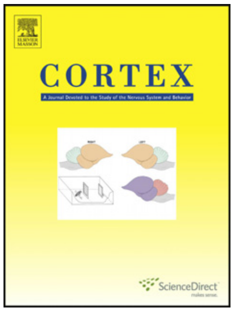
Accepted Manuscript



Title: Picture-word interference and the response-exclusion hypothesis: A response to Mulatti and Coltheart

Authors: Bradford Z. Mahon, Frank E. Garcea, Eduardo Navarrete

PII: S0010-9452(11)00282-6
DOI: 10.1016/j.cortex.2011.10.008
Reference: CORTEX 729

To appear in: *CORTEX*

Received Date: 2 August 2011
Revised Date: 18 October 2011
Accepted Date: 20 October 2011

Picture-word interference and the response-exclusion hypothesis:

A response to Mulatti and Coltheart

Bradford Z. Mahon[1,2], Frank E. Garcea[1], Eduardo Navarrete[3]

[1]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY
[2]Department of Neurosurgery, University of Rochester, Rochester, NY
[3] DPSS – University of Padova, Italia

Address for Correspondence
Bradford Z. Mahon
Meliora Hall
University of Rochester
Rochester, NY. 14627-0268
mahon@rcbi.rochester.edu

Abstract

Mulatti and Coltheart (2011, this issue) review and summarize several findings from the picture-word interference paradigm that the authors argue challenge the Response Exclusion Hypothesis. However, the hypothesis they take to be the Response Exclusion Hypothesis is not that theory—it is an account developed by Mulatti and Coltheart that holds that target naming latencies in the picture-word paradigm are affected only by the process of excluding the distractor word (and by nothing else). We consider some of the background assumptions implicit in Mulatti and Coltheart's discussion that may have led to this misattribution. Finally, we report a replication of an effect originally described by Dalrymple-Alford (1972) that serves as an empirical basis for reiterating the main points of our proposal and outlining the challenges that lie ahead.

In order for a manipulation of the speed of processing of the distractor (e.g. distractor frequency) to affect the time the response to the target accesses the buffer, it is necessary that the response to the target waits for the buffer to be purged. But this has an unwanted implication: If the response to the target has to wait for the distractor to be purged from the buffer, the benefit that derives from the picture name being of high frequency with respect to being low frequency is cancelled. … An example may illustrate the point.

Suppose that Leonardo (low frequency picture) and Michelangelo (high frequency picture) want to talk to Giotto (the buffer), who is in the Scrovegni Chapel. Leonardo and Michelangelo start from the same point to go to Giotto, but Michelangelo can walk much faster than Leonardo, and gets to the Chapel five minutes before Leonardo. If Giotto is immediately available, then Michelangelo will talk to Giotto 5 minutes before Leonardo. But if Giotto is busy, for example he is painting (that is, the buffer is occupied), then Michelangelo will have to wait, and the advantage of being able to walk much faster than Leonardo vanishes.

(Mulatti and Coltheart, 2011, this issue, p XX)

Imagine the scenario in which Michelangelo called Giotto on his cell while passing through the Prato della Valle, and said, "Hey Giotto, I'm coming to see you so finish up what you are doing and don't keep me waiting for too long." And then Leonardo, on his walk, made the same call to say the same thing from the same place. This could result in the situation in which the difference in speed of arrival between Michelangelo (high frequency target) and Leonardo (low frequency target) was not swallowed up by Giotto's work (distractor exclusion), while at the same time, Giotto might make his visitors wait a bit (distractor exclusion slows down target processing).

Another possibility: Giotto does not receive any advance warning from his visitors, but he is the type who can work only when under pressure. So only when he knows that there is a visitor waiting for him does he start his work (the process of excluding the distractor only begins when a target representation has arrived at the buffer). If that were the case, then with respect to the time that he started walking, Michelangelo (high frequency target) would get to speak to Giotto five minutes sooner than would Leonardo (low frequency target).

There are a number of assumptions that Mulatti and Coltheart's argument does not acknowledge but which must be specified. For instance, it is critical to know how long Giotto's work takes relative to the time it takes Michelangelo and Leonardo to walk to him—How long does it take to exclude distractors? It is also important to know whether Giotto starts his work while his visitors are walking, or waits to start his work—can distractors, in principle, be excluded before the target arrives at the buffer? It is also critical to know whether Giotto communicates anything to his visitors about whether he is busy or not, and whether his visitors communicate anything to Giotto about their imminent arrival – is there any exchange of information about the state of lexicalization and distrator exclusion? Does such (potential) information exchange affect processing time?

