit (Miller, 2016). To evaluate the current
115 understanding in the field, we systematically checked recent empirical studies that used the test.
116 For this review, we searched the Web of Science (Clarivate Analytics) for articles published
117 between 2016 and 2018 that had cited either Miller (1982) or Raab (1962). Using this method,
118 we identified 84 studies that tested the RSE empirically. Among these, 50 studies used Miller’s
119 bound², indicating that the test is frequently used. The astonishing finding is that only 16% (8 of
120 50 articles) explicitly mentioned the context invariance assumption. Hence, a majority of recent
121 RSE studies rejects race models, to put it mildly, using an incomplete argument.

122 Given the lack of consideration, it would be vital for the field to know if context
123 invariance can be reasonably assumed, in which case the rejection of race models could still
124 stand. Unfortunately, as pointed out by Luce (1986), testing the context invariance assumption
125 directly is difficult. The reason is that on presentation of redundant signals, it is difficult to know
126 whether a response was triggered by one or the other decision unit. Hence, it is difficult to access
127 the left-hand term of **Equation 2**. As an alternative approach, we propose that it is possible to
128 check context invariance more generally, and critically, in empirically testable scenarios. If
129 context invariance holds in such scenarios, one might argue that it is reasonably safe to assume
130 context invariance to be also true when testing race models with Miller’s bound. On the other
131 hand, if context invariance does not hold more generally, such findings would show that the
132 assumption, as used in the race model test, must be considered with more care.

133 To derive conditions that allow testing for context invariance broadly, we extended the
134 original redundant signals paradigm by using the same signal compositions but with three types
135 of task instructions (**Figure 1a**). As in the original paradigm, we included the typical AUV-

² The count includes nine studies that claimed to have used Miller’s bound, but who have in fact used what is known as the independent race model (for a similar critique, see Gondan & Minakata, 2016)

136 instruction and asked participants to respond on presentation of either an auditory or a visual
137 signal. Hence, responses were required in three trial types (A-, V-, and AV-trials). No response
138 was required on catch trials. In an additional A-instruction, we asked participants to respond
139 only on presentation of an auditory signal. Hence, responses were required in unisensory
140 auditory and in multisensory trials (A- and AV-trials). No response was required on unisensory
141 visual trials (V-trials) and on catch trials. Likewise, in an additional V-instruction, participants
142 were instructed to respond only on presentation of a visual signal (V- and AV-trials, but not on
143 A-trials and catch trials). Using this design, two types of context influence can be tested, which
144 we coin *stimulus context* and *instruction context*.

145 As the first scenario, the role of stimulus context is tested by analysing responses under
146 unisensory instructions (**Figure 1d**). For example, under the A-instruction, we measured
147 performance in both uni- and multisensory trials (A- and AV-trials). The difference between
148 these conditions is that multisensory trials contain a task-irrelevant stimulus (here the visual
149 stimulus). Analogously, the role of stimulus context can be tested under the V-instruction. If
150 performance is generally invariant to the stimulus context, performance in uni- and multisensory
151 trials is expected to remain unchanged.

152 As the second scenario, the role of instruction context is tested by analysing responses
153 on presentation of unisensory targets (**Figure 1e**). For example, on A-trials, we measured
154 performance under both uni- and multisensory instructions (A- and AUV-instructions). The
155 difference between these conditions is that under multisensory instructions, participants are
156 asked to respond also to the second unisensory signal (here the visual stimulus). Critically, on
157 analysed trials, the second signal was not present, hence the physical stimulation in both
158 conditions is the same. Analogously, the role of instruction context can be tested on
159 presentation of V-trials. If performance is generally invariant to the instruction context,
160 performance under uni- and multisensory instructions is expected to remain unchanged.

161 With these two scenarios, we aim to evaluate the plausibility of the context invariance
162 assumption (**Equation 2**). To explore the role of stimulus and instruction context broadly, we
163 tested performance in a simple detection task (as introduced so far) and analogously in a go/no-
164 go tasks. A difference between the two tasks lies in the number of transient signal onsets. In the
165 detection task, A- and V- trials present transient signal onsets in only one modality. In contrast,
166 AV-trials have transient signal onsets in both modalities. Hence, the number of transients is
167 changing. In the go/no-go task, all trials have two transient signal onsets, one in each modality.
168 Hence, the number of transients is here constant. As transient signal onsets are considered to
169 contribute to violations of Miller's bound (Innes & Otto, 2019), we speculate that changes in
170 context might influence these two tasks differently. Consequently, we investigated each
171 scenario using a task (detection vs. go/no-go) and sensory modality (audition vs. vision) factorial
172 design. If performance is constant across changing contexts, our results would provide at least
173 some assurance that assuming context invariance is unproblematic when testing race models
174 with Miller's bound. On the contrary, if performance is not constant, our results would provide
175 a warning message that the context invariance assumption must be considered with more care.

176 **2. Methods**

177 **2.1. Participants**

178 44 participants took part in the study (10 males, 34 females; aged between 18-34 years). All had
179 self-reported normal or corrected-to-normal vision and normal hearing. Participants gave
180 informed consent before the experiment. The study was approved by the *University Teaching*
181 *and Research Ethics Committee* (UTREC; approval code: PS12613).

182 **2.2. Apparatus**

183 The study was carried out in a dimly lit and sound-attenuated room. We controlled the
184 experiment using MATLAB (MathWorks) and the Psychophysics Toolbox extensions (Brainard,
185 1997; Kleiner et al., 2007; Pelli, 1997). Auditory stimuli were delivered to both ears using
186 Sennheiser HD-280 Pro headphones with a sampling frequency of 44.1 kHz. We calibrated the
187 auditory volume using a Bruel and Kjaer sound level meter (Type 2250 light) equipped with an
188 artificial ear adaptor (Type 4153). Visual stimuli were presented using a Dell UltraSharp U2713
189 monitor (resolution: 1920*1080 pixel; refresh rate: 60Hz). The viewing distance was 57 cm as
190 supported by a headstand. We collected responses using a hand-held pushbutton (custom-made)
191 connected to an RTBox V6 (Li, Liang, Kleiner, & Lu, 2010).

192 **2.3. Task and instructions**

193 We employed both a detection and a go/no-go task. In the detection task, a target signal was
194 either present or not. Participants were instructed to respond when they detected a target signal.
195 In the go/no-go task, a signal was presented on each trial. If a target signal was presented,
196 participants were instructed to respond. Nontarget signals did not require a response.

197 For each task, we defined targets using three instructions (**Figure 1a**). In AUV-instruction
198 blocks, participants responded to any auditory or visual target signal. In A-instruction blocks,
199 participants responded only to auditory target signals. In V-instruction blocks, participants were
200 asked to respond only to visual target signals.

201 Within instruction blocks, we presented four different combinations of signals randomly
202 interleaved (**Figure 1a**). In the detection task, these were an auditory signal only (A-trials), a
203 visual signal only (V-trials), both an auditory and a visual signal simultaneously (AV-trials), and
204 no signal at all (catch trials). In the go/no-go task, these were signals with an auditory target and
205 a visual nontarget (A-trials), signals with a visual target and an auditory nontarget (V-trials),

206 signals with both an auditory and a visual targets (AV-trials), and signals with both an auditory
207 and a visual nontargets (catch trials).

208 **2.4. Stimulus design**

209 The RSE has been observed with basically with any combination of salient multisensory signals.
210 For consistency, we here use a stimulus design similar to previous studies from our group (Innes
211 & Otto, 2019; Otto, Dassy, & Mamassian, 2013; Otto & Mamassian, 2012).

212 *Auditory simulation.* All auditory stimuli were generated anew on each trial using
213 Gaussian noise (i.e., a sequence of normally distributed random numbers at a sample rate of
214 44.1 kHz). Auditory detection and go-signals were noise tones filtered using a 2nd order
215 Butterworth filter with cut-off frequencies of 1000 Hz and 1500 Hz. Auditory nontarget signals
216 were noise tones filtered using a 2nd order Butterworth filter with cut-off frequencies of 250 Hz
217 to 750 Hz. All three types of signals were presented at 55.5 dB SPL. In addition, they were
218 embedded in auditory background noise, which was bandpass filtered using a 2nd order
219 Butterworth filter with cut-off frequencies of 200 Hz and 1550 Hz. We presented the background
220 noise at 55 dB SPL.

221 *Visual simulation.* Visual background stimulation consisted of 1500 dots presented on a
222 uniform grey screen. Each dot had the size of a pixel and was randomly either black or white.
223 Dots were uniformly distributed within the area of a ring with an inner/outer diameter of 1°/4°
224 around central fixation. Dots moved linearly in random directions with random speed, normally
225 distributed with mean 1 °/s and SD 0.2 °/s. Dots that moved outside the ring a were randomly
226 relocated within the ring area. Likewise, each dot had a lifetime of 0.1s after which they were
227 randomly relocated within the ring area. For visual detection and visual go signals, 50% of the
228 dots changed from liner motion to rotate clockwise around fixation. The rotation speed of each
229 dot was randomly selected from a normal distribution with mean 0.106 rad/s and SD 0.0106
230 rad/s. For visual nontarget signal, 50% of dots changed to counterclockwise rotation.

254 Each participant performed both the detection and the go/no-go task. The order of tasks
255 was counterbalanced across participants. For each task, four A-, four V- and six AUV-instruction
256 blocks were presented. This procedure ensured that for each target condition and instruction
257 block, 60 valid responses were collected. For example, each A-instruction block contains 15 A-
258 trials and 15 AV-trials (**Table 1**). Presenting four A-instruction blocks thus yields 60 A-trials and
259 60 AV-trials per participant. Instruction blocks were presented in a pseudo-random order.
260 Including breaks, the experiment took around 2.5h to complete. In total, we presented 76,278
261 trials and collected 36,957 correct responses on signal and go trials.

