Evaluation of Cognitive Processing in Redundant Audio−Visual Signals

Elizabeth L. Fox (fox.119@wright.edu)
Joseph J. Glavan (glavan.3@wright.edu)
Joseph W. Houpt (joseph.houpt@wright.edu)
Wright State University, 3640 Colonel Glenn Highway
Dayton, Ohio 45435 USA

Abstract

The goal of the present effort was to revisit Miller's (1982) claim that audio-visual stimuli are processed by a coactive architecture. We replicated Miller's analysis and extended it using both group and individual level measures from Systems Factorial Technology (SFT; Townsend & Nozawa, 1995). Similar to previous findings, some participants exhibited redundancy gain beyond that predicted by independent parallel processing. However, the majority of participants performed no better than would be predicted by an independent parallel model, and some even performed worse than independent parallel. Furthermore, the variation observed across individual participants suggests that individual level performance measures are at least as important as group measures for the robust interpretation of human information processing data.

Keywords: System factorial technology; human information processing; multimodal; race model inequality

Introduction

The rapid growth in the need for task efficiency makes the development of systems that combine multiple sources of information essential. Systems that combine multiple components, or modalities, can enhance user performance by speeding up reaction times or increasing accuracy in a given task. However, "there must be limits for optimality and conditions under which sensory integration is not the best strategy" (Ernst & Bülthoff, 2004). It may be the case that when multiple sources of information involve conflicting cognitive pathways, they increase cognitive workload and potentially harm user performance. In order to hone in on cognitive processing of multisensory information we will specifically focus on situations in which two modalities, each supplying a single target stimulus, co-occur to prompt a single response, a pattern referred to as redundant signals. In nature when an audio and a visual signal co-occur, they are often due to the same cause. Hence, it is plausible that our perceptual processes would combine co-occurring audio and visual evidence. This phenomenon, known as "coactivation," occurs when processes pool separate sources of information toward making a single decision. In contrast, perceptual processes may treat the audio and visual signals independently. One might assume that if people are faster to detect redundant audio-visual stimuli than either single modality stimulus (audio-only or visual-only), then they must be using a coactive process. However, this pattern of results could also follow from independent parallel perception of each modality. Miller (1982) proposed an inequality that could be used to distinguish between independent parallel and coactive processing. Using this inequality, he found group level evidence for coactive processing of audio/visual stimuli. In this study we attempted to replicate Miller's (1982) study in both experimental design and analyses. We then extended the study by using a sophisticated modeling framework, SFT, to determine the underlying cognitive workload capacity. We argue that analysis of individual-level results is vital to establishing the representativeness of group-level findings. We conclude with how a more robust modeling framework provides a clearer description of the underlying components of cognition necessary for future study of more complex environments.

The Race Model Inequality

Redundant signals often lead to faster reaction times than when either stimulus is presented alone (e.g., Duncan, 1980; Kahneman, 1973). This is called the redundant signals effect. Raab (1962) demonstrated that an independent parallel, race model also predicts a redundant signals effect. The race model assumes that when two modalities are processed in parallel, whichever has the faster processing rate will be the modality used in the decision making process. In general, the minimum of two random variables tends to be smaller than either variable alone, so the decision time in a race model tends to be faster with multiple modalities than with any single modality. This is often referred to as statistical facilitation. Therefore, if people respond faster with redundant, cross-modal stimuli than they do with either modality in isolation, the speedup may be merely a product of statistical facilitation rather than cognitive facilitation.

To further distinguish between independent parallel processing and true speedups in perceptual processing, Miller (1982) derived an upper bound on the degree of speedup that can be accounted for by statistical facilitation. This bound is often referred to as the race model inequality (RMI). Whenever responses are faster than the bound, the race model can be rejected for cross-modal stimuli, and a coactive processing model is favored.[1]

[1]Unfortunately, there is not enough information in race model (or the capacity coefficient) to distinguish between separate decisions parallel models and coactive models without making strong assumptions such as context invariance. Another measure of Systems Factorial Technology, the Survivor Interaction Contrast, is able to make more specific conclusions about these competing models.

