

NEURAL CORRELATES OF SPEED-ACCURACY TRADEOFF:
AN ELECTROPHYSIOLOGICAL ANALYSIS

A Dissertation
Presented to
The Academic Faculty

by

Richard P. Heitz

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy

Georgia Institute of Technology

May, 2007

Copyright 2007 by Richard P. Heitz

NEURAL CORRELATES OF SPEED-ACCURACY TRADEOFF:
AN ELECTROPHYSIOLOGICAL ANALYSIS

Approved by:

Dr. Randall W. Engle, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Paul M. Corballis
School of Psychology
Georgia Institute of Technology

Dr. Daniel H. Spieler
School of Psychology
Georgia Institute of Technology

Dr. Eric Schumacher
School of Psychology
Georgia Institute of Technology

Dr. David Washburn
Department of Psychology
Georgia State University

Date Approved: 3/26/2007

ACKNOWLEDGEMENTS

I want to thank all of those people who have supported me throughout my graduate career. My dad Philip and my mom Norma, for their unending encouragement and support. I am here today *only* because you made it possible. My brother Rob and sister Margie for always being there for me. My grandfather, George Heitz for instilling in me the value of education. My uncle Mark Trzaska, for blazing this path many years before I and encouraging me throughout the way. Nash Unsworth, for the countless conversations we've had and the positive effect it's had on my work. To my labmates and good friends Tom Redick, Jim Broadway, and Maggie Ilkowska for always lending an ear and helpful comments. Nate Parks and Matt Hilimire, for teaching me the art of ERP analysis and allowing me to call at all hours for help. Paul Corballis, for endless encouragement, the use of his lab and equipment, and for introducing me to the ideas that led to this dissertation. It has been a pleasure. To the rest of my committee, Dan Spieler, Eric Schumacher, and David Washburn, for being patient and helpful. Randy Engle, you have opened up countless doors – and often shoved me through them. Thanks for the push. I could not have asked for a better mentor.

I dedicate this work to my incredible fiancée Gail. You have given my life new meaning. Thank you for everything.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
SUMMARY	ix
CHAPTER 1 – Introduction	1
The Urn Game	2
Criterion	2
Prior Probability	4
Rate of Information Accumulation	5
Sequential-Sampling Models	6
Evidence for the Neural Implementation of Accumulator Models	10
The Lateralized Readiness Potential	17
CHAPTER 2 – Experiment 1	24
Method	24
EEG and EMG Recording	24
Eriksen Flanker Task	25
Results	28
Behavioral Data	28
Latency Distributions	30
Conditional Accuracy Functions	34
LATER Model	38
Electrophysiological Data	44
EMG Onset Detection	44
Compatible and Incompatible Correct R-LRPs	46
Compatible and Incompatible Median Splits	48
Incorrect Responses	55
Missed Deadlines	57
Discussion	62
Baseline versus Criterion	65
CHAPTER 3 – Experiment 2	69
Method	69
Flanker Task	69
Results	69
Behavioral Data	69
Latency Distributions	71
Conditional Accuracy Functions	72

LATER Model	77
Electrophysiological Data	78
EMG Onset Detection	81
Compatible and Incompatible Correct R-LRPs	82
Compatible and Incompatible Median Splits	83
Incorrect Responses	86
Missed Deadlines	87
Discussion	90
CHAPTER 4 – General Discussion	91
Main Thesis	91
Errors and Missed Deadlines	94
Conclusion	95
APPENDIX A	96
Experiment 1	96
Experiment 2	101
FOOTNOTES	106
REFERENCES	107

LIST OF TABLES

	Page
Table 1: Mean RT and Accuracy Rate by Compatibility and SAT	29
Table 2: % Missed Deadlines and Median RT for Made and Missed Trials	33
Table 3: Median EMG versus RT	46
Table 4: Mean RT and Accuracy Rate by Compatibility and SAT	70
Table 5: % Missed Deadlines and Median RT for Made and Missed Trials	72
Table 6: Median EMG versus RT	81
Table A1: Subject-level RT and Accuracy rate for Experiment 1 (Compatible)	96
Table A2: Subject-level RT and Accuracy rate for Experiment 1 (Incompatible)	96
Table A3: Subject-level RT and Accuracy rate for Experiment 2 (Compatible)	101
Table A4: Subject-level RT and Accuracy rate for Experiment 2 (Incompatible)	101

LIST OF FIGURES

	Page
Figure 1: Effect of criterion variance on the decision process.	3
Figure 2: Effect of prior probability on the decision process.	4
Figure 3: Effect of rate of accumulation on the decision process.	5
Figure 4: Idealized random walk/diffusion model.	7
Figure 5: Recording from area MT using the dot motion task.	11
Figure 6: Neural decision making model of Shadlen et al. (1996)	12
Figure 7: Firing rates of LIP neurons during the dot motion task.	15
Figure 8: Derivation of the LRP.	20
Figure 9: Compatible correct and incorrect latency distributions.	31
Figure 10: Incompatible correct and incorrect latency distributions.	32
Figure 11: Compatible CAFs.	35
Figure 12: Incompatible CAFs.	36
Figure 13: The Reciprobit plot.	40
Figure 14: Compatible Reciprobit Plots.	41
Figure 15: Incompatible Reciprobit Plot.	42
Figure 16: Typical single-trial EMG.	45
Figure 17: Grand-average R-LRP for compatible correct trials.	47
Figure 18: Grand-average R-LRP for incompatible correct trials.	47
Figure 19: Grand-average R-LRP waveforms for median split (compatible).	50
Figure 20: Grand-average R-LRP waveforms for median split (incompatible).	51
Figure 21: Grand-average R-LRP considering only slow half of median split.	53
Figure 22: Effect of two different baseline periods on the R-LRP.	55

Figure 23: Grand-average R-LRP for correct and incorrect trials (compatible).	56
Figure 24: Grand-average R-LRP for correct and incorrect trials (incompatible).	57
Figure 25: Grand-average R-LRP for correct incompatible trials (slow condition).	58
Figure 26: Grand-average R-LRP for correct incompatible trials (med condition).	59
Figure 27: Grand-average R-LRP for correct incompatible trials (fast condition).	59
Figure 28: Grand-average S-LRP for made and missed deadlines.	61
Figure 29: Compatible correct and incorrect latency distributions.	73
Figure 30: Incompatible correct and incorrect latency distributions.	74
Figure 31: Compatible CAFs.	75
Figure 32: Incompatible CAFs.	76
Figure 33: Compatible Reciprobit plots.	79
Figure 34: Incompatible Reciprobit plots.	80
Figure 35: Grand-average R-LRP for compatible correct trials.	81
Figure 36: Grand-average R-LRP for incompatible correct trials.	82
Figure 37: Grand-average R-LRP waveforms for median split (compatible).	84
Figure 38: Grand-average R-LRP waveforms for median split (incompatible).	85
Figure 39: Grand-average R-LRP for correct and incorrect trials (compatible).	86
Figure 40: Grand-average R-LRP for correct and incorrect trials (incompatible).	87
Figure 41: Grand-average R-LRP for made and missed deadlines (slow condition).	88
Figure 42: Grand-average R-LRP for made and missed deadlines (med condition).	88
Figure 43: Grand-average R-LRP for made and missed deadlines (fast condition).	88
Figure 44: Grand-average S-LRP for made and missed deadlines.	89
Figure A1: Subject-level LATER fits (compatible).	97
Figure A2: Subject-level LATER fits (incompatible).	98
Figure A3: Subject-level R-LRPs (compatible).	99

Figure A4: Subject-level R-LRPs (incompatible).	100
Figure A5: Subject-level LATER fits (compatible).	102
Figure A6: Subject-level LATER fits (incompatible).	103
Figure A7: Subject-level R-LRPs (compatible).	104
Figure A8: Subject-level R-LRPs (incompatible).	105

SUMMARY

Recent computational models and physiological studies suggest that simple, two-alternative forced-choice decision making can be conceptualized as the gradual accumulation of sensory evidence. Accordingly, information is sampled over time from a sensory stimulus, giving rise to an “activation function.” A response is emitted when this function reaches a criterion level of activity. Critically, the phenomenon known as “speed-accuracy tradeoff” (SAT) is modeled as a shift in the response boundaries (criterion). As speed stress increases and criterion is lowered, the information function travels less distance before reaching threshold. This leads to faster overall responses, but also an increase in error rate, given that less information is accumulated.

Psychophysiological data using EEG and single-unit recordings from monkey cortex suggest that these “accumulator” models are biologically plausible. The present work is an effort to strengthen this position. Specifically, it seeks to demonstrate a neural correlate of criterion and demonstrate its relationship to behavior. To do so, subjects performed a letter discrimination paradigm under three levels of speed stress. At the same time, electroencephalogram (EEG) was used to derive a measure known as the lateralized readiness potential, which is known to reflect ongoing motor preparation in motor cortex. In Experiment 1, the amplitude of the LRP was related to speed stress: as subjects were forced to respond more quickly, less information was accumulated before making a response. In other words, criterion lowered. These data are complicated by Experiment 2, which found that there are boundary conditions for this effect to obtain.

CHAPTER 1

INTRODUCTION

It is a matter of common knowledge that one's ability to perform a task is hindered by the need to perform it quickly. Essentially, errors are more likely to result in nearly any task when one is in some sense "rushed," compared with performance in an unrushed task. The converse holds as well: when a task must be performed with high accuracy, more time will pass between start and finish. In scientific circles, this phenomenon is known as the speed-accuracy tradeoff (SAT), and it forms an important part of modern decision making theory. Specifically, any complete model of decision making must include a mechanism with which to implement the SAT. To date, the most successful class of models meeting this demand are based on a sequential-sampling process in two-alternative forced choice tasks (2AFC). Accordingly, information is thought to be continually sampled, over time, from a stimulus. This information is accumulated, giving rise to an "activation function" that represents the momentary amount of evidence for option "A" over option "B". A response is produced when this activation function reaches a boundary ("criterion"). As will be shown, movement of the criterion imposes the SAT and accounts for associated changes in accuracy rate and RT (Ratcliff, 1978; Ratcliff & Rouder, 1998). Despite their now widespread use, and their ability to account for a wide range of behavior, it remains to be shown that such models are a plausible depiction of the way decisions are produced in the brain. Following up on pioneering work making this case (e.g., Smith & Ratcliff, 2004; Hanes & Schall, 1996; Shadlen & Newsome, 2001), the present work will explore the mechanism through which the SAT is implemented and the biological plausibility of such a mechanism. The

following section introduces the basics of the sequential-sampling process in the context of a simple game.

The Urn Game

Imagine that you are presented with two urns, each containing some number of quarters and dimes. Urn Q contains mostly quarters, but also some dimes. Urn D contains mostly dimes, but also some quarters. On each “trial,” you are presented with only one urn – you do not know which urn it is, nor can you see inside of it – and you must decide whether it is urn Q or D . To help you decide, you are allowed to sample, one-by-one, coins from each urn. For present purposes, assume that coins are sampled at a fixed rate. You are not told how many items to sample, but two critical facts are understood. First, if you are correct in your decision, you are allowed to keep all of the money in the urn. Thus, it is in your best interest to maintain high accuracy. Second, because sampled coins are thrown away, the fewer the samples, the more money left inside. For this reason, it is also in your best interest to arrive at a decision quickly, sampling as little as possible consonant with an acceptable accuracy rate. Three important factors will influence what decision you make and how long it takes to arrive at that decision.

Criterion

Each time you draw a sample, you accumulate some evidence in favor of one of the two urns. Obviously, you would not want to continue sampling without end; rather, you must establish what constitutes enough evidence to stop and make a decision. In the models considered here, this “criterion” or “threshold” level represents the relative amount of evidence for urn Q versus urn D that must be reached before emitting a

response. In other words, how many quarters *over* dimes must be sampled in order to make a Q decision? Once this criterion is reached, you discontinue sampling and emit a response.

The choice of a criterion level will have a direct impact on your winnings. Figure 1 depicts a simulated record of the accumulation process. The solid upper and lower boundaries represent a high criterion setting; the dashed lines a low setting. The red step-function indicates the relative evidence in favor of either of the urns, each step being another sample.

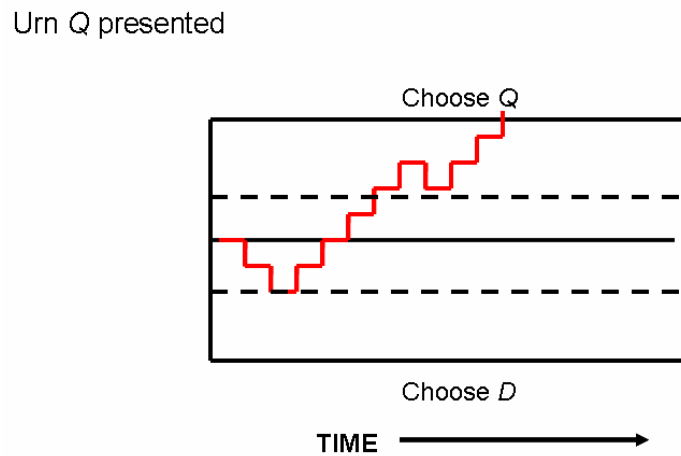


Figure 1. The effect of criterion variance on the decision process in the simple urn game. Note that with a low criterion (dashed horizontal), an incorrect “D” response would be produced at an early point in time. With a high criterion, a correct response is produced, but with a longer RT.

First consider the high criterion case. Because dimes and quarters are intermixed, the information function demonstrates periodic up and down movements. Slowly, the function drifts towards the Q boundary. At some point, a criterion level of evidence is

reached and a Q response is produced. Contrast this with a low criterion case. After only two samples, the D boundary is met, resulting in a very fast, but erroneous, response. In some situations, a low criterion is beneficial – particularly when speed is emphasized over accuracy. This is the basis of the SAT (Ratcliff, 1978; Wickelgren, 1977; Lohman, 1989).

Prior Probability

The base-rate probability of being presented with urn D or Q will also surely impact your decision, and how long you take to make it. In the previous example, it was assumed that urn D and urn Q are presented with equal frequency. For this reason, the information function in Figure 1 begins equidistant from the Q and D boundaries. This need not be the case. For instance, suppose you are told that urn Q is twice as likely to occur as urn D . In this instance, it would be wise to bias your decision process towards D even before the sampling process begins. The net result of this bias is to reach the Q criterion after fewer quarter samples (Figure 2).

Urn Q presented

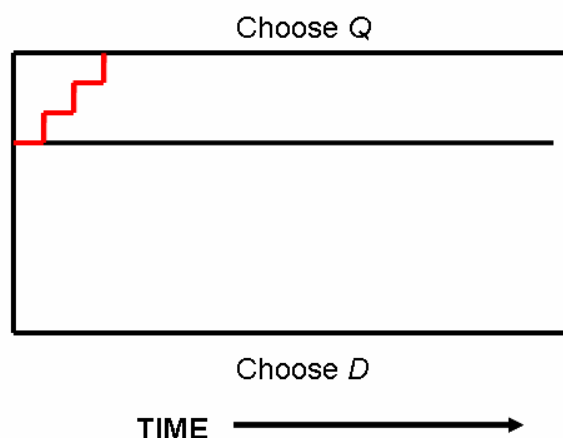


Figure 2. The effect of prior probability on the decision process in the simple urn game. When urn Q is twice as likely, the accumulation process begins “biased” towards the Q boundary, producing shorter RTs for that decision.

Note that on this trial, the Q boundary was reached after only 3 samples. Every so often, of course, this will lead to an error – a run of three quarters is not implausible even when urn D is presented. However, because urn Q is presented twice as often, these errors carry less weight. Thus, most of the time, this bias will speed correct decisions.