These types of considerations mean that, at best, the gloomy forcast for the Response Exclusion Hypothesis fortold by Mulatti and Coltheart (2011, this issue) does not follow with the force of logic. This is because the authors' arguments are built on premises of their own invention—not the assumptions that compose the Response Exclusion Hypothesis (Janssen et al., 2008; Mahon et al., 2007; Miozzo and Caramazza, 2003). Once it is assumed, as do Mulatti and Coltheart, that the time to produce the picture name depends *only* on the time it takes the system to exclude the distractor word, *and nothing else*, then it follows (by definitional fiat) that there will be no effect on naming latencies of any variables known to affect target retrieval. We know that hypothesis, outlined for the first time by Mulatti and Coltheart, is falsified by the many robust demonstrations of variables affecting target picture naming latencies (e.g., the picture frequency effect). Because Mulatti and Coltheart contend that their hypothesis is 'entailed' by the Response Exclusion Hypothesis they are led into the logical error of concluding that rejecting their hypothesis implies rejection of the Response Exclusion Hypothesis. Two background commitments seem to drive the authors' approach in thinking about these issues.

The first background assumption behind Mulatti and Coltheart's argument is that if two cognitive processes are not structurally serial, then they will run in parallel. In other words, anything that can, on structural or logical grounds, run in parallel, will run in parallel. This is too strong of an assumption. Clearing the single channel output buffer of the distractor and processing the target word up until that channel may not happen (entirely, or partly) in parallel, even though there may be no structural constraint that says they cannot happen in parallel. The reason why excluding the distractor may not (by hypothesis) take place during lexicalization of the target could have to do with other (non structural) reasons. For instance, it may be that endogenously allocated attentional resources are required for both processes, and those resources cannot be divided. Or, for instance, it could be that the process of verifying whether a given representation at the output level is the target can operate over only one word at a time, and only happens over the distractor when it becomes necessary to do so—i.e., when the target has 'arrived.'

A second background assumption made by Mulatti and Coltheart (2011, this issue) that seems too strong is that cognitive processes operate on the basis of only information that is logically necessary, or otherwise constitutive to their operation. Consider for instance the observation that motor processes are engaged by perceptual tasks, and that perceptual judgments can be influenced by putatively motor-based processes (cf embodied cognition, for reviews and discussion see Chatterjee, 2010; Hickok, 2009; Machery, 2007; Mahon and Caramazza, 2008). Or consider, for instance, that decisions about phonological similarity among auditorily presented words can be affected by the orthographic similarity of the words (Tanenhaus et al., 1980). Those types of findings suggest a different approach for thinking about how to use the picture-word paradigm to study language production—one in which information is promiscuously exchanged among distinct systems. This leads to a different class of questions then emphasized by Mulatti and Coltheart: How is information about distractors 'bundled' at the level at which they are excluded as potential responses? Does the information about the provenance of a representation at that level have to be 'looked up' or 'indexed'? If so, then do factors that affect the 'speed of rejecting distractors' exert their influence on the actual process of exclusion or rather on the 'looking up' of the relevant information? For instance, in the case of the distractor frequency effect, Miozzo and Caramazza (2003) interpreted the effect as having to do with how fast distractors are available for exclusion. Could it also be due, in whole or in part, to how fast the relevant information about the distractor (provenance, category, etc) can be 'looked up' or otherwise accessed?

Mulatti and Coltheart's review of the picture-word literature is a wonderfully concise summary of findings; that their conclusions regarding the Response Exclusion Hypothesis do not stand is because their starting premises are only a rough caricature of that theory. To their credit, the authors predicate much of their argument on *ceteris paribus* provisos – and it might be that by 'ceteris paribus' they mean to imply a certain set of background assumptions about how the mind (ceteris paribus) *must* work. But it is in no way obvious how issues for instance about parallel versus serial action among 'logically' independent processes match up with expectations of what should be the case, *ceteris paribus*. Thus, while the meta-analytic component of Mulatti and Coltheart's article clearly makes an important contribution in and of itself, revision of their initial assumptions could generate conclusions relevant to the Response Exclusion Hypothesis. As the authors frame the issues, however, the empirical findings that they summarize do not discredit or even embarrass the Response Exclusion Hypothesis. If there were to be any theoretical tension induced by those data, then we would have to explore the available hypothesis space and design the right experiments to tease apart the alternatives.