The RMI is derived from the following logic. First, the race model predicts that response times (RT) in redundant signal trials are determined by the fastest among a sample from each of the single stimulus processing time distributions. Using the inclusion-exclusion principle, this suggests that for all times (t):

$$\begin{aligned} &P(T_1 < t \text{ or } T_2 < t|S_1, S_2) \\ &= P(T_1 < t\,|S_1, S_2) + P(T_2 < t|S_1, S_2) \\ &- P(T_1 < t \text{ and } T_2 < t|S_1, S_2). \end{aligned} \quad (1)$$

Here we use T_1 and T_2 to indicate the time it takes to identify signals S_1 and S_2. The left side of Equation 1 is the cumulative probability density function (CDF) for reaction times on redundant signal trials. Assuming the speed of identifying signal 1 does not change depending on the status of signal 2, the first two terms on the right side can be estimated by the empirical CDFs of the single-signal trials. The final term is the probability that both stimuli have been detected with given time (t). Although the final term cannot be directly observed, it must be positive, and thus the race model inequality can be written as

$$\begin{aligned} &P(T_{12} < t\,|S_1, S_2) \\ &= P(T_1 < t\,|S_1) + P(T_2 < t|S_2). \end{aligned} \quad (2)$$

Inequality 2 is the RMI, an upper bound for the possible speedup in redundant signal trials when using a separate-decisions, parallel process. If responses to redundant signals are faster than this bound, the speedup is greater than that which can be attributed to statistical facilitation, and thus a separate-decisions, parallel model should be rejected. Traditionally, if the race model inequality is violated, it is concluded that the cross-modal stimuli are being processed coactively.

To test for coactive processing of audio-visual stimuli, Miller (1982) used four trial conditions: audio-only, visual-only, both audio and visual simultaneously, and no stimulus. Participants were found to violate the race model inequality for audio/visual stimuli using a group-level analysis (Miller, 1982). We replicated Miller's (1982) study, using the same methods and approaches to data analysis. We hypothesized that, like Miller (1982) and Gondan et al. (2005), we would find violations of the race model inequality. We extended the analyses using SFT (Townsend and Nozawa, 1995) to make more specific, individual-level conclusions about the perceptual process.

Capacity Coefficient

One operationalization of capacity is the degree to which the speed of processing changes as the number of processes changes. The capacity coefficient, one measure of SFT, is the ratio of the cumulative hazard function of response times to redundant signals, $H_{\mathrm{AV}}(t)$, relative to a baseline performance. The baseline is derived from an unlimited capacity, independent, parallel (UCIP) processing model. The capacity coefficient that applies to this experiment is

$$C_{\mathrm{OR}}(t) = \frac{H_{\mathrm{AV}}(t)}{H_{\mathrm{A}}(t)+H_{\mathrm{V}}(t)}. \quad (3)$$

The "OR" in Equation 3 refers to the "first-terminating" structure of the experimental task (the first modality to detect a signal is sufficient to make a response).

The capacity coefficient classifies processing into three different categories: limited, unlimited, and super. Limited capacity refers to a decrease in performance as the number of sources of information increases. Unlimited capacity refers to performance that remains consistent with the baseline performance even as more sources of information are added. Super capacity refers to an increase in performance as the number of sources of information increases. Townsend and Nozawa (1995) demonstrated that when the race model inequality is violated, $C_{\mathrm{OR}}(t) > 1$ for some t, i.e., there will be super capacity for at least some range of times.