262 **2.6. Analysis**

263 We used miss rates and false alarm rates to compare the accuracy of responses. We excluded
264 data of one participant due to an overall miss rate of 17.7% from the analysis. The remaining 43
265 participants had overall miss rates of 2.3% ($\pm 0.3\%$; SEM) and overall false alarm rates of 1.3%
266 ($\pm 0.2\%$). Based on miss and false alarm rates, we computed sensitivity (d'). To deal with the
267 special cases where d' becomes infinity (when hit and/or false alarm rates were equal to 1 or 0),
268 we used the loglinear approach adjust hit and/or false alarm rates (Hautus, 1995).

269 We used RTs to compare the speed of responses. We performed an outlier correction
270 on RTs to reduce the influence of anticipatory responses and of slow responses due to fatigue
271 and/or lack of attention. The correction was based on the 60 valid responses per condition and
272 participant. We transformed RTs to the reciprocal scale ($1/RT$). We considered values falling
273 outside the range of $\pm 3 * 1.4826 * MAD$ around the median as outliers (Leys, Ley, Klein, Bernard,
274 & Licata, 2013; the procedure is implemented by the RSE-box function outCorrect, Otto, 2019).
275 On average, 0.86% of the valid responses were considered as outliers and removed from the
276 analysis.

277 To analyse the influence of stimulus context (**Figure 1d**), we focused on unisensory
278 instructions. For example, in an A-instruction block, responses are required in both unisensory

279 A-trials and multisensory AV-trials. To quantify the effect of context (here, the additional
280 presentation of a visual signal that did not require a response), we computed performance
281 differences in unisensory compared to AV-trials. For example for RTs, we computed the change
282 of RT as

$$\Delta RT = RT_{uni} - RT_{AV}. \quad \text{Equation 3}$$

283 With respect to context invariance, which is typically expressed on the level of RT distributions
284 (**Equation 2**), we computed here the difference in mean RTs as such difference proves a
285 difference in RT distributions. Analogously to RTs, we computed performance differences for
286 misses and false alarms ($\Delta Miss$, ΔFA). For sensitivity, we deducted unisensory from multisensory
287 performance ($\Delta d'$; when computing d' , we used catch trials and trials with only the irrelevant
288 signal as an estimate of false alarm rates). With these definitions, positive Δ -values always
289 indicate better performance in AV- compared to unisensory trials. Absolute Δ -values indicate
290 the size of performance changes.

291 To analyse the influence of instruction context (**Figure 1e**), we focused on unisensory
292 trials. For example, in A-trials, we tested if performance is the same or changes in A- compared
293 to AUV-instructions. To quantify any change due to the change in instruction, we computed
294 performance differences in unisensory compared to AUV-instructions analogously to **Equation**
295 **3** for RTs and miss rates (ΔRT , $\Delta Miss$; we here cannot compute d' as the design lacks separate
296 estimates of false alarm rates). With these measures, positive values always indicate better
297 performance in AUV compared to unisensory instructions.

298 In addition, we analysed the influence of trial history on performance. We labelled a trial
299 as “repeat” if the same trial type was presented on the previous trial. For example, a V-trial that
300 followed a V-trial is a repeat. We labelled a trial as “switch” if the other sensory signal was
301 presented on the previous trial. For example, a V-trial that followed an A-trial is a switch. To

302 quantify the trial history effect, we computed the performance difference in switch compared
303 to repeat trials:

$$History(RT) = RT_{switch} - RT_{repeat}. \quad \text{Equation 4}$$

304 Analogously, we computed the history effect on misses. With these definitions, a positive history
305 effect indicates better performance in repeat compared to switch trials. Analogously to
306 **Equation 3**, we computed changes in history effects in unisensory compared to AUV-instructions
307 (Δ History). With this definition, a positive value indicates smaller history costs in AUV- than in
308 unisensory instructions.

309 Finally, we analysed violations of Miller's bound (Miller, 1982). We followed the
310 approach described with the RSE-box (Otto, 2019). As a first step, we obtained cumulative
311 distribution functions (CDFs) in each condition and for each participant. For this, we ordered RTs
312 from the fastest to the slowest and computed corresponding cumulative probabilities (using the
313 RSE-box function *getCP*). We down-sampled these distributions to 50 quantiles using linear
314 interpolation (using the RSE-box function *interpCDF*). For example, with redundant signals, we
315 obtained a vector of quantile RTs named AV_q (where q is the index of the 50 quantiles). To obtain
316 Miller's bound, we summed up corresponding unisensory CDFs (using the RSE-box function
317 *getMiller*). With this procedure, Miller's bound is obtained at the same 50 quantiles as the
318 empirical CDFs ($Miller_q$). Violations of Miller's bound are quantified using a geometrical
319 approach (Colonius & Diederich, 2006). For this, violations of Miller's bound are averaged at
320 each quantile

$$Violation = \frac{1}{50} \sum_{q=1}^{50} \max(Miller_q - AV_q, 0). \quad \text{Equation 5}$$

321 This final step is obtained using the RSE-box function *getViolation*. While the basic race model
322 inequality test just requires only a categorical decision (Miller's bound violated vs. not violated),

323 we think the geometrical approach (Colonius & Diederich, 2006) is elegant and a quantification
324 of potential violations may be helpful to understand the underlying cause using a comparative
325 approach (Innes & Otto, 2019).

326 The extended paradigm allows computing Miller's bounds with three different pairs of
327 unisensory CDFs (**Figure 5a**). First, we used *A*- and *V*-trials, both tested in *AUV*-instructions
328 (**Figure 5**, dark red). These conditions share the same instruction with the predicted redundant
329 signals condition but differ in the presented stimulus. We refer to this computation as the
330 stimulus-different pair. This option is typically tested in RSE studies. Second, we used *A*-trials in
331 *A*-instructions and *V*-trials in *V*-instructions (**Figure 5**, light red). These conditions differ to the
332 predicted redundant signals condition in both stimulus and instruction, which we refer to as the
333 both-different pair. Third, we used *AV*-trials tested in *A*- and *V*-instructions (**Figure 5**, orange).
334 These conditions share the same stimulus context as the predicted redundant signals condition
335 but differ in instruction. Consequently, we refer to this computation as the instruction-different
336 pair. We computed Miller's bound with each of the three pairs. Hence, this allowed us to check
337 RTs on *AV*-trials tested under the *AUV*-instruction for potential violations of the race model
338 inequality in three different ways.

339 For statistical analysis, we performed all tests in IBM SPSS Statistics 22. When analysing
340 the effects of stimulus context, we used a repeated-measures analysis of variance (ANOVA) with
341 factors *task* (detection, go/no-go) and *instruction* (*A*, *V*). When analysing the effects of
342 instruction context, we used an ANOVA with factors *task* (detection, go/no-go) and *trial* (*A*, *V*).
343 For both, we specifically looked at the intercept term in the ANOVA outcome as a significant
344 intercept can be directly interpreted as a violation of stimulus and instruction context variance,
345 respectively. We further performed a correlation analysis across conditions to investigate the
346 relationship between performance changes due to instruction context and history costs (for
347 both speed and accuracy). When analysing violations of Miller's bound, we used an ANOVA with

348 factors *task* (detection, go/no-go) and *unisensory RT pair* (stimulus-different, both-different,
349 instruction-different). When performing ANOVAs, we applied Greenhouse-Geisser corrections
350 whenever sphericity was violated. The alpha level for all statistical tests was 5%.

351 **3. Results**

352 **3.1. Stimulus context**

353 To test the influence of stimulus context, we focused on responses in unisensory instruction
354 blocks (**Figure 1d**). We used RTs as a measure of speed and sensitivity as a measure of
355 performance accuracy. To measure the influence of task-irrelevant stimuli on performance (e.g.,
356 a visual stimulus in an A-instruction block), we computed the performance difference between
357 unisensory and multisensory trials (e.g., A- and AV-trials; **Equation 3**; for an illustration, see
358 **Figure 2a**). If stimulus context invariance holds, performance in the two conditions should not
359 differ.

360 --- *Figure 2 about here* ---

361 Regarding speed, first, we used a 2x2 ANOVA with factors *task* (detection, go/no-go) and
362 instruction (A-, V-instruction) to analyse changes in RT due to the changing stimulus context (ΔRT ;
363 **Figure 2b**). We found main effects of *task* ($F_{1,42}= 4.140$, $p=0.048$, $\eta_p^2=0.090$) and instruction
364 ($F_{1,42}=4.127$, $p=0.049$, $\eta_p^2=0.089$), as well as an interaction of *task* and instruction ($F_{1,42}=14.742$,
365 $p<0.001$, $\eta_p^2=0.260$). Post-hoc analyses revealed that these effects came only from the V-
366 instruction in the go/no-go task, which had smaller ΔRT values compared to the other three
367 conditions (all $p\leq 0.002$). ΔRT values in the other three conditions were not significantly different
368 from each other (all $p>0.05$). Most critically, we found that ΔRT values were different from zero
369 as indicated by a significant intercept ($F_{1,42}=93.346$, $p<0.001$, $\eta_p^2=0.690$), with an average ΔRT
370 value of 21.3 ms (± 2.2 ms; values following the \pm sign always indicate the SEM). In addition, post-

371 hoc analyses revealed that ΔRT values in all conditions were larger than zero (all $p \leq 0.015$). As
372 these findings show faster responses in AV-trials compared to corresponding unisensory trials,
373 stimulus context invariance does not hold in terms of RTs.

374 To understand the asymmetric effect in the go/no-go task (with a smaller ΔRT in the V-
375 instruction), we noted that there is a corresponding effect in RTs themselves. We found that RTs
376 are faster in V- compared to A-trials ($t_{42} = 2.009$, $p = 0.025$). In comparison, ΔRT was similar across
377 audition and vision in the detection tasks. Correspondingly, RTs are not different here either
378 ($t_{42} = -0.324$, $p = 0.748$). We will come back to this observation in the discussion.