Taking a Step Back

Over and above different ways of fleshing out the Response Exclusion Hypothesis, what are the broader issues at stake? As Mulatti and Coltheart point out, the 'REH is not a theory of lexical access per se …[but rather an] …account of performance in the PWI task.' The theory of lexical access within which the REH was situated was one that dispensed with the need for assuming lexical selection by competition, and instead assumed that the most highly activated word is produced (without regard to the levels of activation of other words). That theory of lexical access offers an explanation of the general pattern of facilitation that is observed in the picture-word interference paradigm when distractor words are in a semantic relationship to target pictures.[1] If lexical selection is facilitatory, then there are exciting new issues to explore.

For instance, it is exciting to think about how to integrate inferences derived from chronometric effects with inferences derived on the basis of errors (Oppenheim et al., 2010). On the account that we have outlined (Janssen et al., 2008; Mahon et al., 2007; Miozzo and Caramazza, 2003), there are only two places at which speech production can be slowed down – the semantic level at which a decision must be made about which concept to lexicalize, and the 'point of no return' at the response level. Lexicalization on this view is a ballistic process that can be sped up, but not slowed down. There is no temporal cost possible at the level of lexical selection—provided that the correct target word is initially selected as being the most highly

[1] IE, an integral part of our general account of behavior in the PWI task is that naming latenceis are modulated by variables that affect target picture retrieval (semantic facilitation). In fact, the motivation for pursuing the Response Exclusion Hypothesis in the first place was the prior inference that lexical retrieval is facilitatory. This makes it all the more puzzling why Mulatti and Coltheart assert that the Response Exclusion Hypothesis maintains that picture naming latencies are affected only by distractor exclusion.

activated. Sometimes, however, the wrong word is lexicalized, and an error is made, either internally and never overtly uttered, or actually spoken. For instance, brain damage can lead to selective impairments at the lexical level in which patients make overt semantic coordinate errors (e.g., Caramazza and Hillis, 1991), and such errors can also occur in spontaneous speech, albeit rarely (Dell and Reich, 1981). The existence of such errors are not in conflict with the idea that lexical selection is facilitatory – those errors arise (by hypothesis) because the wrong word was the most highly activated. This type of view represents a new way of thinking about how to model response time effects in speech production (e.g., Dell et al., 2008).

We see this direction of development as an important step toward bringing together two traditions that have existed somewhat side-by-side in pyscholinguistics throughout the 'reign' of the theory of lexical selection by competition. On the one hand, there have been models primarly designed to explain how the correct word (among the possible alternatives) is selected; that class of models has principally been evaluated in the context of error data. On the other hand, there has been a tradition of models designed to explain *when* the system produces the correct word, *given* a prior stipulation about which word will be selected at the lexical level. In other words, theories that modeled uncertainty as to *which* word would be the target were evaluated more or less independently of theories that modeled uncertainty about *when* the target word would be produced. If we are no longer compelled to model temporal uncertainty at the level of lexical selection, then we can turn our attention toward integrating these two classes of models.

An empirical example

In the classic Stroop task, participants are slower to name the ink color (e.g., "red") of a printed incongruent word (green) compared to a printed congruent word (red). Part of this response time effect is due to slowing down in the incongruent condition and part of this response time difference is presumably due to facilitation in the congruent condition. However, the above described congruent condition is ambiguous as to what level of processing the facilitation effect might be occurring (semantic, lexical, phonological). An overlooked finding (Dalrymple-Alford, 1972) resolves the issue by changing the printed words to fire and lawn. According to the model of selection by competition, fire will compete more for saying the word "red" than will lawn, and should thus lead to slower naming latencies. However, if lexical selection is facilitatory then participants would be faster to name the ink color 'red' when the printed word is fire than when it is lawn. The finding, originally reported by Darymple-Alford (1972), shows that naming latencies are faster with fire as the distractor than with lawn as the distractor. Glaser and Glaser (1989) replicated the effect, although did not test the zero Stimulus Onset Asychrony (SOA) condition. We have replicated the original experiment from Dalyrmple-Alford (1972) with our own materials and obtained the same pattern (see Figure 1 for a reproduction of Dalrymple-Alford's experiment and our replication, Table 1 for response times and error by condition, and Supplemental Online Materials for our experimental procedures).