Assumptions

One important assumption of both the race model inequality and the capacity coefficient is context invariance. Context invariance means that the time required to process any one of the channels is invariant of what is happening in the other channel(s) (Townsend & Wenger, 2004). This assumption implies that for all time (t), when instructed to respond only when S_1 is detected,

$$\mathrm{P}(\mathrm{RT} < t \mid S_1) = \mathrm{P}(\mathrm{RT} < t \mid S_1 \text{ and } S_2). \quad (4)$$

Equation 4 applies likewise when S_2 is the target signal. Context invariance implies that the response time distribution to a single target signal will not vary in the presence or absence of another, non-target signal. For example, if a person is instructed to only respond when an audio signal is presented, the response time distributions for audio-only and redundant signals trials will be equivalent under context invariance. Context invariance, however, is different than stochastic independence. Stochastic independence demands that the processing channels be truly independent of one another, exhibiting no channel correlation within redundant signal trials.

The race model inequality assumes context invariance but not stochastic independence. The capacity coefficient assumes both context invariance and stochastic independence. These assumptions serve as the foundation for the baseline performance of the two measures: RMI – context invariance and parallel processing; capacity coefficient – context invariance (unlimited capacity), stochastic independence, and parallel processing (UCIP model).

Experiment

The current experiment was aimed at replicating the task from Miller (1982) and repeating its analyses for comparison to the capacity coefficient. The experiment was a Go/No-Go detection task with two possible cues, an audio stimulus and a visual stimulus. The presence of either stimulus prompted a response. We hypothesized that we would find a violation of the race model inequality and group-wide observations of super workload capacity.

Methods

Participants In order to achieve a sample size similar to that of Miller (1982), 119 students were recruited from an undergraduate psychology course at Wright State University and received class credit for their participation.

An additional twenty-six members of the Wright State community were recruited to participate with paid compensation in a second paid version of the experiment. The original motivation for conducting this second experiment was to compare Miller's (1982) analyses to additional SFT measures; however, for the scope of this paper we will only discuss results pertaining to the replication of Miller (1982) and the capacity coefficient. From here on, we refer to the participants who received class credit and the participants who received monetary compensation as the first group and the second group, respectively.

All participants had no previous training and were naive to the purpose of the study. All participants self-reported normal or corrected-to-normal vision and hearing.

Materials Stimuli were presented using PsychoPy (Peirce, 2009). Visual signals were presented on a 20" Sony Trinitron monitor. Participants wore Sennheiser headphones throughout the experiment to receive audio signals. Participants responded using a mouse.

Stimuli Audio signals were always a 780 Hz pure tone. Visual signals were always a white asterisk spanning 1.85 degrees of visual angle in the center of a mid-level gray screen. Stimuli presentation types were an exact replication to that of Miller (1982) as indicated in Table 1.

Table 1: Miller (1982) trial types.

	Visual	Ø
Audio	AV	ØA
Ø	VØ	ØØ

AV represents redundant audio-visual signal trials.
ØV represents visual-only trials.
AØ represents audio-only trials.
ØØ represents target absent trials.

Procedures Instructions were explained verbally as well as displayed on the computer screen. The instructions given were "Respond by clicking the mouse as quickly as possible if either the tone or asterisk is presented. Withhold response if neither signal appears."

At the onset of each trial a fixation cross was displayed in the center of the screen for 250 ms to direct the participant's attention to the start of the trial. After the offset of the fixation cross and a delay of 250 ms, one of the four trial types was presented. In redundant-target trials, the asterisk was displayed and the tone was played for 150 ms. In single-target audio (visual) trials, only the tone (asterisk) was presented. In target-absent trials, neither the tone nor the asterisk was presented. The participant was given 2 seconds from the onset of the target to respond by clicking the mouse or to withhold response. Trial duration was kept constant throughout the experiment for a total of 2.5 seconds per trial. If a response was withheld, the participant waited until a fixation cross was displayed at the start of the next trial. Trial order was randomized and consisted of 100 trials of each type, giving a total of 400 trials per participant.