379 Regarding accuracy, second, we used a 2x2 ANOVA to analyse changes in sensitivity due
380 to the changing stimulus context ($\Delta d'$; **Figure 2c**). We found neither main effects of task,
381 instruction, nor a task x instruction interaction (all $F_{1,42} \leq 1.543$, $p \geq 0.221$, $\eta_p^2 \leq 0.035$). Most
382 critically, we found a significant intercept ($F_{1,42} = 49.478$, $p < 0.001$, $\eta_p^2 = 0.541$), with an average $\Delta d'$
383 of $-0.268 (\pm 0.038)$. In addition, post-hoc analyses revealed that $\Delta d'$ values in all conditions were
384 different from zero (all $p \leq 0.007$). As these results show reduced sensitivity in AV-trials than in
385 corresponding unisensory trials, stimulus context invariance is also violated in terms of accuracy.

386 To understand these violations in accuracy in more detail, next, we investigated the
387 components to compute sensitivity, miss and false alarm rates, separately. Firstly, we used a 2x2
388 ANOVA to analyse changes in miss rates ($\Delta Miss$). We only found a main effect of instruction
389 ($F_{1,42} = 6.543$, $p = 0.014$, $\eta_p^2 = 0.135$). $\Delta Miss$ values were higher for A- ($1.1 \pm 0.3\%$) than V-
390 instructions ($0.2 \pm 0.2\%$). We found a significant intercept ($F_{1,42} = 11.220$, $p = 0.002$, $\eta_p^2 = 0.211$),
391 with an average $\Delta Miss$ value of $0.7\% (\pm 0.2\%)$. However, post-hoc analyses revealed that only
392 the $\Delta Miss$ value in the A-instruction of the detection task was significantly different from zero
393 ($t_{42} = 3.294$, $p = 0.002$). In the other three conditions, $\Delta Miss$ values were not different from zero
394 (all $t_{42} \leq 1.371$, $p \geq 0.178$). This suggests that stimulus context invariance in terms of miss rates was
395 only violated in the A-instruction of the detection task. Secondly, we used a 2x2 ANOVA to

420 (Δ IES). We found a main effect for Modality ($F_{1,42}=6.436$, $p=0.015$, $\eta_p^2=0.133$). Δ IES values were
421 higher in A (-36.3 ± 5.0 ms) than in V-instruction blocks (-21.8 ± 2.7 ms). We also observed a
422 significant interaction between task and modality ($F_{1,42}=4.383$, $p=0.042$, $\eta_p^2=0.094$). While there
423 is no significant difference between Δ IES values in audition (35.6 ± 6.6 ms) and vision (29.4 ± 3.5
424 ms) in the detection task ($t_{42}= -0.862$, $p>0.05$), audition Δ IES values (37.1 ± 6.6 ms) is significantly
425 larger than vision (14.2 ± 2.8 ms) in the go/no-go task ($t_{42}= -3.354$, $p=0.001$). Most critically, we
426 found that Δ IES values were different from zero by a significant intercept ($F_{1,42}=107.652$, $p<0.001$,
427 $\eta_p^2=0.719$), which suggests there is an overall increase in IES when changing the from uni- to
428 multisensory stimulus context, with an average Δ IES value of 29.1 ms (± 2.8 ms).

429 In summary, our results revealed that stimulus context invariance does not hold. We
430 found gains in response speed that trade-off with losses in sensitivity when task-irrelevant
431 stimuli were presented. Performance difference in unisensory and AV-trials were neither
432 influenced by task nor instruction, except Δ RT being lower in the visual go/no-go task than other
433 conditions. While false alarms to task-irrelevant stimuli contributed to changes in sensitivity,
434 they cannot directly explain the speed-up of responses.

435 **3.2. Instruction context**

436 To test the influence of instruction context, we focused on responses in unisensory trials (**Figure**
437 **1e**). We used RTs as the measure of speed and miss rates as the measure of performance
438 accuracy. To assess the influence of using unisensory or multisensory instructions on unisensory
439 trials (e.g., A-trials in A- and AUV-instruction blocks), we calculated performance differences
440 between uni- and multisensory instructions (for an illustration, see **Figure 3a**). If instruction
441 context invariance holds, performance in both instruction types should not differ.

442 --- Figure 3 about here ---

443 Regarding speed, we used a 2x2 ANOVA with factors task (detection, go/no-go) and trial
444 (A, V) to analyse changes in RT due to the changing instruction context (ΔRT ; **Figure 3b**). We
445 found a main effect of task ($F_{1,42}=87.999$, $p<0.001$, $\eta_p^2=0.677$), with ΔRT values deviating further
446 from zero in the go/no-go (-96.8 ± 7.1 ms) compared to that in the detection task (-30.2 ± 6.2 ms).
447 Hence, instruction context has a larger influence in the go/no-go than in the detection task. We
448 also found a main effect of trial ($F_{1,42}=30.200$, $p<0.001$, $\eta_p^2=0.418$), with ΔRT values deviating
449 further from zero in A-trials (-83.6 ± 7.7 ms) than in V-trials (-43.5 ± 5.4 ms). Hence, instruction
450 context has a larger influence in A- than in V-trials. The interaction of task and trial was not
451 significant ($F_{1,42}=0.007$, $p=0.934$, $\eta_p^2<0.001$). Most critically, we found a significant intercept
452 ($F_{1,42}=127.739$, $p<0.001$, $\eta_p^2=0.753$), with an average ΔRT value of -63.5 ms (± 5.6 ms). In addition,
453 post-hoc analyses revealed that only the ΔRT value in V-trials in the detection task was not
454 different from zero ($t_{42}=-1.481$, $p=0.146$). For all other three conditions, ΔRT values were
455 significantly different from zero (all $p<0.001$). As responses in the AUV-instructions were slower
456 compared to the corresponding unisensory instructions, instruction context invariance does not
457 hold in terms of RTs.

458 Regarding accuracy, we used a 2x2 ANOVA to analyse changes in miss rates ($\Delta Miss$;
459 **Figure 3c**). We found a main effect of task ($F_{1,42}=6.673$, $p=0.013$, $\eta_p^2=0.137$). $\Delta Miss$ values were
460 deviating further from zero in the go/no-go ($-3.8 \pm 0.6\%$) than in the detection task ($-2.5 \pm 0.5\%$).
461 Hence, instruction context has a larger influence in the go/no-go than in the detection task. We
462 also found a main effect of trial ($F_{1,42}=16.404$, $p<0.001$, $\eta_p^2=0.281$). $\Delta Miss$ values were deviating
463 further from zero in A-trials ($-4.5 \pm 0.7\%$) than V-trials ($-1.8 \pm 0.3\%$). Hence, instruction context
464 has a larger influence on A-trials than V-trials. We also found an interaction of task and trial
465 ($F_{1,42}=8.256$, $p=0.006$, $\eta_p^2=0.164$). $\Delta Miss$ values in A-trials were deviating less from zero in the
466 go/no-go ($-4.2 \pm 0.8\%$) than in the detection task ($-4.8 \pm 0.9\%$), whereas $\Delta Miss$ values in V-trials
467 ($-3.5 \pm 0.6\%$) were deviating further from zero in the go/no-go than in the detection task (-0.1

468 $\pm 0.4\%$). Hence, in *A*-trials, instruction context has a smaller influence on the go/no-go than the
469 detection task, and the opposite pattern revealed in *V*-trials. Most crucially, we found a
470 significant intercept ($F_{1,42}=47.774$, $p<0.001$, $\eta_p^2=0.532$), with an average ΔMiss value of -3.2%
471 ($\pm 0.5\%$). In addition, post-hoc analyses revealed that only the ΔMiss value in the *V*-instruction
472 block in the detection task was not different from zero ($t_{42}=-0.279$, $p=0.781$). For all other three
473 conditions, ΔMiss values were significantly different from zero (all $p<0.001$). As participants
474 missed more signals in *AUV*-instructions than unisensory instructions, instruction context
475 invariance does not hold in terms of accuracy.

476 Overall, in unisensory trials, performance in both speed and accuracy was worse in the
477 multisensory than in the unisensory instructions. In addition, ΔRT and ΔMiss were positively
478 correlated ($r=0.271$, $p<0.001$). The slowdown in speed was associated with the increase in miss
479 rates when using the *AUV*-instruction. This suggests that the *AUV*-instruction is more difficult
480 than the unisensory instructions, possibly as it is more demanding to monitor and detect targets
481 across different sensory modalities. Both ΔRT and ΔMiss values were task and trial dependent,
482 with larger effects in *A*- than in *V*-trials and in the go/no-go than in the detection task.

483 To further investigate the effect of instruction context, we considered the possibility
484 that trial history effects might influence performance differently for different instructions.
485 Performance on a given trial can be dependent on the previous trial, typically with better
486 performance in a repetition (e.g., a *V*-trial presented after a *V*-trial) than a switch trial (e.g., a *V*-
487 trial presented after an *A*-trial). To quantify such history cost, we computed the performance
488 difference between repetition and switch trials in each instruction block, which we express as
489 History values (**Equation 4**). To quantify how instruction context influence history cost, we
490 adopted a similar strategy as before and compared the difference in history cost in unisensory
491 and *AUV*-instruction ($\Delta\text{History}$; for an illustration, see **Figure 4a**). If history costs are unaffected
492 by instruction context, $\Delta\text{History}$ would be zero.

493

--- Figure 4 about here ---

494

495

496

497

498

499

500

501

502

For history cost in terms of speed, we used a 2x2 ANOVA to analyse Δ History (RT) values (**Figure 4b**). We found a main effect of task ($F_{1,42}= 12.257$, $p<0.001$, $\eta_p^2=0.226$). Δ History (RT) values were deviating further from zero in the go/no-go (-99.5 ± 12.3 ms) than in the detection task (-45.5 ± 8.6 ms). Hence, for history cost in terms of RTs, the instruction context has a larger influence in go/no-go than in detection tasks. Most critically, we found a significant intercept ($F_{1,42}=98.399$, $p< 0.001$, $\eta_p^2=0.701$), with an average Δ History (RT) value of -72.5 ms (± 7.3 ms). In addition, post-hoc analyses revealed that Δ History (RT) values in all conditions were different from zero (all $p<0.001$). As history cost in AUV-instructions were larger than that of unisensory instructions, instruction context invariance does not hold for history cost of RTs.