Rate of Information Accumulation

The rate of information accumulation does not refer to how quickly one samples – this was assumed to be some fixed value. Rather, it is akin to d' from signal detection theory – in this case, the proportion of dimes to quarters within each urn. Figure 3 depicts what the accumulation record might look like for a near-random mixture and a more unequal mixture.

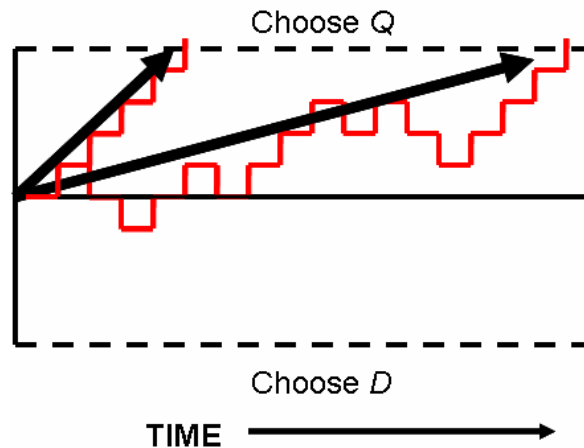


Figure 3. The effect of change in proportion of dimes to quarters on the rate of information accumulation. When there is a large difference, nearly every sample will move the function towards the correct boundary. When there is a more random mixture, it takes longer to reach criterion.

For the near-random mixture, it is evident that the information function requires many samples to reach the Q boundary. Contrast this with a situation where there is a large proportion of dimes to quarters in urn D : nearly every sample will move the information

function towards the correct boundary (Figure 3). Altering the proportion of dimes to quarters in the urns, then, affects how long, on average, it takes for the information function to reach criterion.

The rate at which the function approaches boundaries is conceptually similar to d' in signal detection theory. If we consider the D distribution to be noise and the Q to be signal, then d' indicates the overlap between the two distributions. When there is little overlap, samples from the noise distribution are easily distinguishable from samples from the signal distribution. When there is much overlap (as when dimes and quarters are randomly intermixed), it is difficult, if not impossible, to separate signal from noise¹.

Sequential-Sampling Models and the “Accumulator in the Brain”

The above vignette embodies modern theories of decision making known as “sequential-sampling” models. There are a variety of instantiations in this class of models, some of the more popular being the random-walk (Link & Heath, 1975), the diffusion (Ratcliff, 1978; Ratcliff & Rouder, 1998), the leaky, competing accumulator (Usher & McClelland, 2001), and the Linear Approach to Threshold with Ergodic Rate (“LATER”; Carpenter & Williams, 1995; Reddi & Carpenter, 2000). The success of these models can be traced to four essential properties held in common: 1) the accumulation of evidence, over time 2) some form of bias, manipulable by prior probability, 3) the idea of a drift rate, manipulable by quality of information, and 4) the idea of a criterion level of evidence and its relationship to the SAT. Depicted in Figure 4 is a generalized sequential sampling model. The reader should note that the model is virtually identical to that provided in the above vignette, albeit with different names for the parameters. And, obviously, the information being sampled is no longer quarters and

dimes, but evidence from some sensory event.

Sequential sampling models have been extensively tested and verified in a wide variety of research domains, task paradigms, and subject populations (Luce, 1986; Ratcliff & Smith, 2004). They have become intensely popular not only because they are extraordinarily accurate at predicting behavioral performance, but because they are parsimonious (a model with 3 free parameters can reproduce entire response time distributions for both correct and error trials as well as mean error rates; note the positively skewed RT distribution depicted in Figure 4).

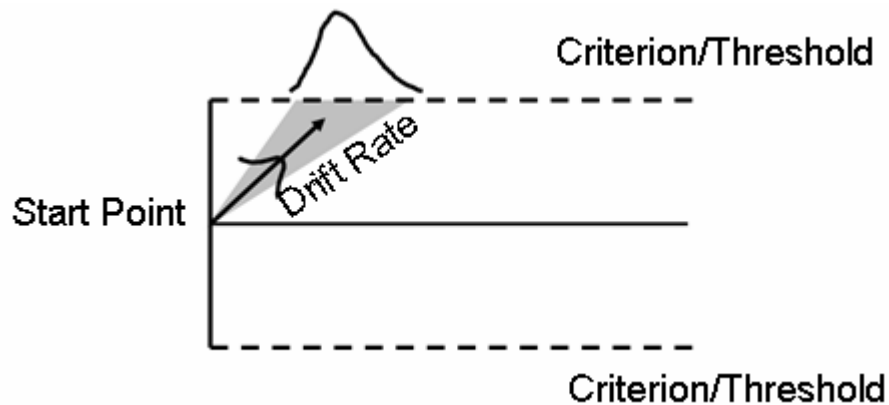


Figure 4. An idealized random walk/diffusion model. Note that the geometry of the decision process gives rise to the positively skewed RT distributions commonly observed.

Showing that a particular mathematical model can reproduce behavioral performance is impressive and certainly provides a framework for conceptualizing the decision process. Nevertheless, it is important to consider whether the models, as described, bear any relationship to the brain mechanisms responsible for decision making. If they do not, then the models are little more than tools of prediction. Although this may be sufficient in some cases, it does not fulfill the ultimate goals of cognitive

psychology and cognitive neuroscience – namely, to discover the cognitive mechanisms that govern behavior as well as understand how they are implemented in the brain.

Fortunately, recent neurophysiological evidence shows that we may not abandon sequential-sampling models in the near future. Quite to the contrary, research has shown that these same models can describe single-unit recordings from monkey cortex. What's more, the time-course of neural activity can be used to predict, with near certainty, the nature of the impending response as well as its latency. Though it should not be surprising that some form of neural activity predicts behavioral responses, what is impressive is the specific pattern of neural activity, its relationship to established behavioral models, and our ability to manipulate that pattern of data in a predictable manner. Together, this provides evidence that sequential-sampling models are biologically plausible (Smith & Ratcliff, 2004; Shadlen & Newsome, 2001; Roitman & Shadlen, 2002; Hanes & Schall, 1996). As will be detailed below, certain cortical neurons appear to exhibit a “ramping” of activity in the presence of stimulation, with the rate of ramping proportional to the quality, or strength, of the signal (Hanes & Schall, 1996). The reader will recall that this is similar to the drift rate parameter discussed previously. And, like the before mentioned mathematical models, it appears that cortical neurons are sensitive to the prior probability of encountering a stimulus; they become more or less active when the probability of encountering a preferred or nonpreferred stimulus, respectively, is high, even *before* any stimulus is presented (Platt & Glimcher, 1999; see also Basso & Wurtz, 1997, 1998; Dorris & Munoz, 1998). Finally, and most important for present purposes, neural activity reaches a fixed, criterion level of activity at the point an overt response is produced, *regardless* of the resulting RT (Hanes &

Schall, 1996; Roitman & Shadlen, 2002).

What has yet to be demonstrated, however, is that criterion – in the neural sense – can be manipulated. Again, sequential-sampling models suggest that it can, through manipulations of SAT instructions, payoffs, deadline procedures, and the like. More specifically, criterion should fall with speed emphasis and rise with accuracy emphasis. But, within any given SAT “setting” – say, a block of “accuracy” trials, criterion should remain constant. To date, research has confirmed only the latter – a fixed criterion. This has been observed not only in single-unit responses, but in neural populations, as evidenced through electroencephalographic recordings (EEG). In neural firing records, research finds that a response (say, an arm or eye movement) is produced at the precise moment certain neurons reach a specified firing rate, regardless of that movement’s latency. Using EEG, research finds that hemispheric lateralization during hand movements reaches a fixed value at the moment an overt movement is produced, again regardless of latency. Whether or not these fixed values can be modulated by changes in SAT setting is unknown, and is the main goal of the present work.

Finding a neural correlate of SAT setting is important in at least two regards. First, it would provide further evidence for the biological plausibility of accumulator models². Second, it would explain the SAT in mechanistic terms. Very often, SAT settings are stated in such nondescript terms as a “cognitive set for accuracy” or “cognitive set for speed.” The present work attempts to discover, quantitatively, a neural basis of these cognitive sets. To maintain proper context for this exploration, the following section reviews evidence for the neural implementation of accumulator models, focusing first on evidence from single unit recordings, followed by neural population

studies using EEG in human subjects as well as functional imaging with fMRI.

Evidence for the Neural Implementation of Accumulator Models

Any neural implementation of an accumulator model first needs a signal to accumulate. In the previous vignette, this signal was a record of sequential samples of quarters and dimes. In more reductionist terminology, this signal should code basic stimulus attributes necessary for task performance. Take, for instance, a simple motion discrimination task in which subjects must decide which direction a field of dots are moving. Each sample (for example, each screen refresh) might represent a binary decision as to whether motion was in a particular direction or not. Over a period of time, one collects this evidence and produces a response when some criterion level of evidence is reached. In a perceptual categorization task, the accumulated signal might code the extent to which a picture resembles, say, a house or a face. Based on the sequential-sampling models described above, this signal should vary in intensity based on the “quality” or strength of the evidence. For instance, a suprathreshold picture of a face would exemplify a strong signal, while a perithreshold stimulus, degraded and presented amidst noise exemplifies a weak signal.

Obviously, the signal that is accumulated must originate from areas of cortex that are specialized for the task at hand. One area of cortex that has proven a fruitful area of inquiry is middle temporal (MT), also known as V5. Cells in MT are motion sensitive and direction selective. Thus, it is possible to find neurons that fire vigorously when motion of a particular direction is presented in its receptive field (RF). Likewise, these same neurons often become suppressed in the presence of motion in another direction, often diametrically opposite to the preferred direction. Importantly, the activity of MT

neurons has a near linear relationship to the strength of the motion stimulus (Britten et al., 1993).

Using this well-defined cortical area, Newsome and colleagues have explored in detail how information provided by MT neurons might be used in the decision process. Their task, which will be referred to throughout this work as the “dot motion” task, presents subjects (in this case, monkeys) with a field of randomly moving dots (Figure 5). Embedded within this field is a coherent motion signal; the “strength” of the motion

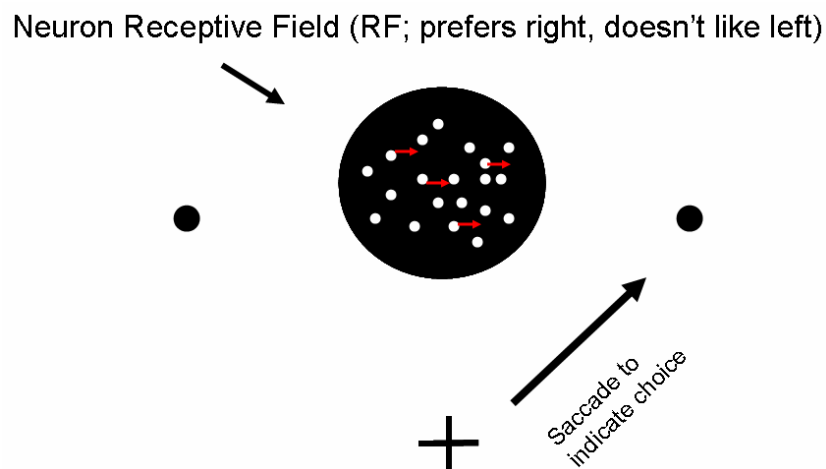


Figure 5. Recording from area MT using the dot motion task. Subjects view a dot stimulus, then, when the stimulus is removed, make a saccade to either of the targets corresponding to the perceived direction of motion.

signal is manipulated by altering the proportion of dots moving coherently. When the proportion of dots moving coherently is high, the motion signal is strong, the output from area MT is large, and psychometric performance is high. Likewise, as coherence is decreased, MT responses decrease (Britten et al., 1993) as well as psychometric performance (Newsome, Britten, & Movshon, 1989). Typically, monkeys respond to the perceived direction of motion with a saccade after some short presentation interval.

Explorations into this cortical area have revealed an exquisitely straightforward

model of the neural decision process. Single-unit experimental (Newsome, Britten, & Movshon, 1989) and modeling studies using actual MT data as input

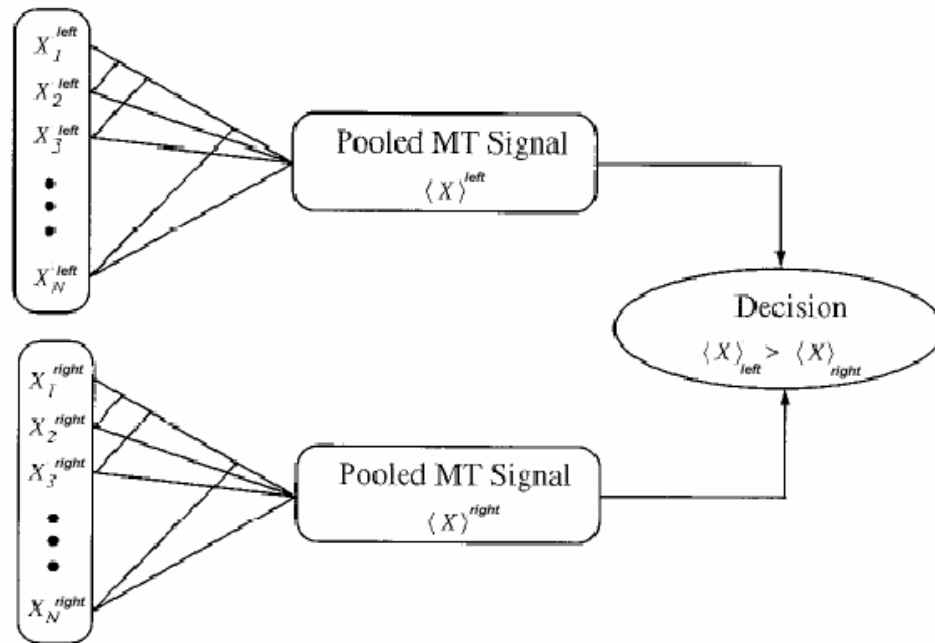


Figure 6. The neural decision making model of Shadlen et al., (1996). Pools of rightward and leftward preferring neurons are pooled, and then compared. The larger of the two dictates the decision.

(Shadlen et al. 1996) liken the decision process to a simple comparison between pools of directionally sensitive MT neurons (Figure 6). For example, imagine a pool of neurons whose RF is selective for rightward motion, and an opposing pool of neurons selective for leftward motion. At the beginning of a trial, all the neurons fire at base rate, and thus exhibit no discrimination. Upon onset the stimuli, however, one group fires significantly more than the other. The simplest possible decision rule would be to choose LEFT when the leftward-selective pool was most active and RIGHT when the rightward-selective pool was most active. Surprisingly, such a simple model (with ancillary assumptions) very accurately reproduces psychometric performance (Shadlen et al., 1996). This

implies that the output of MT neurons is somehow compared during the decision making process. Where this comparison takes place will be discussed later.

Further evidence that the output of two opposing pools of neurons participate in the decision process is the observation that microstimulation of leftward-selective neurons *slows* rightward decisions (Ditterich, Mazurek, & Shadlen, 2003). If only leftward-selective neurons participated in leftward decisions, then there is no reason why stimulation of *rightward*-selective neurons should at all affect the decision process. To summarize, these studies show that 2AFC decisions may be based on the compared output of 2 pools of selective neurons.

Although MT neurons have been shown to play a role in the decision process during the dot motion task, it is unlikely that the decision proper is actually computed there. Two observations make this clear. First, as was mentioned above, if the decision process compares the output of pools of directionally selective MT neurons, then some other area must carry out the comparison. Second, the inherent properties of MT neurons make them unlikely to be the site of higher level integration. Specifically, MT neurons respond only transiently. When stimulation is removed, the cells return to base rate. Yet, monkeys can easily perform a version of the dot motion task with an interpolated delay period. Hence, there must be some region that both accumulates signals from MT and maintains it across delays.