Results

Of the first group of 119 total participants, 27 had lower than 90% overall accuracy and were not included in further analyses. For the remaining 92 participants, the average false alarm rate was 3.75%, miss rate for single target audio was 3.45%, miss rate for single target visual was 2.96%, and the miss rate for redundant targets was 1.98%. Mean correct response times were 491.9 ms (*SD* = 184) for audio-only, 352.7 ms (*SD* = 134) for visual-only, and 329.8 ms (*SD* = 118) for redundant targets (Figure 1). Using a Bayesian t-test (Morey & Rouder, 2013), we found strong evidence for a redundant-target advantage over both audio-alone ($BF = 1.11\text{x}10^{42}$) and visual-alone ($BF = 3.92\text{x}10^{13}$) conditions.

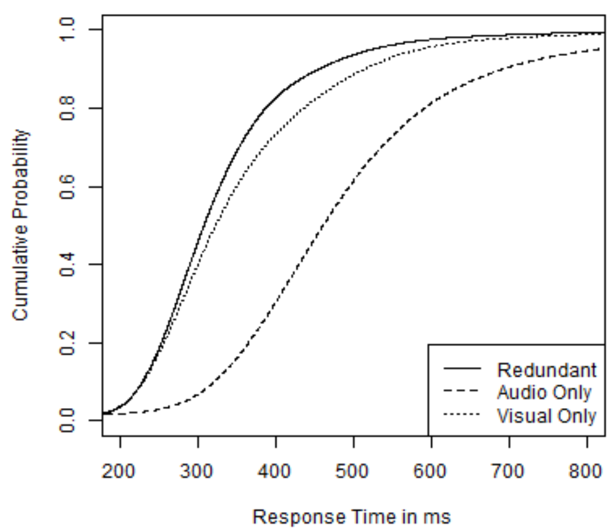

Figure 1: Group level redundant signal, audio-only, and visual-only cumulative distribution functions.

Following Miller (1982), we used the first 20 trials of each presentation type (Table 1) to test for differences in CDFs using t-tests for each quantile from 5% - 95% in 10% increments. These results are shown in Table 2. Note that we present both the standard *t*-test and the Bayes Factor *t*-test from Rouder, Speckman, Sun, Morey & Iverson (2009). The standard *t*-test is included for comparison to earlier results, although we focus our interpretation on the Bayesian analysis. For a comparison to current findings, Miller's (1982) results are reported in Table 2 indicating each quantile found to be significant as well as each quantile that was trending towards a violation of the race model (note that Miller (1982) did not report corresponding t-values).

Despite the clear evidence of a redundant signals effect at the group level, there was, at best, only marginal evidence of a group level violation of the race model inequality (Figure 2).

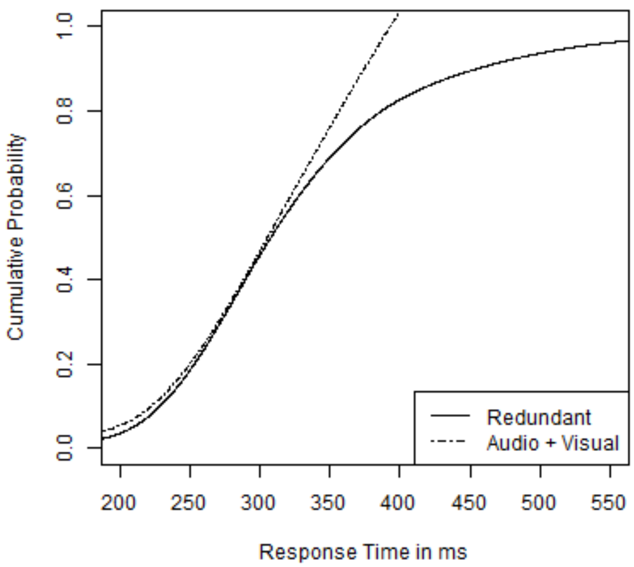

Figure 2: Group level redundant signals CDF and the corresponding RMI bound (audio alone + visual alone).

Table 2: Sequential t-test of the race model inequality.