503

504

505

506

507

508

509

510

511

512

513

514

515

516

For history cost in terms of accuracy, we used a 2x2 ANOVA to analyse Δ History (Miss) values (**Figure 4c**). We found a main effect of trial ($F_{1,42}=11.451$, $p = 0.002$, $\eta_p^2=0.214$). Δ History (Miss) values were more negative in A-trials ($-7.2 \pm 1.4\%$) than V-trials ($-2.0 \pm 0.9\%$). Hence for history cost in terms of miss rates, instruction context has a larger influence on A-trials than V-trials. We also found an interaction of task and trial ($F_{1,42}=5.491$, $p=0.024$, $\eta_p^2=0.116$). Δ History (Miss) values in A-trials were less negative in the go/no-go ($-5.8 \pm 2.0\%$) than in the detection task ($-9.1 \pm 2.1\%$), whereas Δ History (Miss) values in V-trials ($-4.3 \pm 1.7\%$) were more negative in the go/no-go than in the detection task ($0.2 \pm 1.0\%$). Hence for miss rates in A-trials, instruction context has a smaller influence on the go/no-go than the detection task, and the opposite pattern revealed in V-trials. Most crucially, we found a significant intercept ($F_{1,42}=27.033$, $p<0.001$, $\eta_p^2=0.392$), with an average Δ History (Miss) value of -4.7% ($\pm 0.9\%$). In addition, post-hoc analyses revealed that only the Δ History (Miss) value in the V-instruction block in the detection task was not different from zero ($t_{42}=-0.237$, $p=0.814$). For all other three conditions, Δ History (Miss) values were significantly different from zero (all $p\leq 0.014$). As history costs were

517 larger in the AUV-instruction than in unisensory instruction blocks, instruction context
518 invariance does not hold for history cost of miss rates.

519 To explore the link between history cost and actual performance, we analysed the
520 correlation between the difference in performance and the difference in history cost in
521 unisensory and AUV-instructions. Regarding speed, Δ History (RT) values were positively
522 correlated to Δ RT values (**Figure 3b** and **Figure 4b**; $r = 0.178$, $p = 0.020$). Regarding accuracy, we
523 found a significant positive correlation between the Δ Miss and Δ History (Miss) values ($r = 0.309$,
524 $p < 0.001$). Hence, change in history cost is positively link to change in actual performance from
525 uni- to AUV-instructions in both speed and accuracy. Therefore, the violation of instruction
526 context invariance in actual performance can at least partially linked to differences in trial
527 history effects between uni- and AUV-instruction blocks.

528 **3.3. Miller's bound**

529 To test the influence of context on multisensory RT modelling, we investigated violations of
530 Miller's bound. With our extended paradigm, Miller's bound can be calculated using three
531 different pairs of unisensory RTs (**Figure 5a**; for additional details, see **Section 2.6**). First, it can
532 be computed using the CDFs in A- and V- trials under the AUV-instruction, which is the typical
533 scenario used in RSE experiments. We coin the pair of unisensory conditions used for model
534 predictions here as the stimulus-different pair because these conditions differ from the
535 predicted condition with redundant signals only in terms of sensory stimulation (unisensory vs.
536 AV-trials). Second, Miller's bound can be computed using the CDFs in A- and V-trials under the
537 A- and V-instruction, respectively. We coin this pair as the both-different pair because the
538 unisensory condition differs from the predicted redundant condition in terms of sensory
539 stimulation (unisensory vs. AV-trials) and instruction (unisensory vs. AUV-instruction). Third,
540 Miller's bound can be computed using the CDFs in AV-trials under the A- and V-instruction,
541 respectively. We coin this pair as the instruction-different pair because the unisensory condition

542 differs from the predicted redundant condition only in terms of instruction (unisensory vs. AUV-
543 instruction). For all three cases, we computed violations of Miller's bound using a geometrical
544 approach (**Equation 5**; for an illustration, see **Figure 5b**).

545 *--- Figure 5 about here ---*

546 We used a 2x3 ANOVA with factors task (detection, go/no-go) and type of unisensory
547 RT pair (stimulus-, both-, instruction-different) to analyse violations of Miller's bound (**Figure 5c**).
548 We found a main effect of pair type ($F_{1,146,48.124}=98.457$, $p<0.001$, $\eta_p^2=0.701$, Greenhouse-Geisser
549 corrected). Post-hoc pairwise comparisons revealed that all three pairs were significantly
550 different from each other (all $p<0.001$). Violations of Miller's bound decreased from the
551 stimulus-different pair (19.0 ± 1.4 ms) to the both-different pair (4.7 ± 0.92 ms), and further to
552 the instruction-different pair (1.6 ± 0.52 ms). We also found an interaction of task and pair type
553 ($F_{1,142,47.976}=8.986$, $p=0.003$, $\eta_p^2=0.176$). When the both-different pair and the instruction-
554 different pair were used, smaller violations occurred in the go/no-go compared to the detection
555 task. In contrast, when the stimulus-different pair was used, violations of Miller's bound were
556 larger in the go/no-go than in the detection task. When using the both-different pair and the
557 stimulus-different pair in the go/no-go task, violations of Miller's bound were not significantly
558 different from zero (all $t_{42}<1.858$, all $p>0.07$). In comparison, all the other pairs have Miller's
559 bound violation significantly different from zero (all $t_{42}>3.241$, all $p<0.002$).

560 In summary, predictions of multisensory RTs depend on the context of the unisensory
561 RTs that are fed to the model. In this example, violations of Miller's bound vary with the
562 unisensory pair type, which can eventually lead to fundamentally different conclusions. In the
563 extreme, the conclusion can flip from rejecting race models when using one pair to not rejecting
564 race models when using another pair. Hence, it is essential to consider the influence of context.

565 **4. Discussion**

566 This study emphasises the critical role of the context invariance assumption, which is frequently
567 made in multisensory modelling. The assumption states that processing of parallel decision units
568 is invariant to the experimental context, i.e., whether a single or a redundant signal is presented
569 (**Equation 2**). It is crucial to consider context invariance carefully, as its violation can lead to a
570 categorically different interpretation of multisensory benefits, for example whether the basic
571 architecture of race models is rejected, or not. While the assumption was thoroughly discussed
572 in early theoretical work (e.g., Ashby & Townsend, 1986; Colonius, 1990; Luce, 1986), it is silently
573 assumed in 84% of recent empirical contributions as reviewed in the introduction. Unfortunately,
574 it is difficult to test for context invariance empirically (Luce, 1986). To tackle this issue, we
575 assessed context invariance indirectly in two similar scenarios that can be tested: stimulus and
576 instruction context. If performance is generally constant when the experimental context is
577 changing, our results would provide at least some assurance that it is reasonably safe to assume
578 context invariance. In contrast, if performance is variant across changing contexts, our results
579 would highlight that the context invariance assumption is likely to be violated and should
580 therefore be considered more cautiously.

581 **4.1. Stimulus context**

582 As first scenario, we investigated performance under unisensory instructions when the stimulus
583 context is changing (**Figure 1d**). For example, for responses under A-instructions, we compared
584 performance in A- and AV-trials. The stimulus context is changing as the latter presented also a
585 task-irrelevant visual stimulus. We found a trade-off between speed and accuracy, responses in
586 multisensory trials were faster while sensitivity decreased (**Figure 2**). The message from the first
587 experiment is that performance can be variant when the stimulus context is changing.

588 Our findings are not unique, performance changes across different stimulus contexts
589 have frequently been reported in research on selective attention in multisensory perception.
590 For example, Talsma, Doty, and Woldorff (2006) instructed participants to attend to the visual
591 modality. The authors found that responses were slower and less accurate for audio-visual
592 compared to visual objects. As another example, Mozolic, Hugenschmidt, Peiffer, and Laurienti
593 (2008) asked participants to attend to stimuli in one modality. Here, responses were faster and
594 more accurate for audio-visual compared to visual stimuli. Finally, several studies on warning
595 signals have reported performance differences in visual tasks with and without task-irrelevant
596 auditory warning signals (e.g., Gleiss & Kayser, 2013; Kim, Peters, & Shams, 2012; Lippert,
597 Logothetis, & Kayser, 2007; Matusz, Wallace, & Murray, 2017; Murata, Kuroda, & Karwowski,
598 2017; Van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010). Given this broad range of
599 observations, it seems rather difficult to insist that performance in unisensory tasks is invariant
600 to stimulus context.

601 To understand the observed speed-accuracy trade-off in our experiment, we noted that
602 the change in sensitivity was largely due to an increased false alarm rate with task-irrelevant
603 signals. It is therefore possible that the speed-up of responses on AV-trials was caused by fast
604 but false responses to the task-irrelevant stimuli (e.g., the speed-up of RTs on AV-trials under A-
605 instructions could be caused by fast false alarms to the visual stimulus component). To check
606 this possibility, we compared genuine unisensory RTs (e.g., responses on A-trials under A-
607 instructions) to the timing of false alarms in unisensory trials that did not require a response
608 (e.g., on V-trials under A-instructions). However, we found these false alarms to be rather slower
609 than genuine RTs. Thus, the observed speed-up of RTs is not directly due to false alarms to task-
610 irrelevant signals. Some other change in processing is required to explain the change of
611 performance.

612 On hypothesis is temporal preparation in that an earlier presented and/or faster
613 processed signal in one modality facilitates processing of a signal in another modality (Los & Van
614 der Burg, 2013). For example, if the task-irrelevant component on AV-trials (e.g., the visual
615 component under A-instructions) is processed faster than the task-relevant component (e.g.,
616 the auditory component under A-instructions), the former could facilitate processing of the task-
617 relevant signal through an effective preparation period. In contrast, if it is processed slower, no
618 facilitation is expected. Interestingly in this regard, we found RTs in the go/no-go task to be
619 faster on V- compared to A-trials. Consequently, the temporal preparation hypothesis would
620 predict a larger speed-up on AV-trials under A- compared to V-instructions. In contrast, in the
621 detection task, we found no difference in RTs. Consequently, the speed-up is not expected to
622 differ either. Our results are in accordance with these predictions (**Figure 2b**). Hence, the
623 violation of stimulus context invariance could be explained by the temporal preparation
624 hypothesis, which not necessarily requires integration or pooling of sensory evidence.