Previous research has suggested that areas such as prefrontal cortex (PFC) contain neurons that exhibit just such these qualities (e.g., Fuster & Alexander, 1971). And, certain areas of the PFC (frontal eye fields; FEF; see Tehovnik et al., 2000 for a review) as well as parietal cortex (lateral intraparietal area; LIP) are known to be innervated by

area MT and play a role in saccade generation (Schall, 2003). Thus, these sites are likely to play a role in the decision process with eye movements as effectors. Using the dot motion task, Kim and Shadlen (1999) examined the time-course of neural activity in the FEF. They found that FEF neurons exhibit properties needed by an accumulator of a decision variable. First, they maintain activity across delays. Second, their activity is time-dependent; unlike MT cells, their firing rate continually *grows* during motion viewing, as if they were accumulating evidence. And, the rate of this growth is related to the strength of the motion signal originating from area MT (Roitman & Shadlen, 2002; Figure 7, left panel). Third, responses are produced when activity reaches a fixed threshold (Roitman & Shadlen, 2002; Figure 7, right panel; see also Hanes & Schall, 1996). Finally, and quite importantly, these neurons code the observer's *decision* regardless of the true direction of motion. That is, neurons selective for rightward decisions become very active in the presence of leftward motion when an error is about to be made (Roitman & Shadlen, 2002). Kim and Shadlen (1999) rested with the conclusion that FEF neurons accumulate a *difference score* from directionally-tuned MT neurons. A response is produced when the accumulating difference score reaches a constant threshold. Computational models show that this is indeed plausible (Mazurek, Roitman, Ditterich, & Shadlen, 2003).

One further important property concerns the fact that this same type of activity is observed even on trials in which a response is *not* ultimately produced. For instance, Hanes & Schall (1996) recorded from FEF while subjects performed a stop-signal paradigm. On most trials, subjects were to make a saccade towards a peripheral target. These trials showed the typical ramping of activity towards a fixed criterion displayed in

Figure 7. On some trials, a *stop* signal would appear indicating that no response should

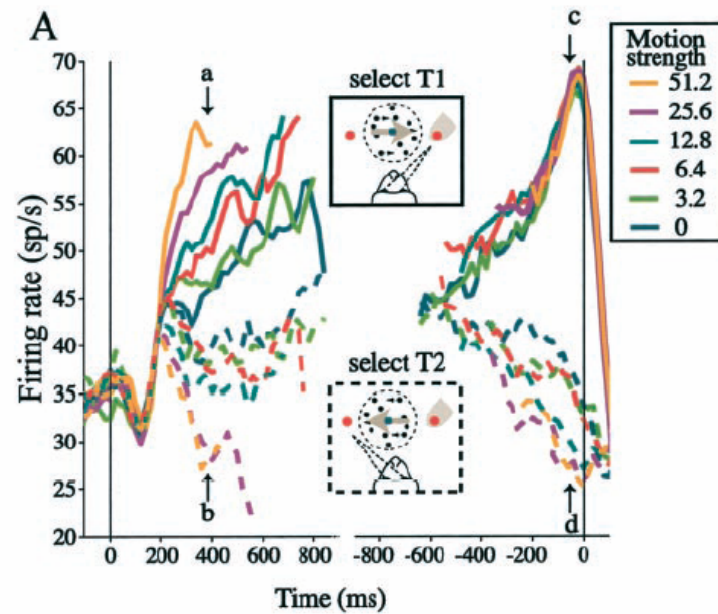


Figure 7. Stimulus-locked (left panel) and response locked (right panel) firing patterns for LIP neurons during the dot motion task. Solid lines depict activity for motion in the preferred direction; dotted lines activity for the non-preferred direction. Note that eye movements were produced at a stereotyped level of activity (right panel).

be made. On these signal trials, they observed that activity begins to ramp, but is inhibited a short time after the stop signal. After estimating response threshold on *go* trials, Hanes and Schall observed that neural activity never reached this level on successfully inhibited *stop* trials but did on non-inhibited *stop* trials. This observation leads to two important conclusions. First it proves that this neural activity is not simply movement-related, but sensitive to cognitive factors (Smith & Ratcliff, 2004). If the neurons were involved only in movement production, they should show no activity on successfully inhibited trials. Second, it provides a critical prediction regarding the SAT. Imagine that due to some SAT manipulation, criterion were lowered so as to stress speed over accuracy. On some *stop* trials, activity might now reach threshold levels before

inhibition set in, leading to an error.

The above work is restricted in the sense that recordings were taken from circumscribed areas of cortex – area MT, known to be motion sensitive and area FEF and LIP known to be important for saccade generation – while presenting a task that produces motion stimuli and takes a saccade as a response. If the accumulation of difference scores to threshold is to be regarded as a general mechanism through which decisions are produced, it is necessary to show that the same general principles hold in other tasks with other response modalities, as well as in other areas of cortex. This consists of a signal generated from specialized areas of sensory cortex (e.g., area MT, somatosensory cortex, etc.), and an area that accumulates the difference between pools of these specialized sensory neurons (e.g., area FEF, LIP, motor cortex, etc.). As well, these same general principles should be exhibited by other subject populations. To date, evidence supporting this is beginning to emerge. Neurons in areas of motor cortex, for example, are known to represent an evolving decision signal when arm or hand movements are used as effectors (Riehle & Requin, 1989; Zhang, Riehle, Requin, & Kornblum, 1997; Salinas & Romo, 1998; Hernandez, Zainos, & Romo, 2002). It is also known that motor cortical neurons, even in primary motor cortex, receive input from lower-level sensory areas, such as somatosensory (Salinas & Romo, 1998) and visual (Fadiga, Fogassi, Gallese, & Rizzolatti, 2000) cortices.

Outside the domain of single-unit recordings, one study in particular makes the case that the accumulation of difference scores to threshold is a general mechanism. Using fMRI, Heekeren, Marrett, Bandettini, and Ungerleider (2004) applied these ideas to human subjects performing a complex object classification task. Subjects were to

classify pictures of faces and houses by button press (2AFC). Importantly, the images were presented amidst varying levels of visual noise. Some images were presented amidst enough visual noise to be nearly subthreshold, while others were presented well above threshold, with little or no visual noise. Previous to performing this task, Heekeren, et al. identified, in each subject, areas of ventral temporal cortex that responded better to pictures of houses than faces, and other areas that preferred faces over houses. Heekeren et al. made two predictions. First, if the PFC is indeed important for the decision making process, its activity should vary *negatively* with the amount of visual noise embedded within the images. Stated differently, when pictures are presented with a high level of noise, the “evidence” being accumulated is weak; thus any accumulator should show a reduced response relative to a trial in which pictures were presented well above threshold. This is conceptually similar to the observation that FEF and LIP neurons “ramp up” at a rate dictated by coherence in the dot motion task (Roitman & Shadlen, 2002; Hanes & Schall, 1996). Second, Heekeren et al. reasoned that if the PFC accumulated a *difference* score between pools of selective neurons, as suggested by Kim and Shadlen (1999), then PFC activity should correlate *positively* with the absolute difference between house and face-selective areas. Difficult discriminations (i.e., a small difference score) should yield, functionally, a smaller difference in activity between the house- and face-selective regions. Thus, a smaller response would be expected. Both of these hypotheses were supported: the only region found to both 1) correlate negatively with visual noise levels and 2) correlate positively with the difference between face- and house-selective regions was a region of the dorsolateral PFC.

The Lateralized Readiness Potential

The Heekeren et al. (2004) study suggests that the accumulation of a difference score emanating from pools of specialized neurons may be a general property of decision making. Furthermore, this study makes an important step in extending monkey single-cell data to human subjects. However, because fMRI has poor temporal resolution, one can not use this methodology to study the SAT. One cannot accurately determine the level of activity at the moment a response is produced. Essentially, one must be able to gauge the level of activity within a very short time window when an overt response is emitted. What is needed is a measure that can ethically be applied to human subjects while retaining an approximation to single-cell recordings. Based on the above review of the neural literature, this measure should 1) reflect pools of response-selective neurons, 2) reflect the integration of evidence, over time, and 3) be sensitive to parameters inherent in sequential-sampling models such as prior probability and criterion. Research indicates that the lateralized readiness potential (LRP) component of the electroencephalogram provides such a measure.

The LRP is computed as the average voltage difference between electrodes C3 and C4 in the 10-20 system (Nuwer et al., 1998). Electrode C3 resides over the left hemisphere, displaced laterally from zenith, while C4 resides at the comparable location over right hemisphere. The LRP takes advantage of the fact that the brain is lateralized – that is, left-hemisphere cortical areas control the right side of the body, and the converse. Thus, on trials where a left hand response is required, the LRP is computed as $(C4 - C3)$; on trials where a right hand response is required, $(C3 - C4)$ is used (see Figure 8). The resulting waveform, when negative, reflects preparation of the correct hand; when positive, the incorrect hand was prepared. The subtractive methodology has the

additional benefit of removing other event-related potentials unrelated to motor preparation. One is left only with a record of how much *more* or *less* active a given hemisphere is, relative to the other, across time. Data from a variety of sources, including intracranial recordings and magnetoencephalography confirms that the LRP is generated in motor cortex (Coles, Smid, Scheffers, & Otten, 1995, p. 99). As well, single-unit recordings in monkey motor cortex exhibit a nearly identical time course as LRP scalp potentials (see Coles, 1989 for a review). As mentioned previously, motor cortical neurons represent an evolving decision, similar to that observed in FEF and LIP. Hence, the LRP reflects the difference in activity between pools of neurons selective for rightward and leftward movements.

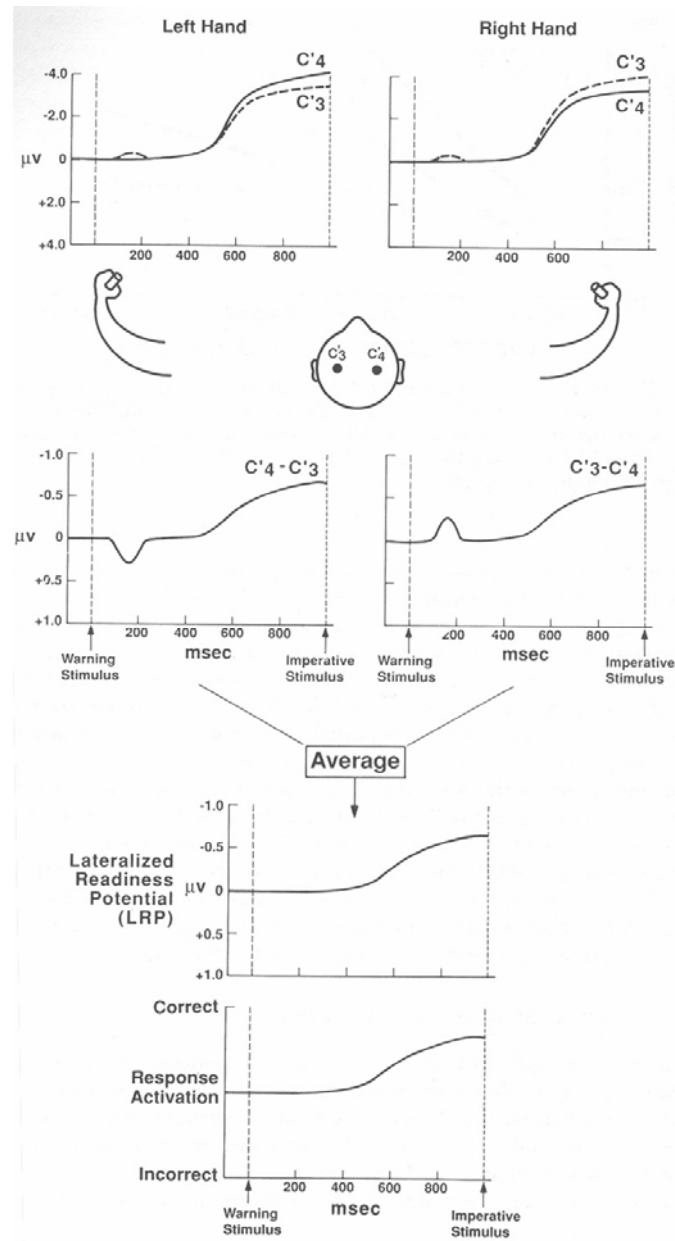


Figure 8. The derivation of the LRP. When a *left* hand response is correct, C4-C3 produces a *negative* LRP. The reverse holds when the *right* hand is correct. If an error is to be produced (as by preparing the left hand when the right hand is correct), this derivation produces a *positive* LRP.

The second test, that the LRP is sensitive to integration of evidence over time, is also strongly supported. In one important study, Gratton et al. (1988) presented subjects with a response-compatibility paradigm known as the “flanker” task (Eriksen & Eriksen,

1974). In this task, subjects viewed strings of 5 letters composed of “H” and “S”, and had to respond to the central letter. Importantly, H’s and S’s were mapped to opposing hands. The relevant data concern performance on incompatible trials, where the central letter was surrounded by response-incompatible letters (e.g., HSHH). Gratton et al. found that at very fast latencies on incompatible trials, responses were more likely to be driven by the flanker letters, leading to below chance performance. The LRP, too, closely matched this, tending to drift towards the incorrect hand at fast latencies. More impressively, they showed this pattern even on *correct* incompatible trials. Thus, even when a trial was eventually responded to correctly, the LRP showed a momentary “dip” towards the incorrect hand, suggesting that the LRP indexed the accumulation of evidence, over time, in a trial.

The third test, that the LRP is sensitive to sequential-sampling parameters is backed by at least two important studies. Gratton et al. (1988) found that overt responses were produced at the same level of LRP lateralization, regardless of the RT. That is, whether a trial was responded to quickly or slowly, the response occurred when the LRP reached a certain level of lateralization. It is not difficult to see the relationship between this observation and the single-cell data reviewed earlier (e.g., Roitman & Shadlen, 2002; Hanes & Schall, 1996). It appears that overt responses are “triggered” by a stereotyped level of neural activity. This was quite clear in the single-cell studies (Figure 7, right panel); the Gratton et al. (1988) study shows that there is an LRP analog of the fixed criterion seen in the single-unit research. This finding has been replicated many times since (e.g., Miller, 1998; Mordkoff & Grosjean, 2001).

The LRP is also sensitive to prior probability, another important factor in

accumulator models. Miller (1998) presented subjects with a simple 2-choice classification experiment. Importantly, he manipulated the probability of encountering the stimuli. The result of this action was to elicit a “pre-lateralization” of the LRP before the critical stimulus even appeared. It is as if subjects were preparing for the stimulus they were led to expect by priming a given hand. This too, emphasizes the fact that the LRP is sensitive to cognitive factors; it is not simply a motor phenomenon.

To summarize, mathematical models of the decision process successfully model the SAT as a shift in criterion. With lower criteria, less evidence supporting a given alternative is required before making a response, leading to a shorter RT. Single-unit recordings in monkey cortex reveal that sequential-sampling models, as described, are biologically plausible. Each of the 3 critical parameters has a neural correlate: neuron activity ramps up at a rate proportional to the quality of evidence, the baseline firing rate of neurons is modulated by prior probability, and a response is produced at the precise moment neural activity reaches a fixed level, regardless of response latency. Using the LRP, a measure of the relative neural activity emanating from left and right motor cortex, research shows many of the same patterns. The LRP can be “pre-lateralized” by manipulating the probability with which one expects a right or left hand response. Like the single-unit data, a response is produced at a stereotyped level of LRP activity, regardless of response latency.