Quantile	t	p-value	BF	Miller (1982)
5%	10.90	1.000	$< 1.0\times10^{-16}$	$p = .10$
15%	5.98	0.999	3.33×10^{-16}	$p < .05$
25%	3.11	0.994	1.60×10^{-3}	$p < .05$
35%	0.15	0.147	0.79	$p < .05$
45%	-0.38	0.088	1.81	$p = .10$
55%	-0.18	0.126	1.31	
65%	1.62	0.749	6.40×10^{-2}	
75%	3.59	0.999	3.28×10^{-4}	
85%	6.54	1.000	$< 1.0\times10^{-16}$	
95%	9.97	1.000	$< 1.0\times10^{-16}$	

Note. H_0: No violation of race model inequality.

To further explore the variations in performance across individuals, the capacity coefficient was calculated for each participant. Individual capacity functions are shown in Figure 3. Only 5 of the 92 participants were significantly super capacity, while 12 were significantly limited capacity.

A Bayesian t-test indicated substantial evidence that the group level capacity z-score would be zero (BF = 4.34), indicating unlimited capacity at the group-level. Despite the evidence against super capacity when the whole capacity function is taken into account, the average capacity function for the group (the thick black line in Figure 3) is above 1 for a small range of time.

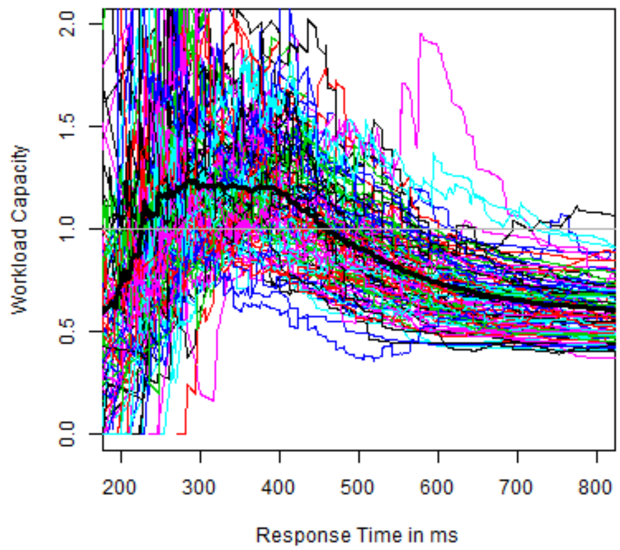

Figure 3: Capacity coefficients for each individual in Group 1 along with the group average capacity (bold black line).

Of the second group of participants, 12 remained after excluding those with low accuracy or, in the case of one subject, data corruption. The average false alarm rate across participants was 1.83%, miss rate for audio only was 1.67%, miss rate for visual only was 1.33%, and the miss rate for redundant targets was 1.17%. Mean correct response times were 471 ms (SD = 130) for audio only, 351 ms (SD = 111) for visual only, and 328 ms (SD = 82.2) for redundant targets, showing consistency with Group 1. There was decisive evidence for a redundant signals advantage over both audio alone ($BF = 2.71\times10^{215}$) and over visual alone ($BF = 1.17\times10^{8}$) in mean correct response times.

Again, following Miller (1982), we tested for differences in CDFs of the first 20 trials of each presentation type using t-tests for each quantile from 5% - 95% in 10% increments (Table 2). In these data, there was more evidence of a violation of the race model inequality. However, note that one should exercise caution when interpreting the Bayes factor in Table 3 because of the dependence among the t-tests.

Table 3: Sequential t-test of the race model inequality.

Quantile	t	p-value	BF	Miller (1982)
5%	2.36	0.981	0.04	$p = .10$
15%	0.78	0.775	0.34	$p < .05$
25%	-0.29	0.389	1.48	$p < .05$
35%	-1.70	0.058	10.70	$p < .05$
45%	-1.59	0.070	9.12	$p = .10$
55%	-1.24	0.121	5.49	
65%	-1.06	0.156	4.28	
75%	-0.40	0.350	1.71	
85%	1.29	0.888	0.17	
95%	3.35	0.997	8.26×10^{-3}	

Note. H_0: No violation of race model inequality.