625 **4.2. Instruction context**

626 As second scenario, we investigated responses to unisensory trials when the instruction context
627 is changing (**Figure 1e**). For example, for responses to A-trials, we compared performance under
628 A- and AUV-instructions. The instruction context is changing as the latter requires to attend also
629 to a second stimulus (which is however only present on other trials). We found that performance
630 under AUV-instructions was slower and had higher miss rates than that under unisensory
631 instructions (except for V-trials in the detection task; see **Figure 3**). The message from the second
632 scenario is that performance with unisensory signals can be variant when the instruction context
633 is changing.

634 Our findings are not unique here either, performance changes under different
635 instructions have frequently been reported in studies exploring the influence of selective
636 attention in multisensory perception. For example, Talsma et al. (2006) recorded responses to

637 visual objects. They found responses were slower and with more errors under the instruction to
638 attend to both audition and vision compared to attending to vision only. In addition, they found
639 a decrease in the P1 component of the ERP signal across the two attentional instructions. As
640 another example, Mozolic, Joyner, et al. (2008) adopted a similar task and found decreased
641 sensitivity to auditory targets under an audio-visual than an auditory attention task. They related
642 this to the suppression of visual cortex activity when participants were instructed to attend to
643 audition only. Given these congruent findings, it seems rather difficult to insist on the
644 assumption of instruction context invariance on presentation of unisensory signals.

645 To understand the performance changes in our experiment, we explored the link to
646 sequential effects. When analysed as a function of trial history, performance is typically better
647 when a modality is repeated (e.g., when a V-trial is preceded by a V-trial) than when modalities
648 are switched (e.g., when a V-trial is preceded by an A-trial), which can be expressed as trial
649 history costs (for examples in RSE studies, see Gondan, Lange, Rösler, & Röder, 2004; Gondan,
650 Vorberg, & Greenlee, 2007; Innes & Otto, 2019; Miller, 1982; Otto & Mamassian, 2012). The
651 question is whether sequential effects account for the performance changes as observed with
652 different instructions (**Figure 3**). Interestingly, we found that the change in history costs basically
653 mirrored the performance changes when the instruction context is changed (**Figure 4**). The
654 changes in sequential effects can be understood from an attentional perspective. Under
655 unisensory instructions, participants can attend to only one modality, which allows to suppress
656 processing in the other (Mozolic, Hugenschmidt, et al., 2008; Mozolic, Joyner, et al., 2008). For
657 example, A-instructions allow participants to focus on audition and to suppress any visual signal.
658 In this case, modality switch costs should be low because there is actually no need to switch
659 attention between modalities. In contrast, AUV-instructions ask participants to divide attention
660 across audition and vision (Spence, Nicholls, & Driver, 2001). In this case, modality switch costs
661 should be higher because attention needs constantly to switch as signals are presented in

662 different modalities. This analysis suggests that the violation of instruction context invariance is
663 at least partly due to changes in sequential effects.

664 An interesting observation is that responses to auditory stimuli are more affected by
665 sequential effects than responses to visual stimuli. It may be tempting to speculate here about
666 a potential link to the Colavita visual dominance effect. When instructed to respond to auditory
667 and visual signals with separate behavioural responses, the observation is that participants
668 respond more frequently only to the visual component when both stimuli are presented
669 (Colavita, 1974; Sinnett, Spence, & Soto-Faraco, 2007). As this paradigm also requires that the
670 two signals are presented in a random order, future research may investigate if sequential
671 dependency contributes to the Colavita effect.

672 **4.3. Considering contextual influence in models of multisensory decisions**

673 As both stimulus and instruction context invariance are violated, our results demonstrate that
674 context invariance cannot generally assumed to be true. To illustrate the importance of the
675 assumption when interpreting multisensory RTs (AV-trials under AUV-instructions in our
676 experiment), we tested for violations of Miller's bound, which is widely used to reject race
677 models (see the many follow-up studies of Miller, 1982). Critically, our experimental design
678 allows deriving Miller's bounds using three different pairs of unisensory RTs, which differ in
679 stimulus context, instruction context, or both (see **Section 2.6** and **Figure 5a**). If context
680 invariance would be generally true, it should not matter which pair is used to compute the bound.
681 However, this is not what we found. Violations of Miller's bound varied when computed with
682 different unisensory pairs, with violations being largest with the most frequently used
683 instruction-different pair (**Figure 5b**). In contrast, no significant violations occurred with the
684 both-different and the stimulus-different pairs in the go/no-go task. If the role of context is not
685 considered, these differences result in conflicting conclusions. While one test seems to suggest
686 that race models must be rejected, the other would see exactly the same behavioural data to

687 be in agreement with race models³. To resolve this issue, it is crucial to consider the influence of
688 context in multisensory models.

689 Unfortunately, current research practice basically neglects contextual effects. As
690 reviewed in the introduction, 84% of the RSE studies published between 2016 and 2018 seem
691 to be unaware of the context invariance assumption (at least it is not mentioned). What is more,
692 if it is considered, the importance of the assumption is rather downplayed. For example, without
693 naming the assumption explicitly, Chandrasekaran (2017) simply argued that when ‘appropriate
694 care is taken to arrange experimental conditions’, race models can be rejected using Miller’s test.
695 However, the approach of how to appropriately arrange experimental conditions is left vague.
696 Alternatively, race models that allow for context variance are criticized as being ‘unfalsifiable’
697 (Miller, 2016). However, we think that this criticism applies more urgently to the large body of
698 studies that reject race models without providing an alternative model that allows for testable
699 predictions (Otto & Mamassian, 2017). Simply stating that the RSE is due to ‘multisensory
700 integration’ does not provide explanation, unless the integration mechanism is explicitly
701 specified by a testable model. The key point made by this study is that any model of the RSE,
702 whether using a race or pooling architecture, needs to consider contextual effects.

703 To understand multisensory interactions, it may be particularly interesting to compare
704 conditions that do and do not violate Miller’s bound. In this respect, as discussed above, no
705 significant violations occurred with the both-different and stimulus-different pairs in the go/no-
706 go task (**Figure 5b**). In contrast, violations occurred with the corresponding pairs in the detection
707 task. Using what they called a comparative approach, Innes and Otto (2019) discussed a link
708 between violations and transient stimulus onsets and a similar explanation could apply here. As
709 presented in the introduction, single and redundant signal conditions differ in the number of

³ Finding satisfaction of Miller’s bound (affirming the consequent) does not show that the race architecture is necessary true. The RSE could be still explained by a pooling model (e.g., Diederich & Colonius, 2004b).

710 onset transients in the detection task (one vs. two transients). On the other hand, the number
711 of onset transients is constant in the go/no-go task (there are two transients in both single and
712 redundant conditions because the former also present a nontarget signal). Using a similar logic,
713 it is further interesting to note that violations in both tasks are largest when the instruction-
714 different pair is used to compute Miller's bound (**Figure 5b**). As these conditions are heavily
715 affected by trial history effects (**Figure 4**), this finding highlights that any modelling approach,
716 be it with a race or pooling architecture, should seriously consider the contribution of sequential
717 dependency to the RSE.

718 **4.4. Conclusion**

719 To advance research on multisensory processing, we argue that the basic race model
720 architecture, despite being frequently rejected, is actually very helpful to examine contextual
721 effects. A first issue is that the basic race model architecture perfectly matches the task demands
722 of the redundant signals paradigm (Otto & Mamassian, 2017). Intriguingly then, the basic race
723 model architecture, for example in form of the independent race model (Raab, 1962), directly
724 allows for predictions on the level of RT distributions. These predictions can be used as
725 benchmark performance. The quest is then not necessarily to reject the basic race model
726 architecture but rather to understand deviances from the benchmark, which can be interpreted
727 as the influence of context as a valid alternative. This approach was possibly first formulated by
728 Townsend and co-workers by quantifying the deviance from the independent race model in
729 terms of processing capacity, which led to *systems factorial technology* as a diagnostic tool (e.g.,
730 Little, Altieri, Fific, & Yang, 2017; Townsend & Nozawa, 1997; Townsend & Wenger, 2004;
731 Wenger & Townsend, 2000; Yang et al., 2018). Instead of discriminating between processing
732 architectures (e.g., pooling vs. racing), this approach classifies performance with respect to
733 workload capacity (e.g., limited vs. super-capacity using the independent race model as baseline;
734 for a recent application to study the effect of aging, see Yang et al., 2019). In some sense

735 elaborating on this approach, Otto and Mamassian (2012) explicitly modelled a violation of the
736 context invariance assumption by adding noise as a single interaction parameter. This work not
737 only demonstrated that the RSE can be explained with a race model despite violations of Miller's
738 bound, but also allows to systematically study deviances from the benchmark using the model
739 as a tool in a comparative approach (Innes & Otto, 2019). In the aggregate, we are convinced
740 that both approaches, systems factorial technology and explicit modelling using the race
741 architecture, will be very helpful to understand the processing of multisensory signals when
742 tested with the redundant signals paradigm.