The study to be reviewed below puts to test the hypotheses that criterion – in the neural sense – can be modulated by changes in response deadline. Response deadlines are implemented as a time limit, on each trial, that a subject has to respond. Deadlines are blocked, so that subjects can predict how much time they have on a given trial.

Missed deadlines are typically marked by a tone or message that instructs the subjects to be faster, even at the expense of accuracy. Experimental evidence exists confirming that enforced response deadlines do in fact affect the SAT (e.g., Rinkenauer et al., 2004).

Evidence supporting this hypothesis will take the following forms. First, like the studies previously mentioned, the LRP should exhibit a constant voltage at the moment a response is produced, regardless of RT, *within a given deadline block*. Thus, holding SAT constant holds the LRP threshold constant. Between deadline blocks, however, criterion LRP voltage should be related to the deadline enforced for that block. LRP voltage at response should be lower for shorter deadlines, and higher for longer deadlines. Stated differently, when subjects adopt, for example, an “accuracy” setting, allowing for longer latencies, more information may be accrued prior to response. This is expected to be reflected in the LRP.

CHAPTER 2
EXPERIMENT 1

Method

Participants

Six subjects were recruited from a pool of Georgia Tech Psychology graduate students. Participants were paid \$10.00/hour in a single 3-hour session. All subjects had normal or corrected-to-normal vision.

EEG and EMG Recording

Electrophysiological signals were recorded using a BioSemi Active-Two 32-channel EEG. Ag/Cl electrodes were securely fastened to the scalp using a nylon cap conforming to the modified 10-20 system (Nuwer et al., 1998). Connections were maintained using conductive electrolyte gel (SignaGel). As the electrodes were “active” (each electrode contained its own amplifier), impedances were not an issue. EEG signals were sampled at 1024 Hz, run through an amplifier and ADC converter, and recorded for off-line analysis. No analog filter was employed; all filtering was accomplished digitally during off-line analysis.

Bipolar electrodes were used to record the electromyogram (EMG) and electrooculogram (EOG). The horizontal EOG (hEOG) provided a measure of eye movement artifacts for later removal. The vertical EOG (vEOG) indicated the presence of blinks, which were later removed using an ocular artifact reduction algorithm (Gratton, Coles, & Donchin, 1983). hEOG electrodes were placed approximately 2-3 cm away from the outer canthus of each eye (corresponding to the area commonly known as the “temple”). vEOG electrodes were placed approximately 2-3 cm above and below each

subjects' left eye. Bipolar electrodes were also used to record EMG activity as a measure of response production. The onset of a manual button press is considerably shorter when measured by EMG than when measured by the closing of an electronic circuit. Some estimates of this differential are as small as 17 ms (Mordkoff & Grosjean, 2001), while others are as large as 80 ms (present data set). Thus, EMG provides a more accurate measure of response onset. The placement of bipolar electrodes for EMG varied with subject. For some subjects, the cleanest signal (estimated by eye) was produced by 2 electrodes placed longitudinally on the inner surface of the wrist, just adjacent to the palm (Stern, Ray, & Quigley, 2001). These electrodes were typically separated by 3-4 cm, but this was not a constant. For other subjects, the cleanest EMG signal arose from electrodes placed on the outer surface of the hand, with one electrode placed directly on the index finger and a reference placed approximately 3-4 cm posterior.

EEG, EOG, and EMG signals were band-pass filtered with a low cutoff of .01 Hz and a high cutoff of 30 Hz. As mentioned previously, blink artifacts were regressed out by algorithm, while trials with eye movements were rejected (segments with activity above or below 75 μ V were rejected). For the stimulus-locked LRP (S-LRP), explained in detail below, epochs lasted 800 ms, beginning 100 ms before the critical stimulus. Each epoch was baseline-corrected using the average amplitude in the 100 ms window prior to the onset of the critical stimulus. For response-locked LRP (R-LRP), epochs lasted 800 ms, beginning 400 ms before manual response and ending 400 ms after response. These waveforms were baseline corrected using the interval 400 ms to 300 ms before response onset (Rinkenauer et al., 2004).

Eriksen Flanker Task

Participants performed a version of the flanker task originally developed by Eriksen and Eriksen (1974). The task parameters were quite similar to that reported by Heitz & Engle (in-press). Subjects viewed sequences of 5 letters, composed only of *H* and/or *S* characters. Subjects were to indicate, by button press, the identity of the central letter. If the center letter was *S*, subjects pushed the “z” key with their left hand, while the “m” key was pressed if the center letter was *H*. On compatible trials, all letters were identical (HHHHH or SSSSS). On incompatible trials, the outer letters were consistent with the incorrect response (HSHHH or SSHSS). A fixation dot above the central letter was always visible before the array appeared to aid in localization. Stimuli were presented in light gray against a black background.

Each trial began with the presentation of a 50 ms warning tone, followed 1000 ms later by the critical stimulus, subtending approximately 2.5 degrees of visual angle. The critical array remained visible until button press or 1500 ms. Following response, a random interval of 1000 or 1500 ms preceded the onset of the next trial.

The SAT manipulation was enforced by three response deadlines. The *slow* condition used a 550 ms deadline, the *medium* a 450 ms deadline, and the *fast* a 375 ms deadline. These deadlines were selected based on pilot testing and previous research using this task (Heitz & Engle, in press). Deadlines were enforced by presenting the message “Deadline missed...Faster!” after each RT that surpassed the designated cutoff. Critically, subjects did not know if a deadline was missed or not until after making a response; in other words, subjects were allowed to make a 600 ms response even in the 375 ms deadline. The stimulus remained visible for a maximum of 1500 ms. To ensure that subjects did not disregard the deadline messages, the program kept a running total of

the proportion of missed deadlines. Every 10 trials, this average was checked. If the proportion of missed deadlines fell under 66%, a new message appeared for 5 seconds: “You are missing too many deadlines. You must respond faster even if errors result.”

Prior to beginning the experiment, subjects were informed of the meaning of SAT and were told that they would be tested at three levels of speed stress. They were told to attempt to “move” along the SAT continuum, even if the response deadlines did not appear to change. Following this and general instructions, subjects performed 20 trials of mapping practice, which familiarized them with the response-mapping. Only compatible trials were presented during mapping practice and feedback was provided immediately after response. Next, subjects began the experimental block. Although it was not detectable to the subject, the first 20 trials of the first block were considered “deadline practice.” This allowed subjects to learn what time-band they were expected to respond in. No feedback was provided in the experimental block. Subjects completed 5 blocks of trials at the *slow* deadline, 5 blocks at the *medium* deadline, and 5 blocks at the *fast* deadline, in that order. While this led to the confounding of SAT with practice, there is good reason to structure the task in this manner. First, it is unclear whether or not subjects can easily return to a *slow* cognitive set after performing hundreds of trials in comparably faster cognitive sets. Thus, as a starting point, it makes sense to begin with the *slow* condition and gradually increase speed stress. Second, unpublished work has revealed that the peak voltage of the LRP is unaffected by extensive practice (up to 4 hours; personal communication with Gabriele Gratton, 6 February, 2007).

Each block of trials consisted of 120 trials: 60 of which were compatible and 60 of which were incompatible (50/50). Of these, half were left hand responses and half

were right hand responses. Because each block was run 5 times in each speed stress condition, this led to a total of 150 trials/SAT condition/compatibility condition/hand. Hence, the total number of experimental trials was 1800. Trials were coded for SAT condition, compatibility condition, correct/incorrect, made or missed deadline, and left/right hand.

Results

Primarily, averaged data shall be reported in the main text, but abbreviated subject-level data are presented in Appendix A. With few exceptions, similar patterns were evident at the subject-level. Thus, the reader can be assured that the below patterns do not arise from an aggregation bias, nor are they carried by a small set of subjects. The data will be presented in two main sections; the first presents RT and accuracy rate analyses including time-course decomposition (Heitz & Engle, in press) and model fits to the LATER sequential-sampling model (Carpenter & Williams, 1995). The main purpose of this first section is to establish that the deadline manipulation successfully induced a speed-accuracy tradeoff and to demonstrate that the patterns are consistent with the sequential sampling models described above. The second section presents electrophysiological analyses, focusing on the LRP.

Behavioral Data

Data were first structured to support a time-course analysis, to be reported later. To do so, a Vincentizing procedure was used: Each subject's RT was rank-ordered separately for each compatibility and SAT condition. Ten *n*tile bins were then computed, the first bin corresponding to the fastest 10% of trials, the second bin to the next fastest 10% of trials, and so forth. The analyses to be reported below were run in the context of

a 2 (compatibility: compatible vs. incompatible) x 3 (SAT: fast, medium, slow) x 10 (RT bin) repeated-measures ANOVA. Huynh-Feldt corrections were employed for violations of sphericity.

The speed-accuracy tradeoff was apparent in the behavioral data (i.e., RT and accuracy rate) as reflected in Table 1. Subject-level data are presented in Table A1 and A2.

Table 1.

Mean RT (ms) and Accuracy Rate by Compatibility and SAT				
Condition		Slow	Med	Fast
Compatible	RT	433	377	320
	ACC	.96	.88	.74
Incompatible	RT	477	400	325
	ACC	.83	.66	.50

For RT, there was a main effect of SAT condition, $F(2,10) = 110.79, p < .001$, partial $\eta^2 = .96$, and a main effect of compatibility, $F(1,10) = 34.45, p < .01$, partial $\eta^2 = .87$. Repeated contrasts revealed that the *fast* condition was significantly faster than the *med* condition, $F(1,10) = 102.63, p < .001$, partial $\eta^2 = .95$, and the *med* condition was significantly faster than the *slow* condition, $F(1,10) = 69.72, p < .001$, partial $\eta^2 = .93$. The main effect of compatibility, reported above, is commonly known as the “flanker effect” and reflects the slowing of incompatible responses due to response conflict (Eriksen & Eriksen, 1974). There was also a compatibility x SAT interaction, $F(2,10) = 37.86, p < .001$, partial $\eta^2 = .8$; as Table 1 suggests, the effect of SAT condition in slowing RTs was somewhat stronger for incompatible than compatible trials. Similar patterns were evident for accuracy rate: There was a main effect of SAT condition, $F(2,10) = 98.60, p < .001$, partial $\eta^2 = .95$, indicating that accuracy rates tended to decline as speed stress increased. Evident also was a main effect of compatibility, $F(1,10) =$

88.52, $p < .01$, partial $\eta^2 = .95$, indicating that incompatible trials were less accurate than compatible trials. Unlike the RT analyses, the SAT x compatibility interaction did not attain significance for accuracy rate.

It is worthy of mention that the mean accuracy rate for the *fast*-incompatible condition is .50 (see Table 1). Generally, this might be taken as evidence that subjects were randomly guessing. However, there is good reason to doubt this supposition. First, if subjects were randomly guessing in this condition, it stands to reason that the accuracy rate for compatible trials should also approach .50. It is difficult to envision a situation where subjects randomly guess only on incompatible trials. Second, as will be shown in the below time-course analyses, accuracy rates for incompatible trials typically fall *below* chance during a specific time-band (Gratton et al., 1988; Heitz & Engle, in press) and then *recover* to .50 a bit later. Below chance performance reflects a process that is accumulating useful information (i.e., subjects are not guessing), but at the point of response, contains more evidence in favor of the incorrect response than the correct response (see Heitz & Engle, in press, for a review). The RT associated with .50 performance (325 ms) is consistent with the “recovery” time-band in previous research.

Latency Distributions

Figures 9 and 10 present latency distributions for compatible and incompatible trials, respectively, as a function of SAT condition. Observations < 150 ms were discarded from analyses as anticipations. Doing so eliminated less than .005% of the data. The upper panels depict correct trial latencies, the bottom panels error latencies.

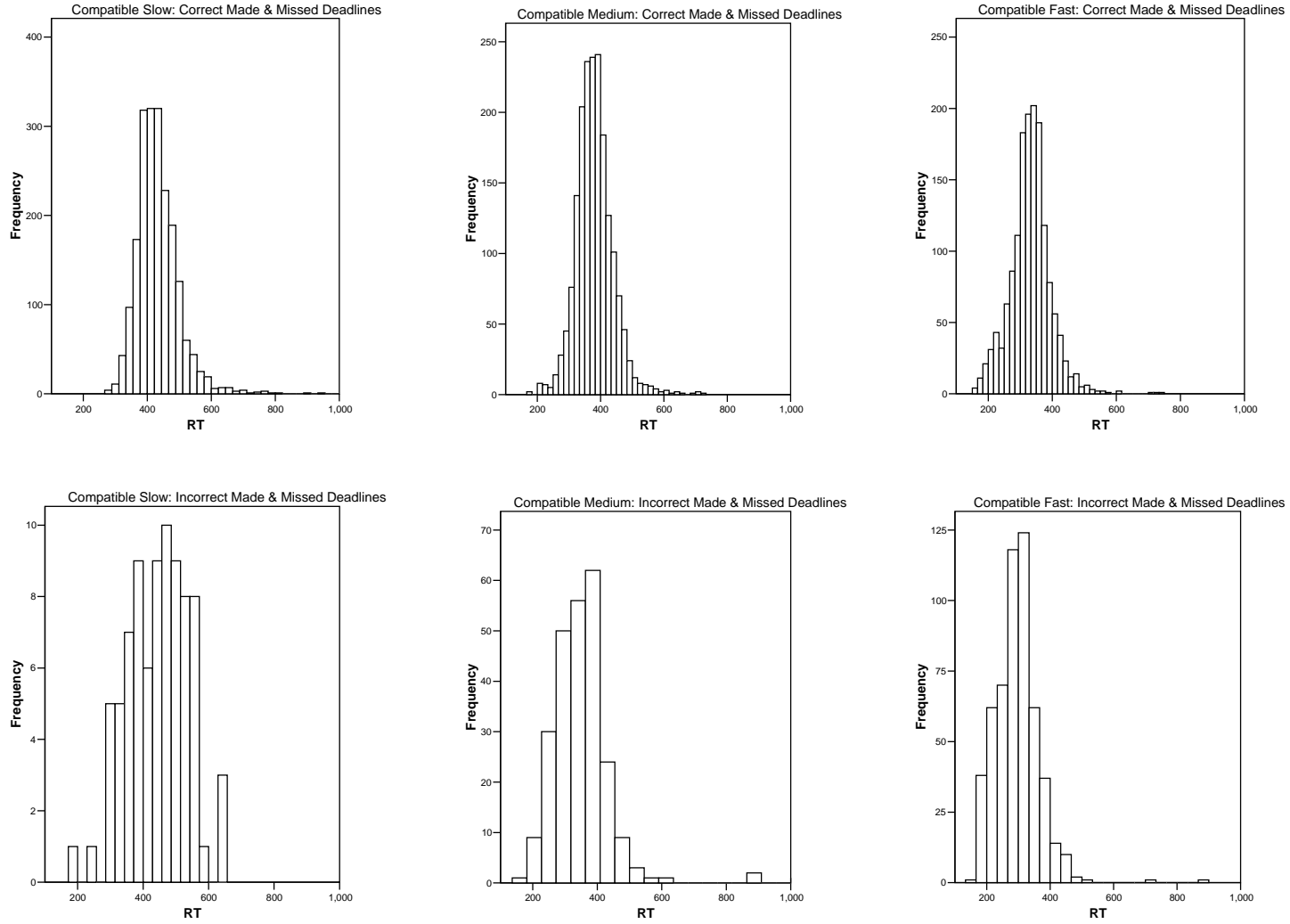


Figure 9. Compatible correct (upper row) and incorrect (lower row) latency distributions.