The observation, in the original Stroop task, that naming latencies ("red") are slower with incongruent color word distractors (green) than congruent color word distractors (red) is naturally explained by the Response Exclusion Hypothesis: the prepotent response engendered by the distractor (green) must be excluded as not being the intended response to the ink color, and excluding the distractor incurs a cost because the distractor satisfies response relevant criteria (its is a color name in a color naming task). However, the observation that ink naming latencies are facilitated in the congruent non-color word condition (fire) compared to the incongruent non-color word condition (lawn) is not due to an effect of distractor exclusion: neither 'fire' nor 'lawn' satisfy relevant criteria of the response 'red' in an ink naming task.

Rather, the distractor 'fire' primes the response 'red' at the semantic level, but the distractor 'lawn' does not—thus, the phenomenon can be explained only if one dispenses with the idea of competitive lexical selection. This invites the prediction that under delayed naming conditions (Janssen et al., 2008), the difference between the distractors 'red' and 'green' in naming the ink color 'red' will be largely preserved, while the difference between the distractors 'fire' and 'lawn' will dissappear.

Where does this leave us?

We have argued that Mulatti and Coltheart's arguments do not have the dire consequences for the Response Exclusion Hypothesis that the authors forecast. That said, their discussion is well taken and adds new dimensions to what a theory should address.

As a field, we know that the 'old view' of what is happening in the picture-word paradigm does not work. Further assumptions can be made to build upon the old theory in the context of challenging results, such as restrictions on which words may enter into competition (Bloem and La Heij, 2003; but see Navarrete and Costa, 2005), or the stipulation that there needs to be a critical mass of activated nontarget words for those words to compete with the target (Abdel Rahman and Melinger, 2009; but see Mahon and Caramazza, 2009). Thus, we have new versions of the selection by competition hypothesis and an alternative account – the Response Exclusion Hypothesis— and a lot of exciting future experiments in between.

Supplemental Online Materials

Methods

Participants. 8 University of Rochester students (2 male; all right handed (Edinburgh handedness questionnaire, average handedness coefficient = 0.95) ranging from 20 to 29 years (*M* = 22.1 years, SD = 2.9 years) participated in the study in exchange for payment. They all had normal or corrected-to-normal vision, and gave written informed consent in accordance with University of Rochester participant review boards.

Materials. The stimuli were words denoting colors (e.g., "RED"), objects (e.g., "FIRE"), or strings of X's that corresponded to the number of characters in the word distractor conditions (e.g., "XXX" for "RED"). The words or characters were capitalized, presented in a white Arial font, and bathed in one of six colors (red, blue, green, purple, pink, yellow). The distractors were red, blue, green, purple, pink, and yellow for the color word distractor condition, and fire, ocean, dollar, eggplant, ham, and corn for the noncolor word distractor condition. The stimuli were 150x80 pixels so as to require that participants viewed the target color within a region also occupied by the distractor word (see Supplemental Figure 1 for schematic). For example, if a color word, object word, or string of X's had 7 characters, that stimulus would be smaller than a stimulus that had 4 characters.

Testing Apparatus. The experiment was run on a desktop computer with a monitor resolution of 1920x1080 pixels. Stimulus presentation and response latencies were controlled and measured with DMDX (Forster and Forster, 2003). Participants sat approximately 60 centimeters away from the computer screen. They named the color of the image presented to them into a microphone. Stimuli subtended ~ 3° of visual angle.

Design. The design of the experiment was a 3 (distractor type: color, object, or XXX) by 2 (congruency: congruent, incongruent) within-subjects design.[*] Each participant took part in 5 blocks of 180 trials (~10 minutes per block). Of the 180 trials, 90 were congruent trials and 90 were incongruent. Of the 90 congruent trials, 30 were color-congruent trials, 30 were object-congruent trials, and 30 were XXXX trials (same distribution for the incongruent trials). Each participant completed the design five times, for a total of 900 trials per participant, or 7,200 total trials.

Procedure. The experiment began with the experimenter explaining to the participant that on each trial they must name the color of the stimulus presented to them (ignoring the colors black and white). They were told that the only allowable responses were red, blue, green, purple, pink, and yellow; it was made explicitly clear that a word or a string of X's was to be presented within the stimulus, but to only identify the background color. The experiment then began. On each trial, a fixation cross appeared on the screen for a jittered amount of time (450, 550, 650, 750 ms). Next, the stimulus was presented centrally and stayed on the screen until a response was recorded or two seconds had passed.