743 To advance multisensory modelling, we think that huge progress can be made by
744 comparing different accounts of the RSE. As a first step in this endeavour, it is important to be
745 clear about definitions. For example in the process of model development, it is crucial to decide
746 about the cognitive architecture that is involved when making multisensory decisions (Otto &
747 Mamassian, 2017). Is multisensory evidence pooled in a single decision unit, or is it a race
748 between two parallel decision units coupled by an OR gate? Unfortunately, definitions are not
749 always clear with respect to this partition of the model space. For example, violations of Miller's
750 inequality are considered as evidence for coactivation (Miller, 1982). In the following decades,
751 the term "coactivation" has been used rather consistently, at least with some connotation, to
752 describe models with a single decision unit (e.g., Blurton et al., 2014; Colonius & Diederich, 2017;
753 Diederich, 1994; Miller, 1982; Mordkoff & Yantis, 1991; Schwarz & Ischebeck, 1994; Townsend
754 & Wenger, 2004; Ulrich & Giray, 1986; Yang et al., 2018; Zehetleitner et al., 2009). Only recently,
755 Miller (2016), who originally coined the term in 1982, clarified that the term should not be used
756 to describe basic model architecture, which is the reason why we use the term "pooling" to
757 explicitly describe models with a single decision unit. With clarified definitions at hand, as a next
758 key step, it is of course of major interest to directly compare pooling and racing models and to
759 trigger in multisensory research a true competition: Which model explains the data better?

760 5. Acknowledgements

761 We would like to thank Bobby Innes and Cleo Pike for their feedback on an earlier draft of the
762 manuscript. Yue Liu was supported by studentship sponsored by the School of Psychology and
763 Neuroscience, University of St Andrews.

764 6. Reference

- 765 Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: from
766 physiology to behaviour. *Seeing and perceiving*, 23(1), 3-38.
767 doi:<http://doi.org/10.1163/187847510X488603>
- 768 Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological*
769 *review*, 93(2), 154-179. doi:<https://doi.org/10.1037/0033-295X.93.2.154>
- 770 Blurton, S. P., Greenlee, M. W., & Gondan, M. (2014). Multisensory processing of redundant
771 information in go/no-go and choice responses. *Atten Percept Psychophys*, 76(4), 1212-
772 1233. doi:<https://doi.org/10.3758/s13414-014-0644-0>
- 773 Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the
774 inverse efficiency score (IES) a better dependent variable than the mean reaction time
775 (RT) and the percentage of errors (PE)? *Psychologica Belgica*, 51(1), 5-13.
- 776 Chandrasekaran, C. (2017). Computational principles and models of multisensory integration.
777 *Current Opinion in Neurobiology*, 43, 25-34.
778 doi:<http://doi.org/10.1016/j.conb.2016.11.002>
- 779 Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, 16(2), 409-412.
780 doi:<https://doi.org/10.3758/BF03203962>
- 781 Colonius, H. (1990). Possibly Dependent Probability Summation of Reaction-Time. *Journal of*
782 *Mathematical Psychology*, 34(3), 253-275. doi:[http://doi.org/10.1016/0022-](http://doi.org/10.1016/0022-2496(90)90032-5)
783 [2496\(90\)90032-5](http://doi.org/10.1016/0022-2496(90)90032-5)
- 784 Colonius, H., & Diederich, A. (2006). The race model inequality: interpreting a geometric
785 measure of the amount of violation. *Psychological review*, 113(1), 148-154.
786 doi:<http://doi.org/10.1037/0033-295X.113.1.148>
- 787 Colonius, H., & Diederich, A. (2017). Measuring multisensory integration: from reaction times to
788 spike counts. *Sci Rep*, 7(1), 3023. doi:<https://doi.org/10.1038/s41598-017-03219-5>

789 Colonius, H., & Diederich, A. (2018). Formal models and quantitative measures of multisensory
790 integration: a selective overview. *European Journal of Neuroscience*. doi:
791 <https://doi.org/10.1111/ejn.13813>

792 Crosse, M. J., Foxe, J. J., & Molholm, S. (2019). Developmental Recovery of Impaired
793 Multisensory Processing in Autism and the Cost of Switching Sensory Modality. *bioRxiv*,
794 565333. doi:<https://doi.org/10.1101/565333>

795 Diederich, A. (1994). A diffusion model for intersensory facilitation of reaction time. In
796 *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 207-
797 220): Springer.

798 Diederich, A., & Colonius, H. (2004a). Bimodal and trimodal multisensory enhancement: effects
799 of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, 66(8),
800 1388-1404. doi:<https://doi.org/10.3758/BF03195006>

801 Diederich, A., & Colonius, H. (2004b). Modeling the Time Course of Multisensory Interaction in
802 Manual and Saccadic Responses.

803 Giray, M., & Ulrich, R. (1993). Motor coactivation revealed by response force in divided and
804 focused attention. *Journal of Experimental Psychology: Human Perception and*
805 *Performance*, 19(6), 1278-1291. doi:<https://doi.org/10.1037/0096-1523.19.6.1278>

806 Gleiss, S., & Kayser, C. (2013). Eccentricity dependent auditory enhancement of visual stimulus
807 detection but not discrimination. *Frontiers in integrative neuroscience*, 7, 52.
808 doi:<http://doi.org/10.3389/fnint.2013.00052>

809 Gondan, M., Lange, K., Rösler, F., & Röder, B. (2004). The redundant target effect is affected by
810 modality switch costs. *Psychonomic bulletin & review*, 11(2), 307-313.
811 doi:<https://doi.org/10.3758/BF03196575>

812 Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention,*
813 *Perception, & Psychophysics*, 78(3), 723-735. doi:[http://doi.org/10.3758/s13414-015-](http://doi.org/10.3758/s13414-015-1018-y)
814 [1018-y](http://doi.org/10.3758/s13414-015-1018-y)

815 Gondan, M., Vorberg, D., & Greenlee, M. W. (2007). Modality shift effects mimic multisensory
816 interactions: an event-related potential study. *Experimental brain research*, 182(2), 199-
817 214. doi:<http://doi.org/10.1007/s00221-007-0982-4>

818 Gordon, I. E. (1967). Stimulus probability and simple reaction time. *Nature*, 215(5103), 895-896.
819 doi:<http://doi.org/10.1038/215895a0>

820 Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated
821 values of d' . *Behavior Research Methods, Instruments, & Computers*, 27(1), 46-51.
822 doi:<http://doi.org/10.3758/bf03203619>

823 Hershenson, M. (1962). Reaction-Time as a Measure of Intersensory Facilitation. *Journal of*
824 *Experimental Psychology*, 63(3), 289-&. doi:<http://doi.org/10.1037/h0039516>

825 Innes, B. R., & Otto, T. U. (2019). A comparative analysis of response times shows that
826 multisensory benefits and interactions are not equivalent. *Scientific reports*, 9(1), 2921.
827 doi:<http://doi.org/10.1038/s41598-019-39924-6>

828 Kim, R., Peters, M. A., & Shams, L. (2012). $0 + 1 > 1$: How adding noninformative sound improves
829 performance on a visual task. *Psychological science*, 23(1), 6-12.
830 doi:<http://doi.org/10.1177/0956797611420662>

831 Kinchla, R. (1974). Detecting target elements in multielement arrays: A confusability model.
832 *Perception & Psychophysics*, 15(1), 149-158. doi:<https://doi.org/10.3758/BF03205843>

833 Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard
834 deviation around the mean, use absolute deviation around the median. *Journal of*
835 *Experimental Social Psychology*, 49(4), 764-766.
836 doi:<http://doi.org/10.1016/j.jesp.2013.03.013>

837 Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by
838 a simultaneous sound. *Brain research*, 1173, 102-109.
839 doi:<http://doi.org/10.1016/j.brainres.2007.07.050>

840 Little, D., Altieri, N., Fific, M., & Yang, C. T. (2017). *Systems factorial technology: A theory driven*
841 *methodology for the identification of perceptual and cognitive mechanisms*: Academic
842 Press.

843 Los, S. A., & Van der Burg, E. (2013). Sound speeds vision through preparation, not integration.
844 *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), 1612-
845 1624. doi:<http://doi.org/10.1037/a0032183>

846 Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*:
847 Oxford University Press on Demand.

848 Lunn, J., Sjoblom, A., Ward, J., Soto-Faraco, S., & Forster, S. (2019). Multisensory enhancement
849 of attention depends on whether you are already paying attention. *Cognition*, 187, 38-
850 49. doi:<http://doi.org/10.1016/j.cognition.2019.02.008>

851 Matusz, P. J., Wallace, M. T., & Murray, M. M. (2017). A multisensory perspective on object
852 memory. *Neuropsychologia*, *105*, 243-252.
853 doi:<http://doi.org/10.1016/j.neuropsychologia.2017.04.008>

854 Miller, J. (1982). Divided Attention - Evidence for Co-Activation with Redundant Signals.
855 *Cognitive Psychology*, *14*(2), 247-279. doi:[http://doi.org/10.1016/0010-0285\(82\)90010-](http://doi.org/10.1016/0010-0285(82)90010-X)
856 [X](http://doi.org/10.1016/0010-0285(82)90010-X)

857 Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception &*
858 *Psychophysics*, *40*(5), 331-343. doi:<https://doi.org/10.3758/BF03203025>

859 Miller, J. (2016). Statistical facilitation and the redundant signals effect: What are race and
860 coactivation models? *Attention, Perception, & Psychophysics*, *78*(2), 516-519.
861 doi:<http://doi.org/10.3758/s13414-015-1017-z>

862 Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of*
863 *Experimental Psychology: Human Perception and Performance*, *17*(2), 520-538.
864 doi:<https://doi.org/10.1037/0096-1523.17.2.520>

865 Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific
866 selective attention attenuates multisensory integration. *Experimental brain research*,
867 *184*(1), 39-52. doi:<https://doi.org/10.1007/s00221-007-1080-3>

868 Mozolic, J. L., Joyner, D., Hugenschmidt, C. E., Peiffer, A. M., Kraft, R. A., Maldjian, J. A., &
869 Laurienti, P. J. (2008). Cross-modal deactivations during modality-specific selective
870 attention. *BMC neurology*, *8*(1), 35. doi:<https://doi.org/10.1186/1471-2377-8-35>

871 Murata, A., Kuroda, T., & Karwowski, W. (2017). Effects of auditory and tactile warning on
872 response to visual hazards under a noisy environment. *Applied ergonomics*, *60*, 58-67.
873 doi:<http://doi.org/10.1016/j.apergo.2016.11.002>

874 Murray, M. M., Eardley, A. F., Edgington, T., Oyekan, R., Smyth, E., & Matusz, P. J. (2018). Sensory
875 dominance and multisensory integration as screening tools in aging. *Scientific reports*,
876 *8*(1), 8901. doi:<http://doi.org/10.1038/s41598-018-27288-2>

877 Näätänen, R. (1972). Time uncertainty and occurrence uncertainty of the stimulus in a simple
878 reaction time task. *Acta Psychologica*, *36*(6), 492-503. doi:[http://doi.org/10.1016/0001-](http://doi.org/10.1016/0001-6918(72)90029-7)
879 [6918\(72\)90029-7](http://doi.org/10.1016/0001-6918(72)90029-7)

880 Otto, T. U. (2019). RSE-box: An analysis and modelling package to study response times to
881 multiple signals. . *The Quantitative Methods for Psychology*, *in press*.