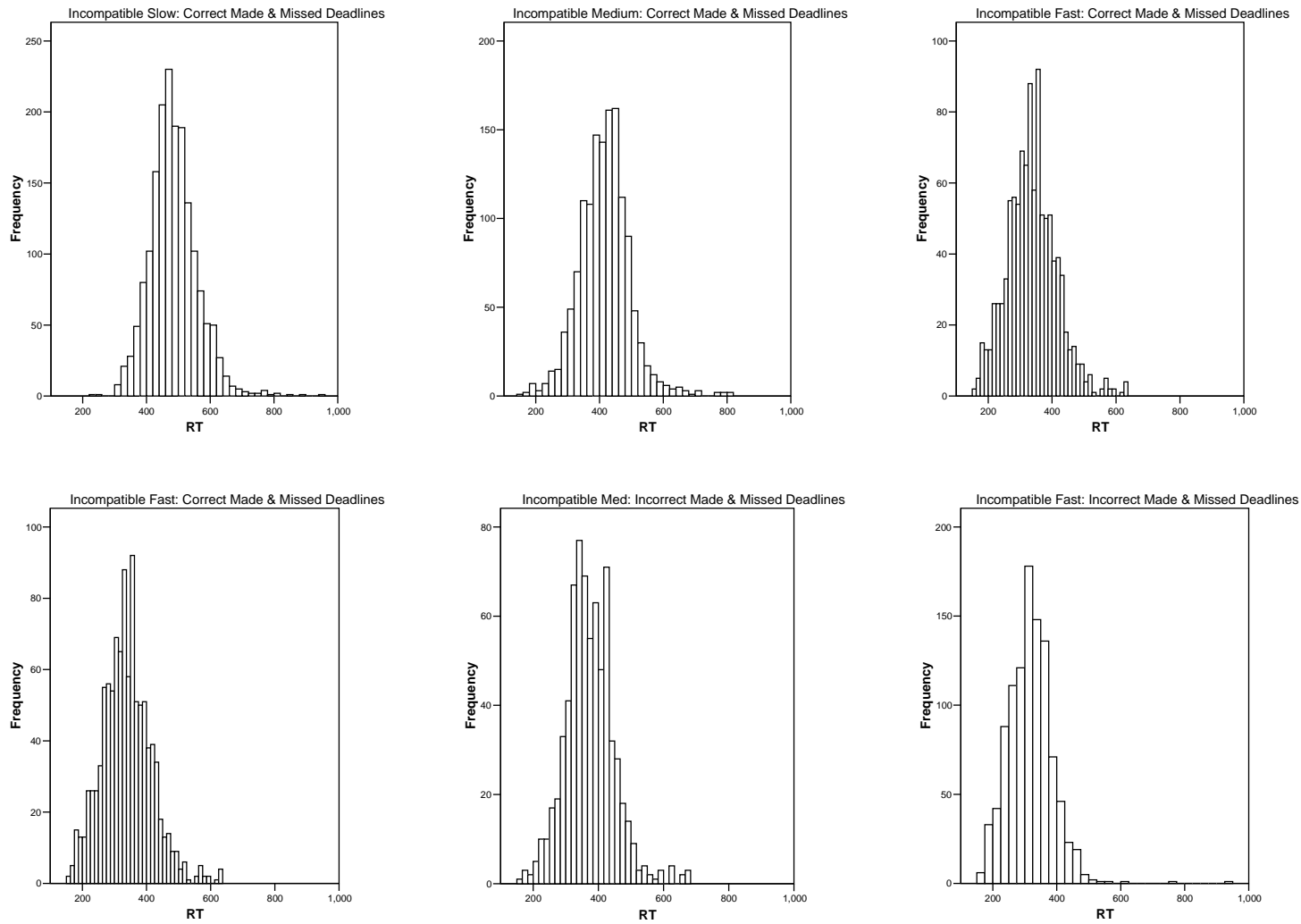


Figure 10. Incompatible correct (upper row) and incorrect (lower row) latency distributions.

Although it is impossible to measure exactly one's internal response criterion using behavioral data alone, it is possible to glean some information regarding its properties. It is thought that one sets their response criteria in such a way as to maximize reinforcement rate (Lo & Wang, 2007; Shadlen & Newsome, 2002; Bogatz, 2007; Bogatz et al., 2006). In the present context, an optimal response criterion would allow for as much processing time as possible (ensuring an accurate response) while beating the deadline on a large proportion of trials. Presented in Table 2 are the mean proportion missed deadlines along with the median RT associated with made and missed deadlines for each SAT condition.

Table 2.

% Missed Deadlines and Median RT (ms) for Made and Missed Trials: Experiment 1						
Condition	Slow (dead = 550)		Med (dead = 450)		Fast (dead = 375)	
Compatible	4.7%		10.1%		16.9%	
	422	590	368	474	312	404
Incompatible	15.4%		24.1%		22%	
	463	593	376	484	306	413

Given the nontrivial proportion of missed deadlines, it is reasonable to suggest that subjects tended to use as much time as they were allotted. In other words, their criterion was set in such a way as to use as much processing time as possible while allowing for some acceptable proportion of missed deadlines. Although it would have been possible for subjects to adopt a criterion that beat the deadline on all trials, this would come at a cost of accuracy due to a loss in processing time. A 2 (compatibility) x 3 (SAT condition) ANOVA confirmed that the proportion of missed deadlines was larger for incompatible than compatible trials, $F(1,5) = 28.63, p < .01$, partial $\eta^2 = .85$ and tended to increase with speed stress, $F(2,10) = 11.05, p < .05$, partial $\eta^2 = .689$. A compatibility x

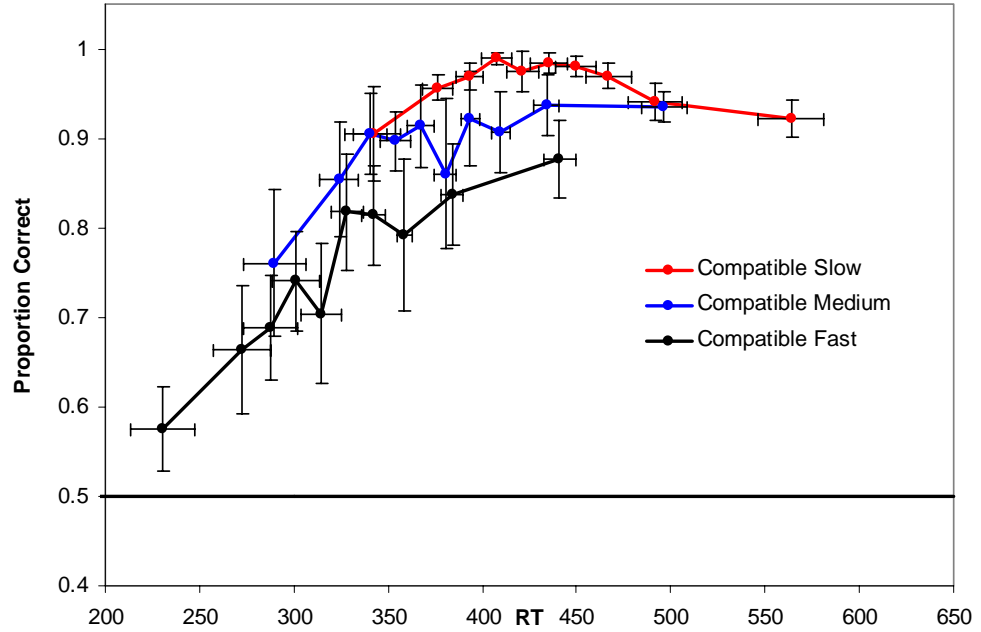
SAT quadratic interaction reflects the fact that missed deadlines were about equally likely in the incompatible *med* and incompatible *slow* conditions, but monotonically increased for compatible trials, $F(2,10) = 4.60$, $p < .05$, partial $\eta^2 = .48$.

Conditional Accuracy Functions

Conditional accuracy functions (CAFs) plot accuracy rate conditional on RT (Wood & Jennings, 1976). These plots depict how performance changes within a condition due to variability in processing time, and will provide some indication of whether or not subjects adopted different response criteria in the different SAT conditions. One can readily assess the extent to which SAT criteria changed by comparing accuracy rates at a given RT. When response criteria change, then the accuracy rate associated with an RT of, say, 500 ms will be dependent on SAT condition, decreasing as speed stress increases. Alternatively, if the deadline manipulation was ineffective, accuracy rates should be identical, holding RT constant (i.e., the accuracy rate associated with 500 ms will be the same in the *fast*, *med*, and *slow* conditions).

Figure 11 depicts these functions for compatible trials. The CAFs are presented in the upper panel of Figure 11; the lower panel plots observed RT as a function of *ntile* bin and SAT condition. Comparable plots for incompatible trials are presented in Figure 12. To analyze the upper-panel CAFs using ANOVA, it must first be established that it is tenable that the latency distributions are equal between the latency bins. In other words, an ANOVA assumes that the data points for each CAF (i.e., for the *fast*, *med*, and *slow* lines) are aligned with respect to the *x*-axis. If this is not the case, then ANOVA is technically inappropriate.

Compatible: SAT x RT



Compatible: Observed RT by Vincentized RT Bin

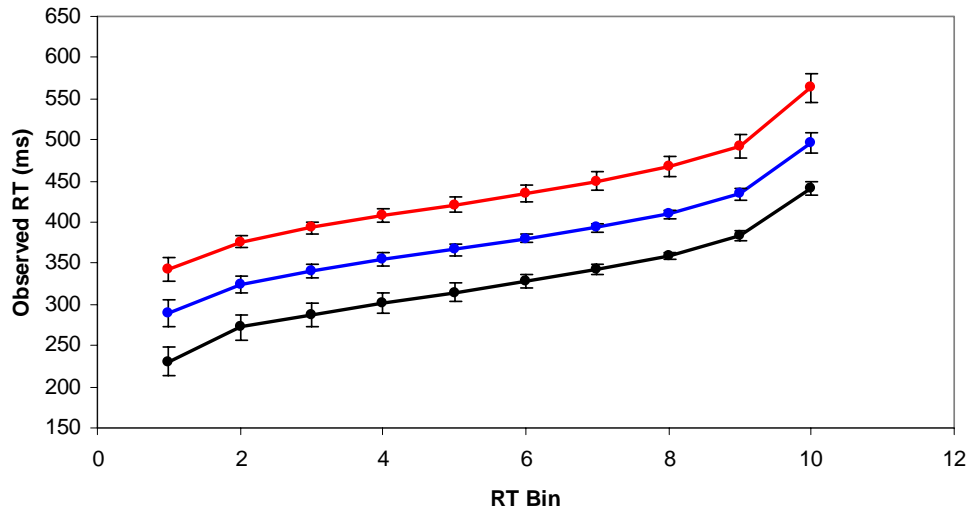


Figure 11. Upper panel: Conditional accuracy functions for compatible trials. Horizontal bars depict ± 1 S.E.M for RT, vertical bars depict the same for accuracy rate. Lower panel: Latency bin RTs. Vertical bars depict ± 1 S.E.M.

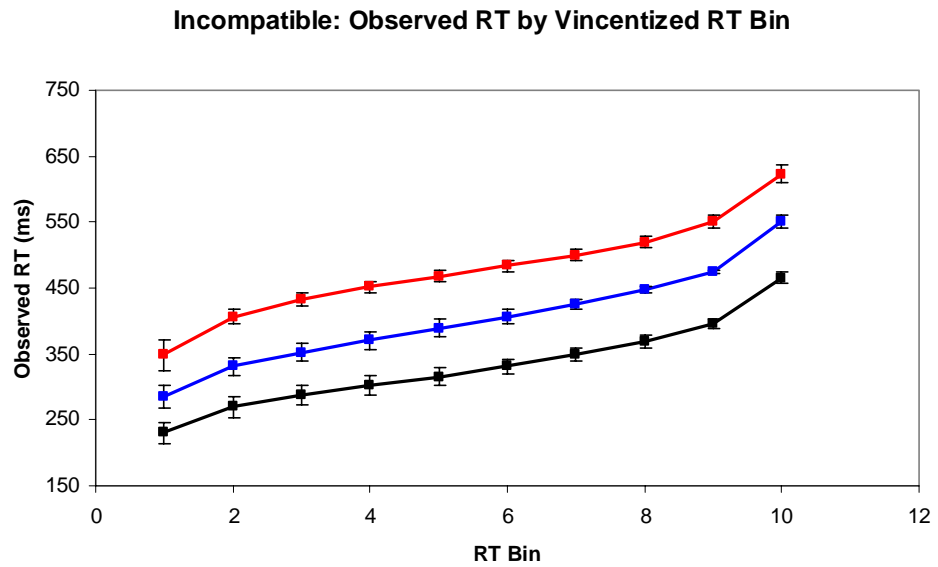
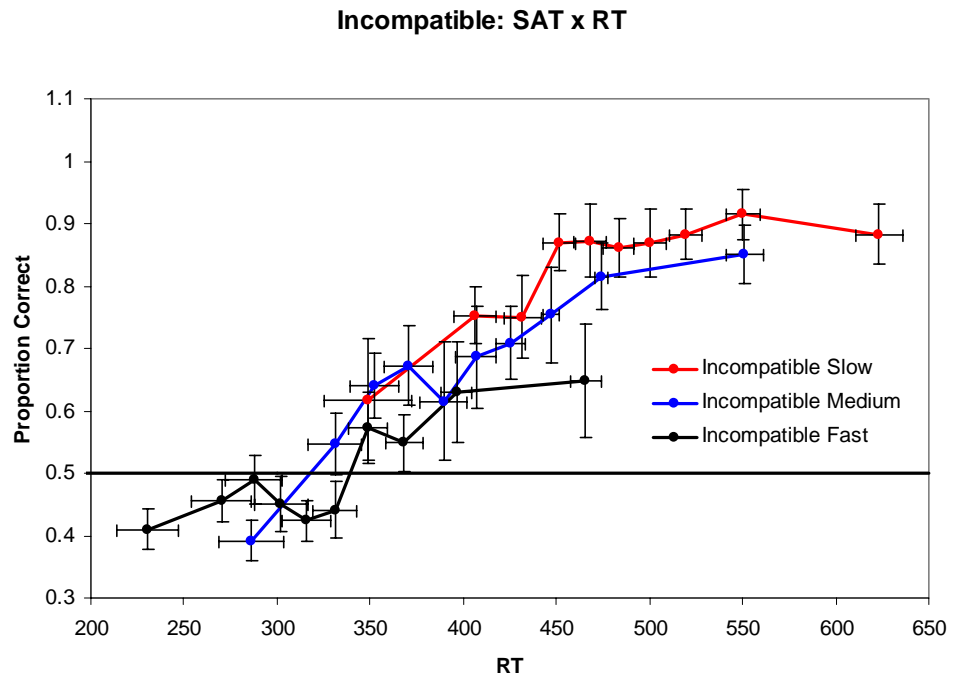


Figure 12. Upper panel: Conditional accuracy functions for incompatible trials. Horizontal bars depict ± 1 S.E.M for RT, vertical bars depict the same for accuracy rate. Lower panel: Latency bin RTs. Vertical bars depict ± 1 S.E.M.

A brief look at Figures 11 and 12, however, indicate that this is sure to be a problem; the data points vary quite widely along the x -axis. This is confirmed by a RT bin \times compatibility \times SAT interaction $F(18,90) = 3.31, p = .051, \text{partial } \eta^2 = .40$, although the main effect of SAT condition on RT, reported earlier, is sufficient to render ANOVA inappropriate. In lieu of omnibus tests, t -tests were used to evaluate specific mean contrasts, while the Friedman nonparametric test was used as a surrogate for a one-way repeated-measures ANOVA.