Results

Only latencies from correct trials were analyzed (~2% of all trials were errors). Latencies from correct trials were cleaned by excluding RTs faster than 200 ms, and greater than or less than two standard deviations above and below the mean for each participant, calculated across conditions (of the remaining correct responses, ~5% of trials were removed). Mean latencies, standard deviations, and error rates for all trials are displayed in Table 1.

---

[*] The XXX condition was arbitrarily divided into 'congruent XXX' and 'incongruent XXX' following the same design and distribution as for the color word and non-color word distractor conditions. This factor was was collapsed to form a single XXX condition against which to measure distractor effects.

A 3X2 ANOVA was performed with the factors Condition (three levels: color, object, X-type) and Congruency (two levels: congruent, incongruent). There were main effects of condition (F (2, 14) = 18.51, MSE = 650.84, $p < .001$, η^2 = .73) and congruency (F (1, 7) = 51.99, MSE = 441.41, $p < .001$, η^2 = .88); furthermore, the interaction of the two factors was significant (F (2, 14) = 49.27, MSE = 275.04, $p < .001$, η^2 = .88).

Planned comparisons (t-tests, two-tailed) tested for effects of congruency, separately for color word and non-color word distractor conditions. For color word distractors, color incongruent stimuli were named slower than color congruent stimuli (t(7) = 7.78, $p < .001$; interference effects ranged from 36 to 158 ms, mean, 110 ms; SEM, 14 ms). For noncolor word distractors, incongruent noncolor word distractors slowed naming latencies compared to congruent trials (t(7) = 3.03, $p < .05$; interference effects ranged from -7 ms to 45 ms; mean, 19 ms; SEM, 6 ms). In addition, compared to the baseline of a string of X's, while there were no differences for the congruent color word condition (t < 1), there were interference effects for the congruent noncolor-word distractor condition (t(7) = 4.15, $p < .01$; interference effects ranged from 3 to 63 ms, mean, 27 ms; SEM, 6 ms), the incongruent color-word distractor condition (t(7) = 6.90, $p < .001$; interference effects ranged from 35 to 163 ms, mean, 109 ms; SEM, 16 ms), and the incongruent noncolor-word distractor condition (t(7) = 6.92, $p < .001$; interference effects ranged from 10 to 68 ms, mean, 46 ms, SEM, 7 ms).

Figure Caption

Figure 1. Naming latencies for Dalrymple-Alford's (1972) original study and our replication. The red bar represents the RT for the 'XXXX' distractor condition

Table

Table 1. Mean latency (in ms), standard deviation (SD in ms) and error rate (%) by factor and congruency in Experiment 1.

	Color			**Object**			**X-type**		
	Latency	SD	%	Latency	SD	%	Latency	SD	%
Congruent	621	94	0.01	650	93	0.01	623	87	0.01
Incongruent	732	121	0.05	669	90	0.01	-	-	-

References

Abdel Rahman R and Melinger A. Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes,* 24 (5): 713-734, 2009.

Bloem I and La Heij W. Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language,* 48 (3): 468-488, 2003.

Caramazza A and Hillis AE. Lexical organization of nouns and verbs in the brain. *Nature,* 349 (6312): 788-790, 1991.

Chatterjee A. Disembodying Cognition. *Language and Cognition,* 2 (1): 79-116, 2010.

Dalrymple-Alford EC. Associative facilitation and interference in the Stroop color–word tests. *Perception and Psychophysics,* 11 (4): 274-276, 1972.

Dell GS, Oppenheim GM, and Kittredge AK. Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes,* 23 (4): 583-608, 2008.

Dell GS and Reich PA. Stages in Sentence Production - an Analysis of Speech Error Data. *Journal of Verbal Learning and Verbal Behavior,* 20 (6): 611-629, 1981.

Glaser WR and Glaser MO. Context effects on Stroop-like word and picture processing. *Journal of Experimental Psychology: General,* 118 (1): 13-42, 1989.

Hickok G. Eight Problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience,* 21 (7):1229-1243, 2009.

Janssen N, Schirm W, Mahon BZ, and Caramazza A. Semantic interference in a delayed naming task: evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 34 (1): 249-256, 2008.

Machery E. Concept empiricism: A methodological critique. *Cognition,* 104 (1): 19-46, 2007.

Mahon BZ and Caramazza A. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris,* 102 (1-3): 59-70, 2008.

Mahon BZ and Caramazza A. Why does lexical selection have to be so hard? Comment on Abdel Rahman and Melinger's Swinging Lexical Network Proposal. *Language and Cognitive Processes,* 24 (5): 735-748, 2009.