882 Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *Journal of*
883 *Neuroscience*, *33*(17), 7463-7474. doi:<http://doi.org/10.1523/JNEUROSCI.4678-12.2013>

- 884 Otto, T. U., & Mamassian, P. (2010). *Noise vs. multisensory integration: The return of the race*
885 *model*. Paper presented at the 11th International Multisensory Research Forum (IMRF),
886 Liverpool, UK.
- 887 Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision
888 making. *Current Biology*, 22(15), 1391-1396.
889 doi:<http://doi.org/10.1016/j.cub.2012.05.031>
- 890 Otto, T. U., & Mamassian, P. (2017). Multisensory Decisions: the Test of a Race Model, Its Logic,
891 and Power. *Multisensory Research*, 30(1), 1-24. doi:[http://doi.org/10.1163/22134808-](http://doi.org/10.1163/22134808-00002541)
892 [00002541](http://doi.org/10.1163/22134808-00002541)
- 893 Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York*
894 *Academy of Sciences*, 24(5), 574-590. doi:[http://doi.org/10.1111/j.2164-](http://doi.org/10.1111/j.2164-0947.1962.tb01433.x)
895 [0947.1962.tb01433.x](http://doi.org/10.1111/j.2164-0947.1962.tb01433.x)
- 896 Schwarz, W., & Ischebeck, A. (1994). Coactivation and statistical facilitation in the detection of
897 lines. *Perception*, 23(2), 157-168. doi:<https://doi.org/10.1068/p230157>
- 898 Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: the Colavita
899 effect revisited. *Percept Psychophys*, 69(5), 673-686.
900 doi:<https://doi.org/10.3758/BF03193770>
- 901 Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory
902 modality. *Perception & Psychophysics*, 63(2), 330-336.
903 doi:<https://doi.org/10.3758/BF03194473>
- 904 Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the
905 perspective of the single neuron. *Nature Reviews Neuroscience*, 9(4), 255-266.
906 doi:<http://doi.org/10.1038/nrn2331>
- 907 Talsma, D., Doty, T. J., & Woldorff, M. G. (2006). Selective attention and audiovisual integration:
908 is attending to both modalities a prerequisite for early integration? *Cerebral cortex*,
909 17(3), 679-690. doi:<https://doi.org/10.1093/cercor/bhk016>
- 910 Todd, J. W. (1912). *Reaction to multiple stimuli*: Science Press.
- 911 Townsend, J. T., & Nozawa, G. (1997). Serial exhaustive models can violate the race model
912 inequality: Implications for architecture and capacity. *Psychological review*, 104(3), 595-
913 602. doi:<http://doi.org/10.1037/0033-295x.104.3.595>
- 914 Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: new capacity
915 measures and predictions for a response time inequality series. *Psychological review*,
916 111(4), 1003-1035. doi:<http://dx.doi.org/10.1037/0033-295x.111.4.1003>

917 Ulrich, R., & Giray, M. (1986). Separate-activation models with variable base times: testability
918 and checking of cross-channel dependency. *Percept Psychophys*, 39(4), 248-254.
919 doi:<https://doi.org/10.3758/bf03204931>

920 Ulrich, R., Miller, J., & Schröter, H. (2007). Testing the race model inequality: an algorithm and
921 computer programs. *Behavior Research Methods*, 39(2), 291-302.
922 doi:<https://doi.org/10.3758/BF03193160>

923 Ursino, M., Cuppini, C., & Magosso, E. (2014). Neurocomputational approaches to modelling
924 multisensory integration in the brain: a review. *Neural Networks*, 60, 141-165.
925 doi:<http://doi.org/10.1016/j.neunet.2014.08.003>

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

929 Van der Stoep, N., Van der Stigchel, S., Van Engelen, R. C., Biesbroek, J. M., & Nijboer, T. C. W.
930 (2019). Impairments in Multisensory Integration after Stroke. *Journal of cognitive*
931 *neuroscience*, 31(6), 885-899. doi:http://doi.org/10.1162/jocn_a_01389

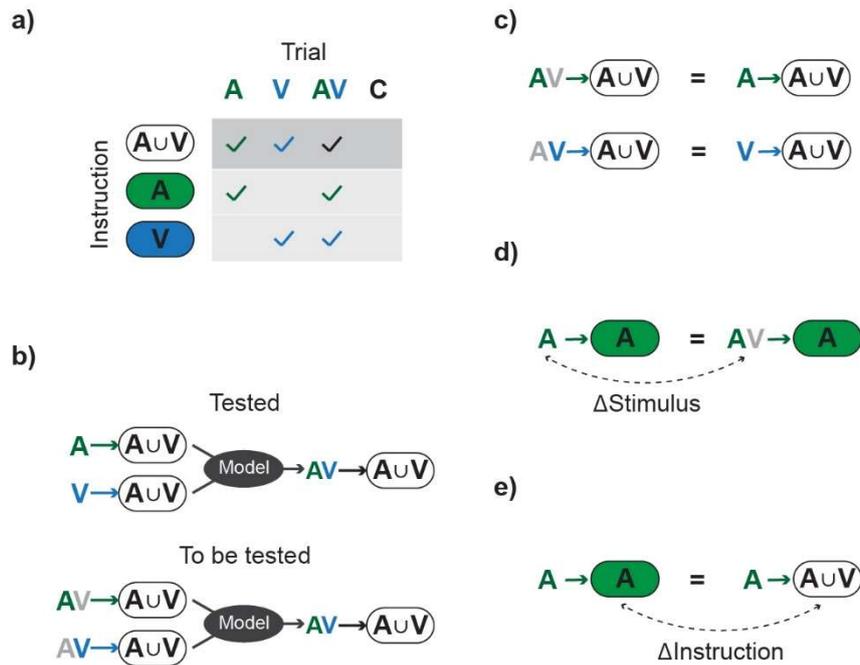
932 Wenger, M. J., & Townsend, J. T. (2000). Basic response time tools for studying general
933 processing capacity in attention, perception, and cognition. *J Gen Psychol*, 127(1), 67-99.
934 doi:<https://doi.org/10.1080/00221300009598571>

935 Yang, C. T., Altieri, N., & Little, D. R. (2018). An examination of parallel versus coactive processing
936 accounts of redundant-target audiovisual signal processing. *Journal of Mathematical*
937 *Psychology*, 82, 138-158. doi:<http://doi.org/10.1016/j.jmp.2017.09.003>

938 Yang, C. T., Hsieh, S., Hsieh, C. J., Fifić, M., Yu, Y. T., & Wang, C. H. (2019). An examination of age-
939 related differences in attentional control by systems factorial technology. *Journal of*
940 *Mathematical Psychology*, 92, 102280. doi:<https://doi.org/10.1016/j.jmp.2019.102280>

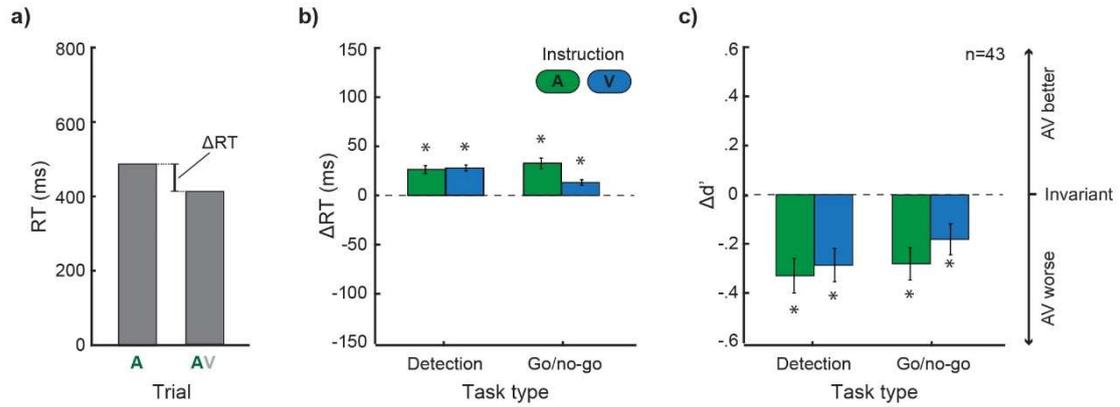
941 Zehetleitner, M., Krummenacher, J., & Müller, H. J. (2009). The detection of feature singletons
942 defined in two dimensions is based on salience summation, rather than on serial
943 exhaustive or interactive race architectures. *Attention, Perception, & Psychophysics*,
944 71(8), 1739-1759. doi:<http://doi.org/10.3758/APP.71.8.1739>