Considering first the compatible trials, two observations are noteworthy. First, accuracy rates tend to increase with processing time, as would be expected. Second, the accuracy rate associated with a given RT is dependent on the SAT condition. This was verified statistically using the Friedman nonparametric test. Three approximate time points were chosen that appeared to contain data from all three SAT conditions. The first test corresponded to an RT of between 430 and 450 (*slow* bin 6, *med* bin 9, *fast* bin 10). This test was significant, $\chi^2(2) = 9.33, p < .01$. The second test considered RTs between 370 and 385 ms (*slow* bin 2, *med* bin 6, *fast* bin 8); this test was also significant, $\chi^2(2) = 8.82, p < .05$. Finally, RTs between 340 and 342 were tested (*slow* bin 1, *med* bin 3, *fast* bin 7), however this test was not significant $\chi^2(2) = 4.26, ns$, likely due to the overlap between the *slow* and *med* condition at this point.

Incompatible CAFs showed similar patterns: Performance gradually increased over time, and performance levels at a given RT were dependent on SAT condition. Verifying what is evident in Figure 12, a Friedman nonparametric test considering *slow* bin 5, *med* bin 9, and *fast* bin 10 (RTs between 465 and 475 ms) yielded significance, χ^2

(2) = 7.04 $p < .05$, as did a similar test within an RT time band of 368 – 370 ms (*slow* bin 2, *med* bin 4, and *fast* bin 8), $\chi^2(2) = 6.33, p < .05$. Also noteworthy is the “dip” below chance performance evident for half of the *fast* condition and bins 1 and 2 of the *med* condition. As mentioned previously, this pattern is suggestive of a process that has begun to accumulate useful information, but has not yet filtered incompatible distractor letters which tend to lead to the opposite response (Heitz & Engle, in press). This pattern was significant for the first RT bin in the *fast* condition, $t(5) = -2.78, p < .05$ and for the *med* condition, $t(5) = -3.38, p < .05$. That it was the first bin that fell below chance is noteworthy. In particular, it indicates that the .50 accuracy rate observed later (bins 2-5 for the *fast* condition and bins 1-3 for the *med* condition) are not indicative of random guesses. Rather, it is simply that at these time points, interference from incompatible distractors balances the information from the target letter.

To summarize the above, mean comparisons and time-course analyses suggest that the deadline manipulation did in fact alter subjects’ response criteria. However, this does not specifically speak to how these data are accommodated by a sequential sampling process. As mentioned in the introduction, sequential sampling models make specific predictions regarding how criterion shifts should affect latency distributions. The next section considers fits to the LATER model (Carpenter & Williams, 1995).

LATER Model

Behavioral data (for correct trials only) were fit to the LATER model (Carpenter & Williams, 1995) using SPIC software (Carpenter, 2007). The LATER model is a linear rise-to-threshold model that has been used extensively to dissociate manipulations

that affect drift rate and manipulations that affect criterion (Carpenter & Williams, 1995; Reddi & Carpenter, 2000; Reddi, Asrress, & Carpenter, 2003; Hanes & Carpenter, 1999). The fitting procedure is straightforward but makes use of a plot known as a “reciprobit.” As this is a non-intuitive method of data presentation, some description is warranted. The first step in deriving a reciprobit stems from the observation that RT distributions generally follow a Gaussian distribution when they are plotted on a reciprocal scale. As a rule, latency distributions plotted on a normal scale have a positive skew, leading many to argue that they are best fit by an exponential-Gaussian rather than the Normal (Spieler, Balota, & Faust, 2000). After converting the x -axis to a reciprocal scale, the distribution is transformed into a cumulative density function (CDF) on a probit scale. Essentially, the probit scale “stretches” the CDF into a straight line (Figure 13).

The LATER model makes predictions at the distribution level, but is inferior to more developed models such as the diffusion (Ratcliff, 1978; Ratcliff & Rouder, 1998), as it cannot account for error latencies or error rates. However, the LATER model has been heralded for its ease of use, ability to fit many data sets, and clearly dissociate criterion effects from drift rate effects. Manipulations that affect drift rate (e.g., the perceptual quality of the stimuli) have the effect of changing the skew of a latency distribution when plotted on a traditional scale. On reciprobit, a manipulation that affects drift rate will cause a horizontal, parallel shift of lines. Hence, if two lines are laterally shifted, while slope remains invariant, the effect was in drift rate. Manipulations that

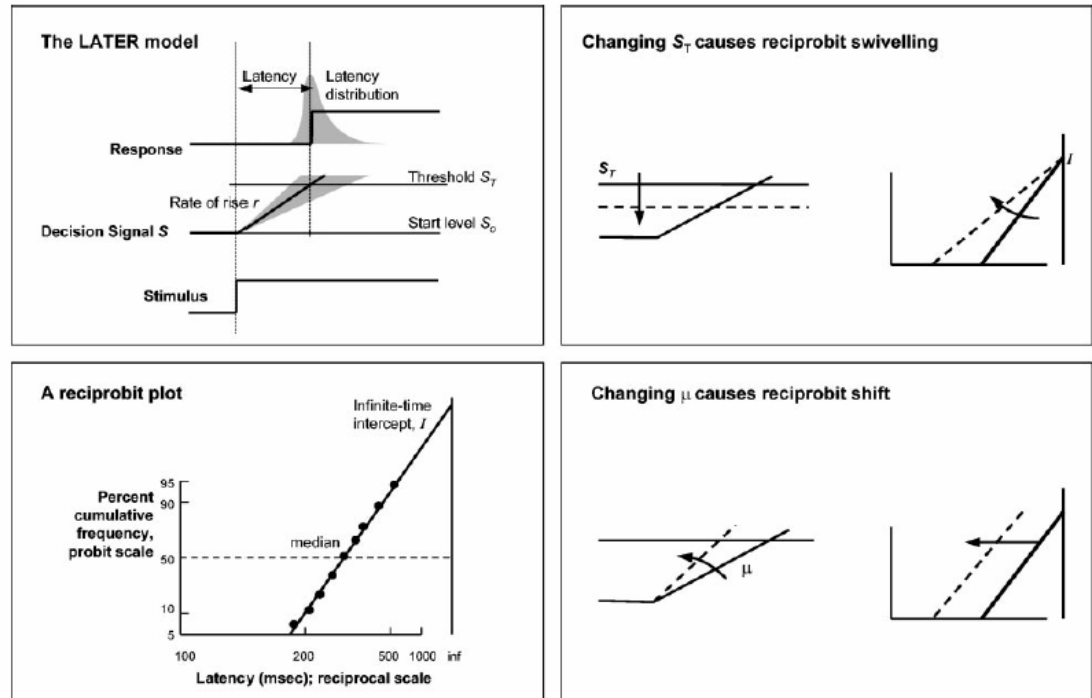


Figure 13. The effect of changes in drift rate and threshold on the reciprob plot (From Reddi, Asress, & Carpenter, 2003). Upper left: The shape and placement of a latency distribution is a function of both drift rate and criterion. Lower left: A reciprob plot. The median of the distribution intercepts the 50% cumulative frequency. Upper right: changes in threshold causes a reciprob “swivel.” Lower right: changes in drift rate causes a reciprob “shift.”

affect the distance between baseline and criterion (as in a rising or lowering criterion) tend to exert its effect primarily on the placement of the response time distribution. On reciprob, this leads to a “swiveling” of lines about a constant intercept. In other words, the slopes will change (Figure 13). Previous research (Reddi & Carpenter, 2000; Reddi, Asress, & Carpenter, 2003; Hanes & Carpenter, 1999) shows that manipulations such as SAT instructions lead to reciprob “swivel,” whereas manipulations such as stimulus discriminability lead to reciprob “shift.” The “shifting” versus “swiveling” model

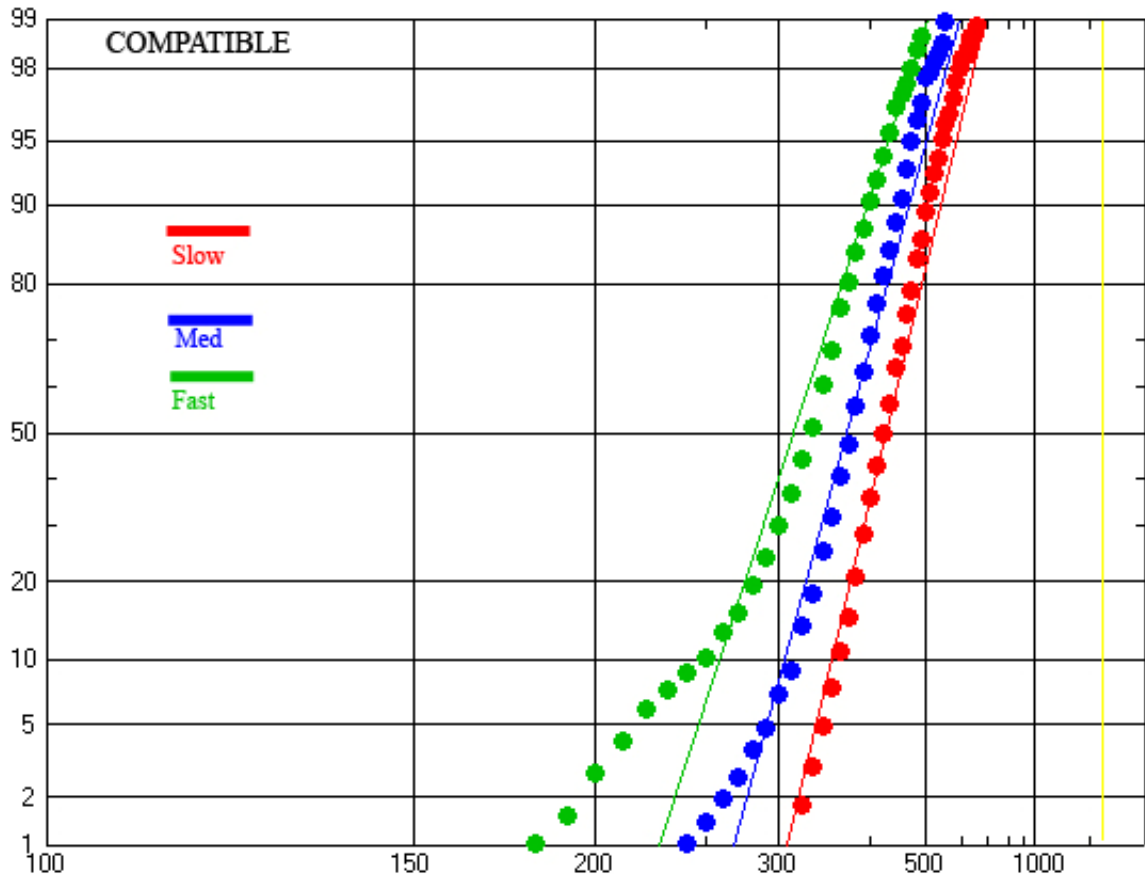


Figure 14. Reciprobit plot for compatible trials. The swiveling of the lines about a common intercept is consistent with a change in criterion/threshold.

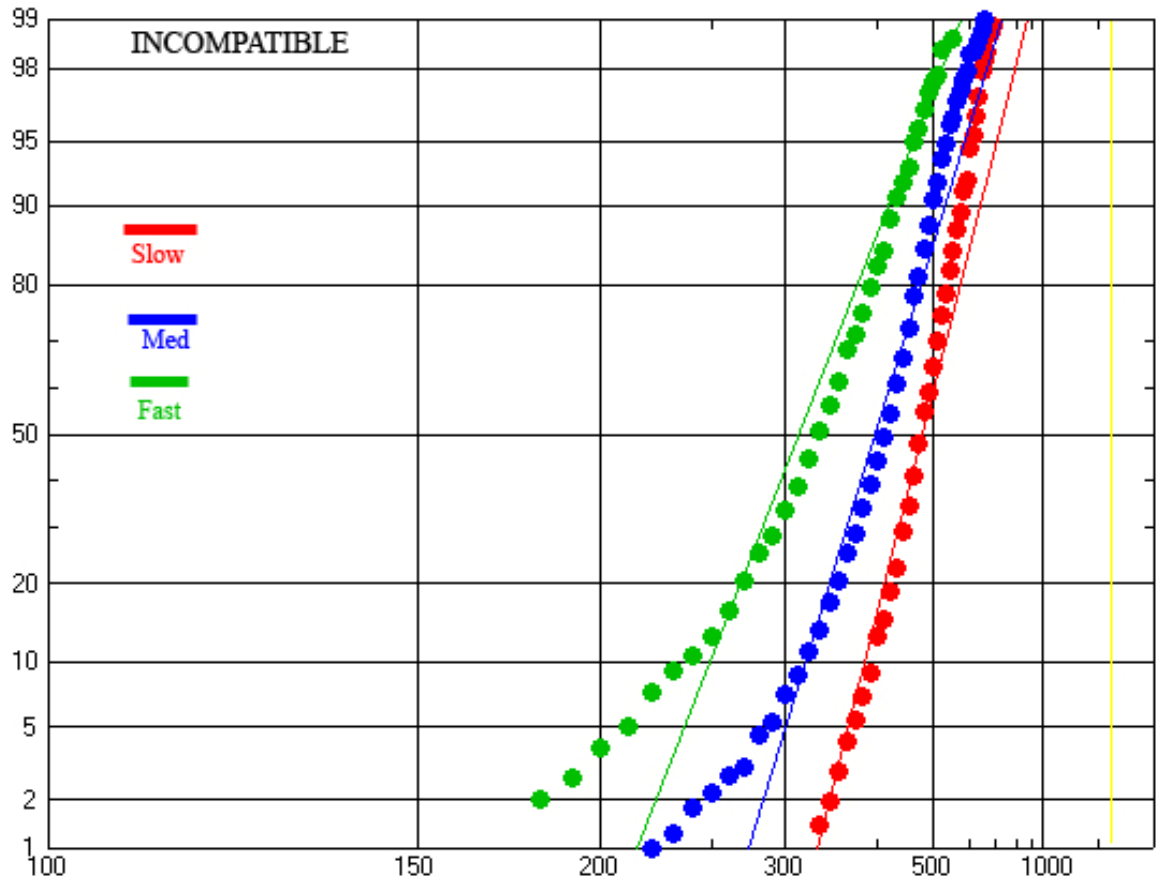


Figure 15. Reciprobit plots for incompatible trials. The swiveling of the lines about a common intercept is consistent with a change in criterion/threshold.

is specifically tested using a log likelihood ratio test (implemented in SPIC software, Carpenter, 2007). Figures 14 and 15 present reciprobbit plots for compatible and incompatible trials, respectively. As is evident, the SAT manipulation invoked a reciprobbit “swivel,” indicating a change in criterion/threshold, for both compatible and incompatible trials. These observations were confirmed by log likelihood ratio test, for compatible trials, $LLR = 17959, p < .001$; for incompatible trials, $LLR = 15106, p < .001$. This held when tested on a subject-level as well (see Appendix for subject-level tests and plots).

The above behavioral data are in agreement that the deadline manipulation affected subjects' criterion placement. As subjects were pushed to respond faster, criterion was lowered, leading to a decrease in RT but also an increase in error rate. LATER model fits were also consistent with the hypothesis that subjects accumulate information over time until a critical threshold (criterion) is reached, at which point a response is emitted. The deadline manipulation had the effect of lowering or raising this criterion, as evidenced by reciprobital "swivel."

Although these results are interesting in and of themselves, they are not altogether surprising. A long history of research has shown comparable effects of deadline manipulations and fits to the LATER model and other sequential sampling models (e.g., diffusion, leaky accumulator, etc.). The next section addresses the primary question of interest – namely, is there a neural correlate of SAT, and can it be observed to move in predictable ways? The argument is that the LRP is a measure of evidence accumulation; its amplitude at response is a measure of criterion/threshold. The theory of continuous flow (Eriksen & Schultz, 1979) argues that response channels are continually primed as information is accumulated (Gratton et al., 1988). Because the LRP reflects the priming of motor cortex preceding a hand movement, it stands to reason that the LRP reflects this accumulation process. Specifically, the prediction was that the threshold for movement onset, as measured by the voltage of the LRP at the point EMG activity is detected, should *decrease* as speed-stress *increases*. This follows from the hypothesis that as criterion lowers, less information is required before a response is emitted. It is for this reason that errors are more likely to result in a high speed-stress environment.