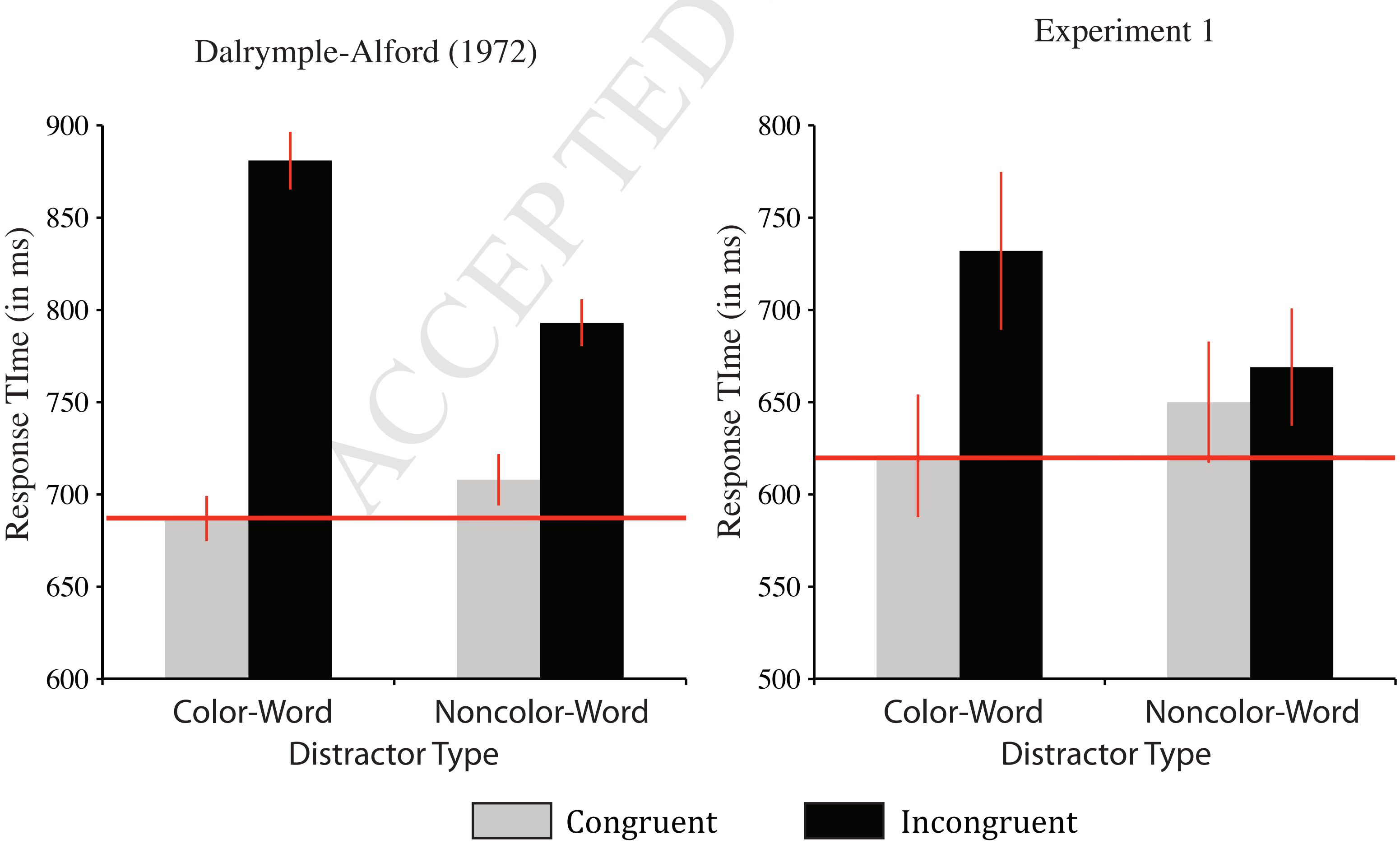
Mahon BZ, Costa A, Peterson R, Vargas KA, and Caramazza A. Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33 (3): 503-535, 2007.

Navarrete E. and Costa A. Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, 53 (3): 359-377, 2005.

Oppenheim GM, Dell GS, and Schwartz MF. The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition,* 114 (2): 227-252, 2010.

Tanenhaus MK, Flanigan HP, and Seidenberg MS. Orthographic and phonological activation in auditory and visual word recognition. *Memory and Cognition,* 8 (6): 513-520, 1980.

Figure 1. Response Time Effects By Experimental Condition



Supplemental Figure 1