Three of the twelve participants were significantly super capacity and one was significantly limited. A Bayesian t-test indicated evidence slightly favoring a group mean z-score of zero (*BF* = 2.27). The general shapes of the capacity coefficients in Figure 4 are similar to those in Figure 3. Again, there is a range over which the mean capacity function is positive.

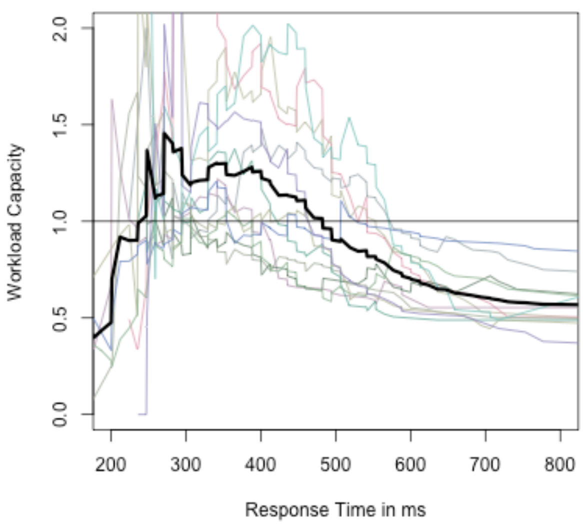

Figure 4: Capacity coefficients for each individual in Group 2 along with the group average capacity (bold black line).

General Discussion

By comparing individual and group CDFs we found many participants violated the race model inequality but only in small increments, not enough to be statistically significant. The few participants who did not violate the race model inequality were significantly slower than the race model bound. Hence, had it not been for these few participants, we may have replicated Miller's (1982) results and demonstrated a violation of the race model inequality. As discussed, with a group level analysis it is difficult to examine the influence of each participant in the overall findings. We examined both a larger sample (Group 1) and a smaller sample (Group 2) while still finding consistent group and individual level results. In both samples the majority of participants did no better or worse than an independent race model, with a few participants showing limited capacity and a few showing super capacity. It should be noted that participants across the two groups have slightly different miss and false alarm rates with paid participants (Group 2) having lower errors rates. In comparison to Miller (1982), both groups have substantially lower false alarm rates and higher miss rates. This indicates that participants were more biased toward responding than Miller's (1982) participants (5.66% false alarms). This bias is one possible explanation for the difference in results across the two studies.

Differences in RMI and Capacity Coefficient

Because the RMI is an upper bound, performance of limited capacity may satisfy the race model and conclude cognitive processing analogous to the baseline assumptions of the RMI. However, the Grice bound (Grice, Canham, & Boroughs, 1984) provides a lower bound on performance relative to the race model indicating an increase of response times, or decrement of performance, when more sources of information are added. For sake of simplicity and replication of Miller (1982) we did not include the Grice Inequality in the analyses of this paper (for more on Grice bound relative to SFT see Townsend and Eidels, 2011; Townsend and Wenger, 2004).

Being group level analyses, both the Miller and Grice inequalities are sensitive to individual variability. With this said, if cognitive processing varies among participants, multiple grouping effects may be disguised resulting in a weak or nonexistent violation. While there is at least one individual level test for violations of the race model inequality (Maris & Maris, 2003), that test requires a very particular experimental design, which deviates from Miller's (1982) original design and would conflict with our goal of replication. The capacity coefficient allowed for the replication of Miller's (1982) experimental design and supplies insight into individual workload variability among participants. A violation of the RMI indicating coactive processing with audio/visual information has been replicated since Miller's original paper (e.g., Gondon, 2005) yet may not be a particularly robust effect given our failure to replicate it. From a theoretical standpoint, it is imperative to analyze individual level performance when weaker evidence is found at the group level to determine whether unimodal (group level) effects or multimodal (subgroups) effects are responsible for the weaker group evidence. When a weak effect is observed, it may be the case that the majority is truly performing better than baseline but a few severely limited participants are dragging the group level observations down. To further advance our knowledge of the redundant signals phenomena we must study individual performance so as to adequately characterize group level interpretations.