Electrophysiological Data

LRPs were derived for each combination of compatibility, SAT, correct versus incorrect, and made versus missed deadlines. Stable LRPs are achieved with approximately 60 trials/hand (Rinkenauer et al., 2004). Although the number of observations per hand for each condition was large (150), there were many instances where too few observations were collected to derive an LRP. These instances primarily concerned incorrect compatible trials and compatible trials with missed deadlines. All waveforms include only made deadlines, unless otherwise noted. All LRPs were computed so that correct activation results in a positive deflection, while incorrect activity produces a negative deflection.

EMG Onset Detection

The present research is primarily concerned with what are known as response-locked LRPs (R-LRP). The R-LRPs presented here consist of an epoch lasting 400 ms before and after response onset. Time 0 is thus the point of response. The LRP, although known as a “slow” potential, lasts only a few hundred milliseconds. Thus, accurately measuring the exact moment of response onset is problematic; if measurement of this moment is in any way delayed, the estimated measure of LRP amplitude needed for response production is flawed. For this reason, EMG activity was used to time-lock R-LRPs. EMG onset detection was accomplished by establishing, for each subject and each hand, a criterion voltage that clearly differentiated muscle activity from baseline noise. Although this criterion was unsystematically determined by eye, EMG activity tends to be very regular and very large with respect to background noise, so there is little question

as

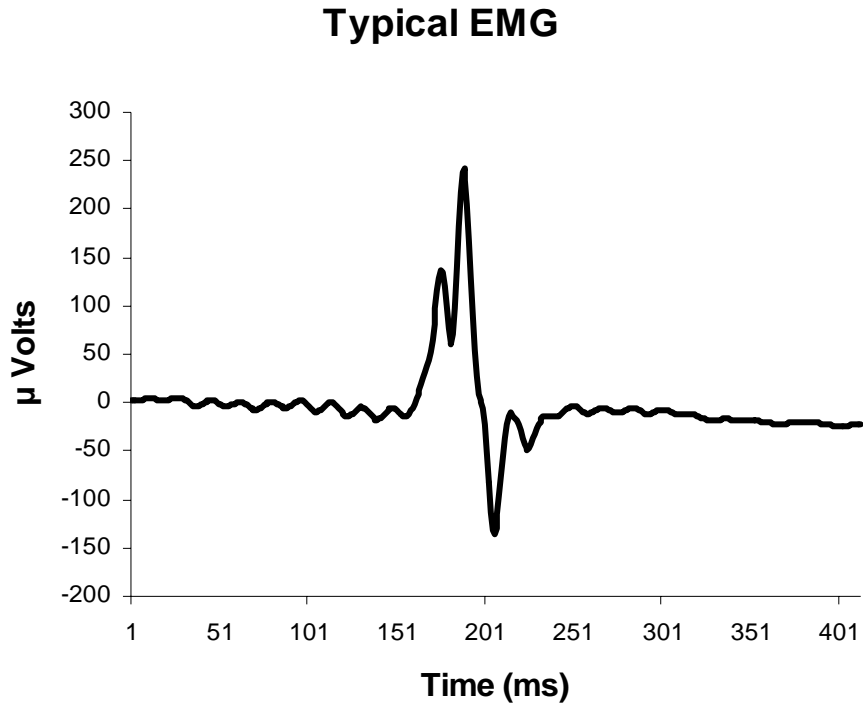


Figure 16. A typical single-trial EMG. For this example, a criterion voltage of approximately 50 μV would reliably dissociate background noise from muscle activity.

to when it began (see Figure 16). For one subject (subj #3), EMG results were contaminated by a tendency to produce a finger extension just prior to finger flexion. This introduced a noticeable yet smaller peak approximately 50 ms before the “main” flexion peak. Unfortunately, these movements were idiosyncratic and occurred on only some trials. For this reason, EMG onset detection was accomplished by hand for this subject. On average, EMG medians were 75.3 ms faster than RTs. To illustrate, Table 3 presents EMG and RT medians as a function of compatibility and SAT condition (for correct trials on which the deadline was met).

Table 3.

Median EMG versus RT and (standard deviation): Correct & Deadline Met						
	Slow (dead = 550)		Med (dead = 450)		Fast (dead = 375)	
	EMG	RT	EMG	RT	EMG	RT
Compatible	347 (85)	422 (51)	295 (62)	371 (45)	248 (57)	319 (48)
Incompatible	389 (94)	467 (52)	307 (73)	387 (55)	237 (58)	309 (53)

Compatible and Incompatible Correct R-LRPs

Figure 17 and 18 present, respectively, grand average R-LRP waveforms for compatible and incompatible trials (see Appendix for subject-level plots). It is immediately obvious that the predicted patterns emerge: The LRP amplitude is higher in the *slow* condition, somewhat lower in the *med* condition, and smallest in the *fast* condition. This pattern emerges for both the compatible and incompatible trials. As well, LRP amplitude at response onset is similar across compatibility conditions for each level of SAT (e.g., the criterion voltage for the *slow* condition is similar for compatible and incompatible trials). A 2 (compatibility) x 3 (SAT) repeated measures ANOVA was conducted to verify these trends. Consistent with what is apparent in Figures 17 and 18, there was a main effect of SAT condition, $F(2,10) = 14.23, p < .01, \text{partial } \eta^2 = .74$. Main effect contrasts revealed that the *med* condition was marginally lower than the *slow* condition $F(1,5) = 5.02, p < .08, \text{partial } \eta^2 = .50$, and the *fast* condition was significantly lower than the *med* condition, $F(1,5) = 29.09, p < .01, \text{partial } \eta^2 = .85$. Consistent with the observation that criterion voltage is similar across compatibility conditions and SAT levels, there was no main effect of compatibility, nor a compatibility x SAT interaction.

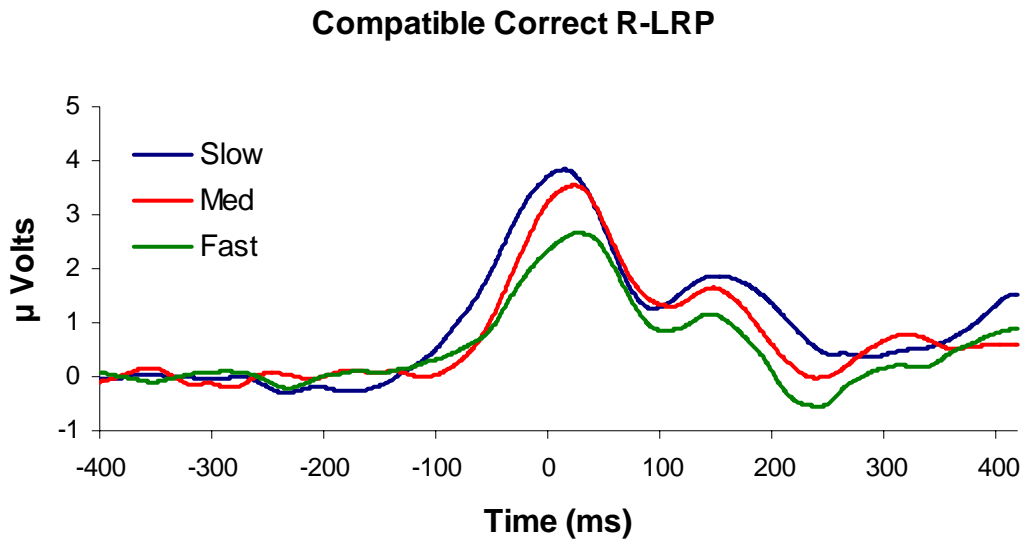


Figure 17. Grand average R-LRP waveforms for compatible, correct trials.

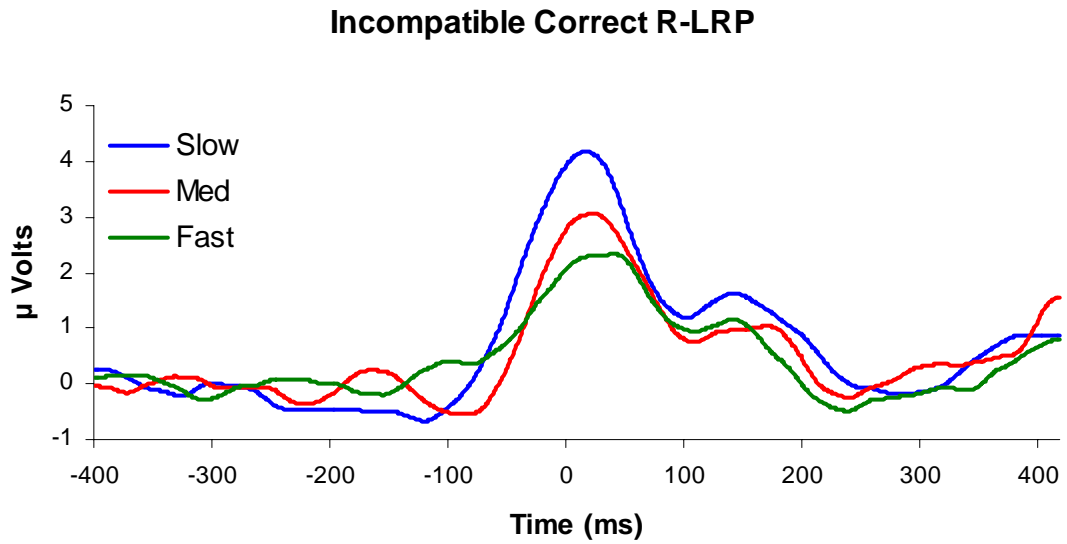


Figure 18. Grand average R-LRP waveforms for incompatible, correct trials.

Compatible and Incompatible Median Splits

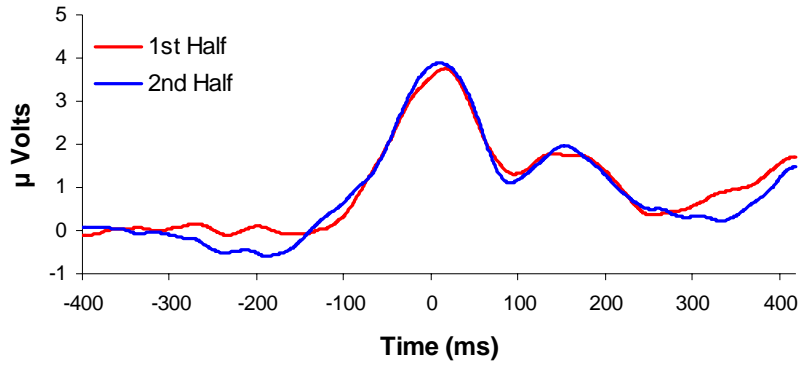
As mentioned in the introduction, previous research has shown a constant LRP threshold *within* a given SAT setting (Gratton et al., 1988; Mordkoff & Grosjean, 2001). To see if this held true, subjects' RTs were median split separately for compatible and incompatible trials for each SAT condition. Figures 19 and 20 present these data. For the *slow* condition, it is quite clear that LRP threshold is identical for the fast and slow halves. Though the *slow* condition was technically speeded in that a deadline was enforced, it may be similar to a condition that did not enforce a deadline. Thus, consistent with previous research (Gratton et al., 1988; Mordkoff & Grosjean, 2001), a constant criterion is observed for both fast and slow trials. However, the story is not so clear when considering the *med* and *fast* SAT conditions. Plots of these data seem to suggest that the threshold is lower for the fast half of the median split. A 2 (compatibility) x 2 (median: first vs. second half) x 3 (SAT) repeated-measures ANOVA revealed that, indeed, there was a marginal main effect of median, $F(1,5) = 5.32, p < .07$, partial $\eta^2 = .52$, and a main effect of SAT, $F(2,10) = 13.75$, partial $\eta^2 = .73$, although no interactions emerged. It should be noted that although the grand averages in Figures 19 and 20 are rather clear, on a subject level, the effect is idiosyncratic. This is at least partially due to the loss in signal-to-noise ratio incurred when performing a median split. It was not uncommon for the *fast* condition to yield only 20-25 observations/hand (sometimes less); it is known that stable LRPs are attained with a minimum of more than double that number (Rinkenauer, et al., 2004). Unsystematic review of the subject-level median split R-LRPs revealed that the first half tended to be “noisier” than the second

half, though the basis for this is unclear.

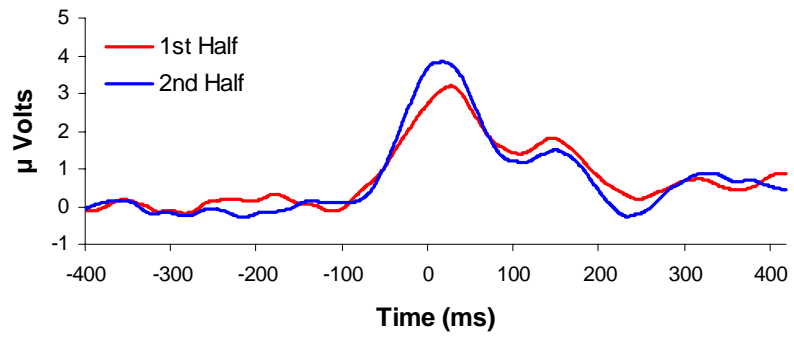
The fact that the second half of trials led to smaller waveforms in the *med* and *fast* conditions is problematic for two major reasons. First, it invalidates a strict version of the constant criterion hypothesis. Rather, there are small shifts in this amplitude due to a trials' latency (i.e., fast half vs. slow half). This may reflect what some have called the "micro-SAT" (Osman et al., 2000; Rinkenauer et al., 2004; Wickelgren, 1977).

Accordingly, response criteria vary widely by SAT ("macro-SAT") condition, but also vary on a trial-by-trial basis due to momentary fluctuations in criteria ("micro-SAT"). Faster trials involve a somewhat lower threshold and slower trials a higher threshold. Second, and more damaging to the central thesis, one could suggest that the pattern depicted in Figures 17 and 18 is an artifact of the introduction of more and more "guess" trials as speed stress increases. Although the CAF data reported above suggest that low accuracy rates were *not* due to random guessing, it remains possible that *some* trials were random (i.e., the response was predetermined), and that these trials contaminated the R-LRPs presented above. Rinkenauer et al. (2004) have developed a sophisticated method by which the influence of such fast guess trials can be removed from the LRP waveforms. One is then left with so-called "subject controlled" LRPs, free from any contamination due to guessing. However, this procedure is based on the "fast guess" model (Ollman, 1966), which the authors themselves admit is wrong (p. 262). The fast guess model holds that subjects produce either slow responses with high accuracy or guess responses at chance accuracy. The SAT results by changing the mixture of subject controlled and guess trials. In a high speed-stress condition, it is thought that a larger

Compatible Correct Slow: Median Split R-LRP



Compatible Correct Med: Median Split R-LRP



Compatible Correct Fast: Median Split R-LRP

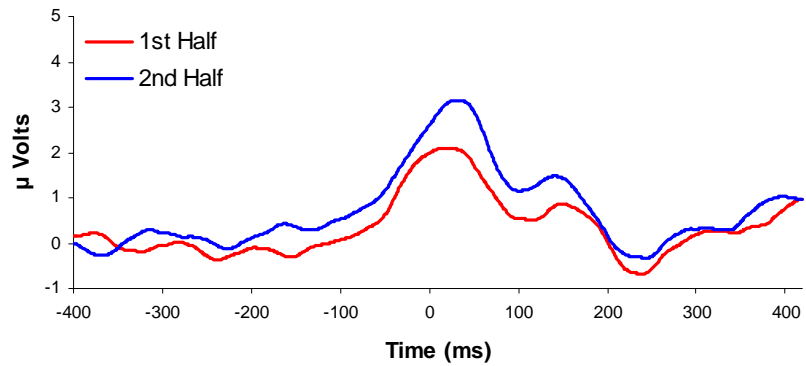


Figure 19. Grand average R-LRP waveforms after median split for compatible trials.