945
946



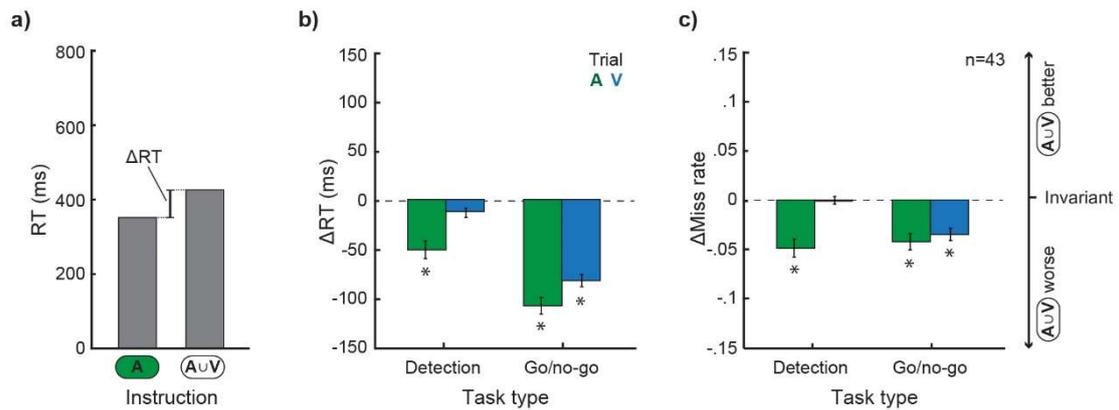
947

948 **Figure 1.** Assumptions and models of multisensory decision making. **a)** Experimental paradigm. Trials
 949 contained either an auditory (A), a visual (V), or a combined audio-visual signal (AV). Catch trials (C)
 950 presented no signal. We used three different instructions. In the AUV-instruction, which is the classic
 951 redundant signal paradigm (dark grey area), participants responded to any signal (ticks) and withheld a
 952 response on catch trials (no tick). In the A-instruction (V-instruction), participants responded only to
 953 auditory (visual) signals. The extended paradigm allows comparing unisensory auditory (visual) decisions
 954 in three scenarios as highlighted by green (blue) ticks. **b)** Testing race models. The top panel shows what
 955 is tested with Miller's (1982) race model inequality (Equation 1), which uses the unisensory conditions in
 956 the AUV-instruction to derive race model predictions for redundant signals. The bottom panel shows what
 957 actually needs to be tested to reject the basic race model architecture as an explanation of the RSE. This
 958 test should use the unobserved unisensory decision times to the auditory and visual components of the
 959 AV signal (grey letters indicate that the other component is physically present but not used for the
 960 decision in the other modality). **c)** Context invariance assumption. When using Miller's test to reject race
 961 models as an explanation of the RSE, the context invariance assumption is made (Equation 2). The
 962 assumption allows to link top and bottom panels in b). **d)** Stimulus context. The extended paradigm allows
 963 comparing unisensory decisions when the stimulus context is changed (e.g., from A- to AV-trials). **e)**
 964 Instruction context. Likewise, the extended paradigm allows comparing unisensory decisions when the
 965 instruction context is changed (e.g., from A- to AUV-instructions).



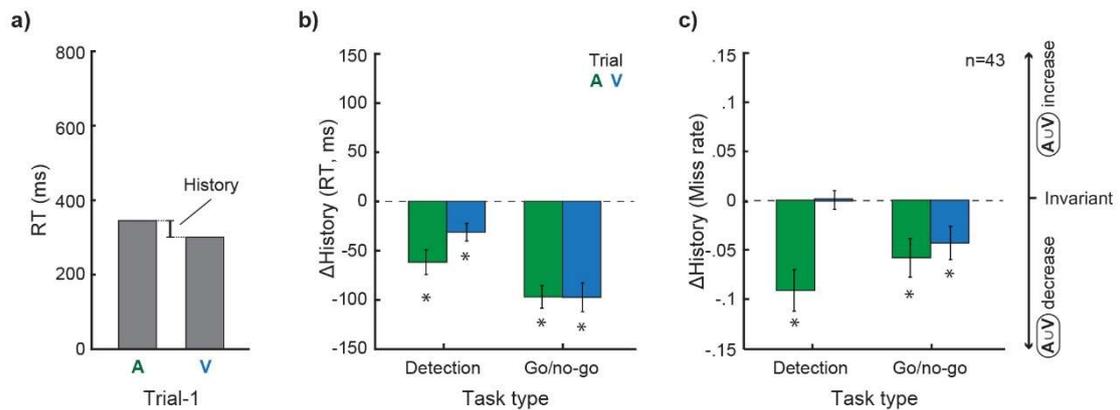
966

967 **Figure 2.** Stimulus context effects. **a)** Computation of the change in RT (ΔRT) when the stimulus is changing
 968 from unisensory to AV. The graph shows example data from a participant in the A-instruction in the
 969 detection task. Changes in sensitivity ($\Delta d'$) are calculated in analogously. **b)** ΔRT and **c)** $\Delta d'$ as a function
 970 of task and unisensory instruction. If Δ -values are zero, performance is invariant to the stimulus context.
 971 Positive (negative) values indicate better (worse) performance in AV-trials. Mean and SEM of 43
 972 participants.



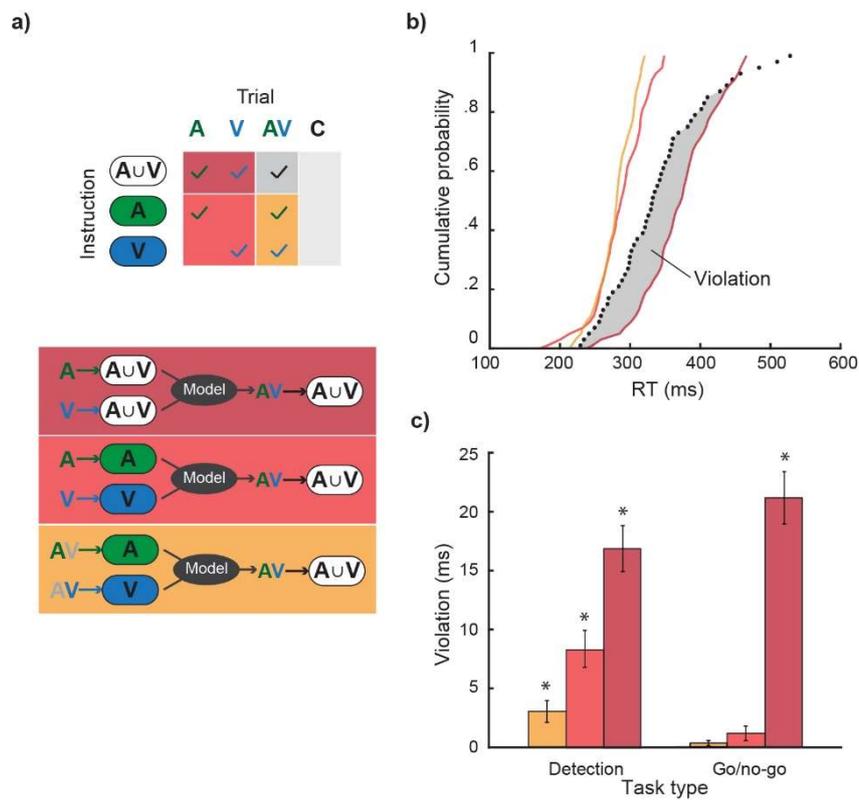
973

974 **Figure 3.** Instruction context effects. **a)** Computation of the change in RT (ΔRT) when the instruction is
 975 changing from unisensory to AUV. The graph shows example data from a participant in A-trials in the
 976 detection task. Changes in miss rates ($\Delta Miss$) are calculated analogously. **b)** ΔRT and **c)** $\Delta Miss$ as a function
 977 of task and unisensory trial type. If Δ -values are zero, performance is invariant to the instruction context.
 978 Positive (negative) values indicate better (worse) performance in the AUV-instruction. Mean and SEM of
 979 43 participants.



980

981 **Figure 4.** Instruction context effects on trial history costs. **a)** Computation of trial history cost. The graph
 982 shows example V-trial RTs from a participant performing the detection task in the V-instruction. History
 983 costs are computed as the difference in RTs on trials preceded by A- and V- trials, respectively. History
 984 costs in miss rates are computed analogously. **b)** Change in history effect on RTs and **c)** miss rates as a
 985 function of task and unisensory trial type. If Δ History values are zero, history costs are invariant to the
 986 instruction context. Positive (negative) values indicate reduced (increased) history costs in the AUV-
 987 instruction. Mean and SEM of 43 participants.



988

989 **Figure 5.** Testing race models. **a)** Different unisensory RT pairs can be used to compute Millar’s bound,
 990 which provides an upper bound for RTs to redundant signals (AV-trials, AUV-instruction). The stimulus-
 991 different pair (dark red) differs from the predicted condition with redundant signals only in stimulation
 992 (unisensory trials are used to predict AV-trials). The both-different pair (red) differs both in stimulation
 993 (as before) and in instruction (unisensory instructions are used to predict responses under the AUV-
 994 instruction). The instruction-different pair (orange) differs only in instruction. **b)** Miller’s bound computed
 995 using the three different unisensory RT pairs. The empirical RT distribution with redundant signals is given
 996 by black dots. Violations of Miller’s bound are quantified by the area between the empirical distribution
 997 and the bound (grey area). Example data from a participant performing in the detection task. **c)** Violations
 998 of Miller’s bound as a function of task and unisensory RT pair. Mean and SEM of 43 participants.

999 **Table 1.** Number of trials per condition in one block. Each block was composed of 30 target and 30 non-
 1000 target trials.

Instruction	Target trials			Non-target trials			Total
	A	V	AV	A	V	Catch	
AUV	10	10	10	-	-	30	60
A	15	-	15	-	15	15	60
V	-	15	15	15	-	15	60

1001 **Table 2.** Unisensory RTs compared to corresponding false alarm times (mean \pm SEM). The comparison is
 1002 performed only for participants who have committed at least one false alarm.

Task	Instruction	Valid responses		False alarms			t	df	p
		Trial	RT (ms)	Trial	N [#]	Timing (ms)			
Detection	A	A	430 \pm 86	V	34	501 \pm 74	-3.159	33	0.002
	V	V	418 \pm 98	A	26	499 \pm 82	-2.517	25	0.009
Go/no-go	A	A	452 \pm 77	V	36	463 \pm 75	-0.551	35	0.292
	V	V	432 \pm 94	A	27	488 \pm 83	-3.331	26	0.002

1003 [#] Number of participants who committed false alarms.