Capacity Coefficient

The UCIP model that is used as a baseline in the capacity coefficient is more constrained than the general class of race models tested by the RMI, so evidence for unlimited capacity is evidence against a violation of the RMI. Indeed, Townsend & Nozawa (1995) demonstrated that if the RMI is violated, then the capacity coefficient must be larger than one for at least some range of times. The capacity statistic from Houpt & Townsend (2012) tests an aggregate value of capacity across time, so it does not directly test if the capacity value is *ever* different from one. In both Group 1 and Group 2, there was a trend toward a violation of the RMI, although it was not strong evidence in either case. The capacity coefficient plots in Figures 3 and 4 seem to have a similar indication: in both plots the mean capacity function and many of the individual functions are above one for at least some time. Nevertheless, despite five (or eight if the second group is counted) replications, we did not find conclusive evidence that participants were generally better than the UCIP model, let alone better than any race model.

Conclusions

We found evidence for the redundant signals effect that did not violate the race model inequality, i.e. evidence that could be explained by statistical facilitation, a result inconsistent with Miller (1982). That study found violations of the race model inequality in two separate experiments, while we did not find a single violation in multiple comparisons. We hypothesized that using SFT (Townsend & Nozawa, 1995; Houpt & Townsend, 2012) techniques would provide additional evidence for coactive processing as proposed by Miller (1982) and for super workload capacity. Instead we found considerable evidence for an unlimited workload capacity. We stress that cognitive processing of audio-visual signals varies across individuals and as such researchers should be wary of conclusions about cognitive workload that are based solely on group analyses.

Acknowledgments

This work was supported by AFOSR Grant FA9550-13-1-0087 awarded to JWH.

References

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87,* 272-300.

Ernst, M., & Bülthoff (2004). Merging the senses into a robust percept. *TRENDS in Cognitive Sciences*, *8(4).*

Gondan, M., Niderhaus, B., Rosler, F., Roder, B. (2005). Multisensory processing in the redundant-target effect: A behavioral and event-related potential study. *Perception and Psychophysics*, *67(4),* 713-726.

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., Townsend, J. T. (2013). Systems Factorial Technology with R. Behavior *Research Methods.* Advance online publication. doi: 10.3758/s13428-013-0377-3

Houpt, J. W. & Townsend, J.T. (2012). Statistical measures for workload capacity analysis. *Journal of Mathematical Psychology, 56,* 341-355.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Maris, G., & Maris, E. (2003). Testing the race model inequality: A nonparametric approach. *Journal of Mathematical Psychology*, *47,* 507-514.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive psychology*, *14(2),* 247-279.

Nickerson, R. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement. *Psychological Review, 80,* 489-509.

Otto, T., & Mamassian, P. (2012). Noise and Correlations in Parallel Perceptual Decision Making. *Current Biology*, *22,* 1391-1396.

Peirce, JW. (2009). Generating stimuli for neuroscience using PsychoPy. Front. *Neuroinform*. 2:10. doi:10.3389/neuro.11.010.2008

Raab, D. Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences,* 1962, 24, 574-590.

Rouder, J., Speckman, P., Sun, D., Morey, R., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review, 16*, pp. 225-237.

Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual changes result from bimodal integration. *Psychophysiology*, *35*, 755-759.

Smith, E.E. (1998) Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin, 69*, 77-110.

Townsend, J. (1990). Truth and consequences of ordinal differences in statistical distributions: toward a theory of hierarchical inference. *Psychological Bulletin, 108,* 3, 551-567.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology, 39*, 321-360.

Townsend, J.T., & Wenger. (2004). A Theory of Interactive Parallel Processing: New Capacity Measures and Predictions for a Response Time Inequality Series. *Psychological Reivew,111,* (1003-1035).

Watson, A. & Pelli, D. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33(2)*, 113-120