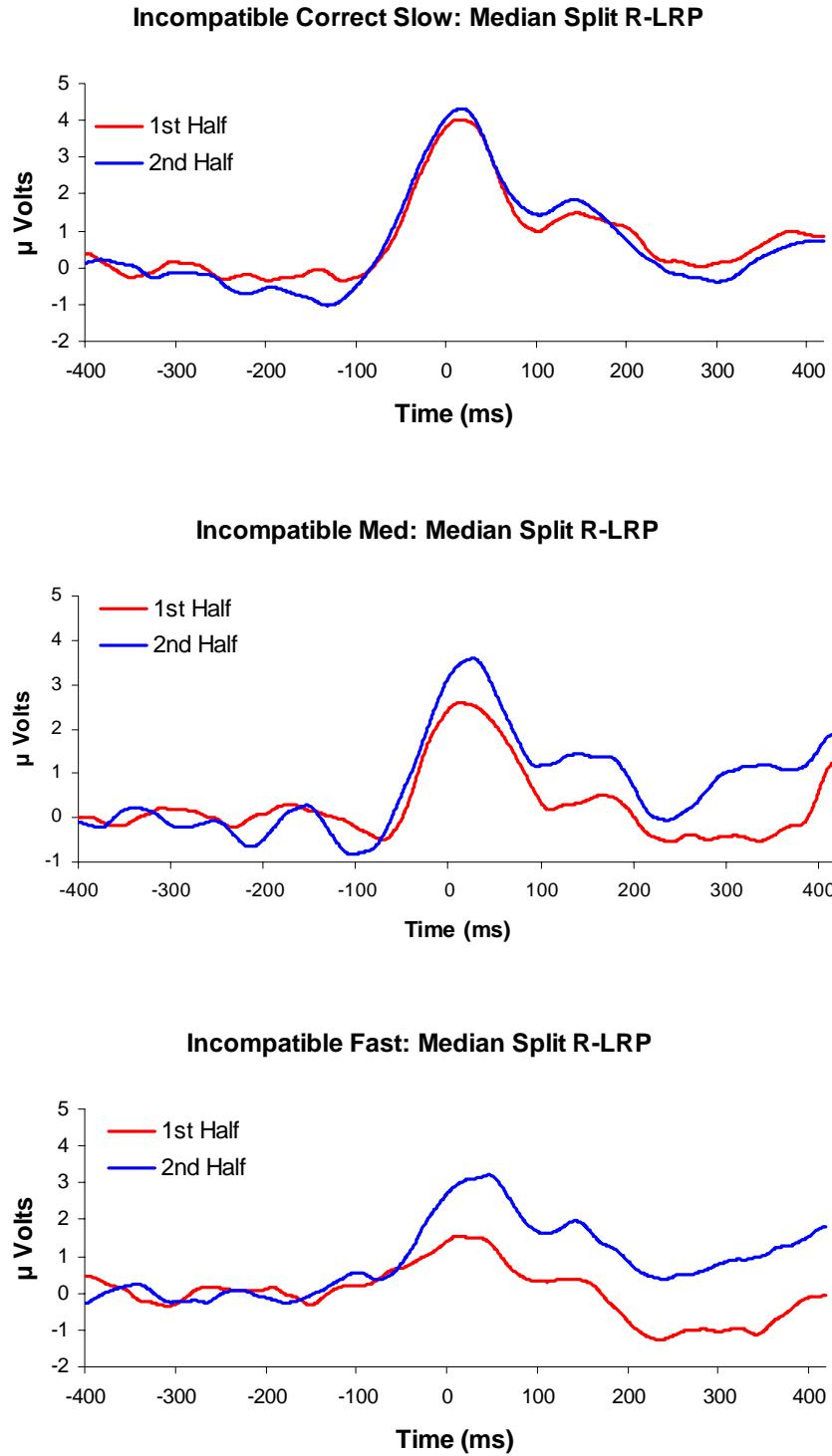


Figure 20. Grand average R-LRP waveforms after median split for incompatible trials.

proportion of trials are guesses, which serve to lower accuracy rate and speed mean RT. Research has shown that this model is untenable. For instance, it cannot accommodate the observation that error RTs are often slower than correct RTs. The most serious problem in using the fast guess model as a basis for LRP correction stems from the necessary assumption that errors are due primarily, if not solely, to guess responses (which will be correct 50% of the time and incorrect 50% of the time). By estimating the LRP waveform elicited by errors (which are by definition guesses), Rinkenauer et al. (2004) argues that the influence of guesses can be removed from the so-called subject controlled waveforms. However, as argued above and elsewhere (Heitz & Engle, in press), error responses need not be guesses – errors are reflective of the type of information accumulated thus far in a trial. It is for that reason that accuracy rates drop below chance on incompatible flanker trials (see Figure 12).

Because of these difficulties, a different approach was used to remove the possible contamination of guess responses. One need only assume that guess responses, when they occur, tend to be faster than non-guess trials. This is a reasonable assumption and is backed by time-course analyses (Heitz & Engle, in press; Gratton et al., 1988). Hence, if we look only at the second half of each median split, they should be relatively uncontaminated from guess trials. Figure 21 depicts the second half for compatible and incompatible trials, respectively. It is evident that the same general pattern emerges when considering only the second half of each median split. Thus, it is unlikely that the overall pattern is due to an increasing contamination due to guess trials. A 2 (compatibility) x 3 (SAT) ANOVA confirmed a

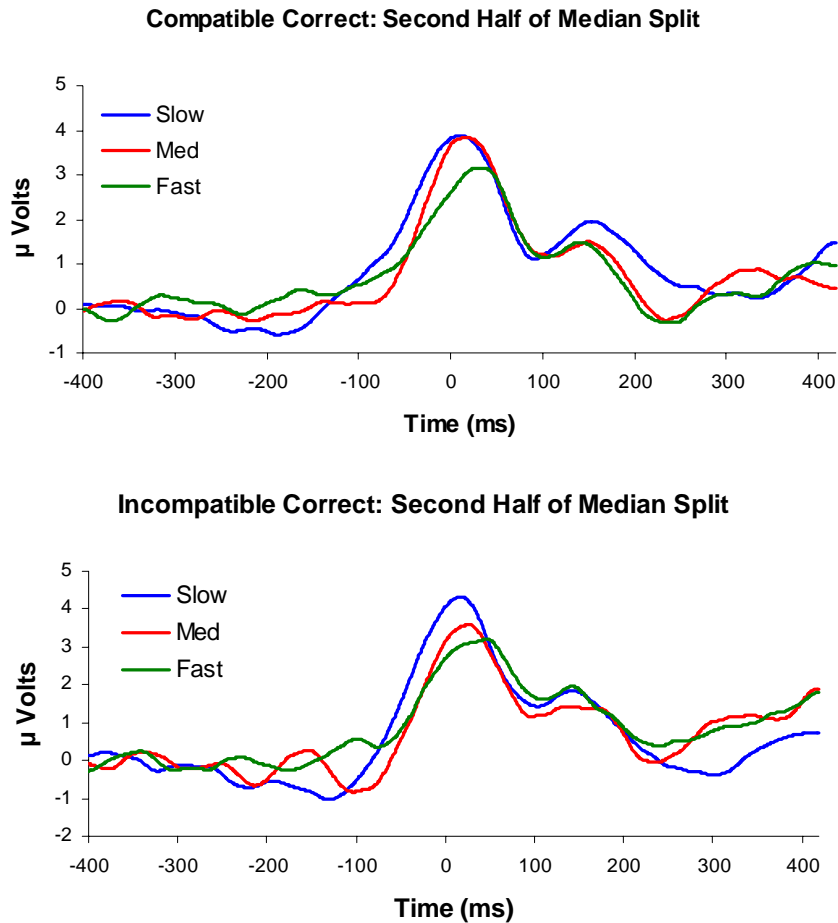


Figure 21. Grand average R-LRP waveforms including only the second (slower) half of a median split on RT. Upper panel depicts compatible trials, lower panel, incompatible trials.

main effect of SAT, $F(2,10) = 4.33, p < .05$, partial $\eta^2 = .46$. As before, there was no main effect or interaction involving compatibility.

Though this alleviates some of the concern for potential confounds in the overall pattern, it does *not* eliminate the possibility that some random guess trials contaminated R-LRPs (particularly, for the first half of each median split). This could be responsible for the lower amplitudes observed in the 1st half of each median split in the *med* and *fast*

blocks. Despite the fact that it is impossible to determine with any certainty which trials were guesses and which trials were simply inaccurate, it is possible to speculate on how they could potentially contaminate the LRP. In a strict sense, it is unlikely that a true guess trial would elicit *no* LRP. Although there may be little cognitive processing behind such trials, they still require a manual response; thus, an LRP *must* be produced. Contamination, however, can result from the combination of two factors: baseline correction of the R-LRP coupled with advance preparation of a response channel.

If we assume that a true guess is produced by pre-selecting a response prior to stimulus onset, it is reasonable to suggest that subjects also prepare that response (hand) prior to stimulus onset. Gratton et al. (1988) presented such evidence: trials that were responded to extremely quickly with chance accuracy showed advance lateralization of the LRP. This becomes problematic in light of the methods used to derive R-LRPs. To derive an R-LRP, one must perform some type of baseline correction. In other words, the beginning of the waveform is set to 0; this corrects for unrelated voltage changes (e.g., electrical noise, skin conductance changes, etc.). Here, the baseline was computed as the average activity 400 to 300 ms *before* response onset. Hence, if guess trials were already pre-lateralized, baseline correcting these waveforms would make them appear artificially small. One way to deal with this would be to use a baseline correction well before the stimulus onset. To do so, R-LRPs for the compatible *fast* condition were re-computed (due to a timing issue, Subject 3 had to be left out of this analysis). Instead of baseline correcting 400 to 300 ms prior to response, a 100 ms period following the onset of the *warning tone* was used. Recall that each trial began with the presentation of a warning

tone, followed 1000 ms later by the critical stimulus. If the smaller R-LRPs in the *fast* condition were due to pre-lateralization, then using this new baseline correct should drastically increase *fast* R-LRPs. Figure 22 shows the results of this analysis.

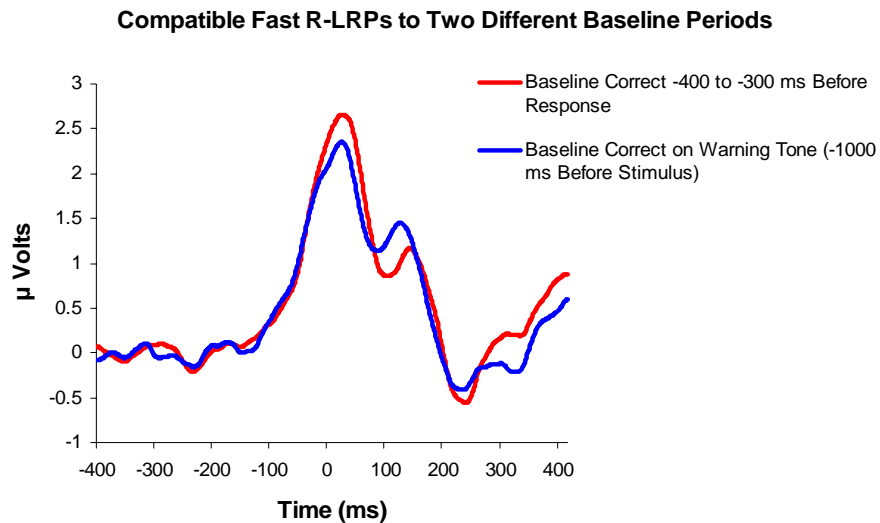


Figure 22. Red line: original baseline correction (-400 to -300 prior to response). Blue line: new baseline period approximately 1000 ms before stimulus onset.

It is evident that the smaller R-LRPs observed in the *fast* condition are not due to pre-lateralization. Had this occurred, then correcting this baseline (Figure 22, blue line) should have led to a much larger R-LRP. A paired *t*-test confirmed that the amplitude at response onset was statistically equivalent for both baseline periods.

Incorrect Responses

The question of what produces an error response is interesting, and one that continues to be addressed by sequential-sampling models. In the diffusion model, an error is produced when the information accumulation function approaches the wrong boundary, perhaps due to noise in the processing system or noise deriving from the

signal. It is often assumed (but not required) that the boundaries for each response alternative are equidistant from the starting point. Accordingly, error trials must be produced when the accumulation function reaches the *same* threshold level, but opposite in sign (i.e., in the wrong direction). Figure 23 and 24 depict the waveforms for compatible and incompatible trials in the *med* and *fast* conditions (there were too few observations in the *slow* condition).

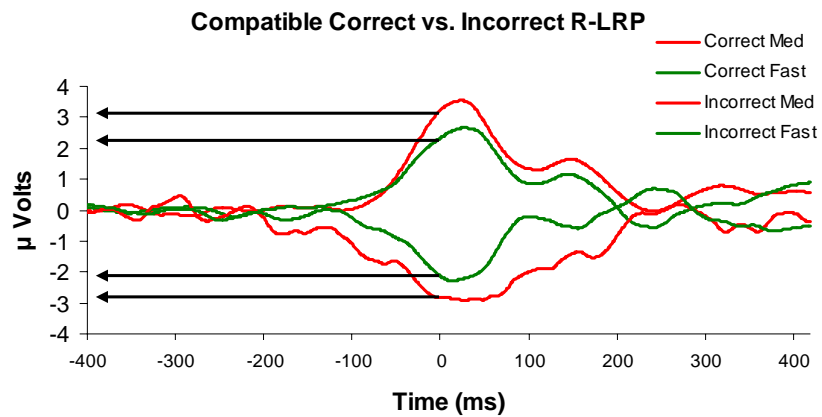


Figure 23. Grand average R-LRP waveforms for the compatible *med* and *fast* conditions. Positive deflections depict correct trials, negative deflections incorrect trials.

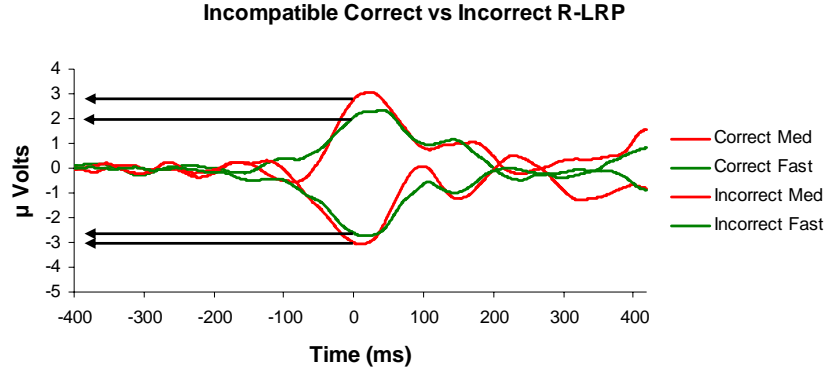


Figure 24. Grand average R-LRP waveforms for the incompatible *med* and *fast* conditions. Positive deflections depict correct trials, negative deflections incorrect trials.

As Figures 23 and 24 show, the voltages at response onset are remarkably similar for correct and incorrect trials, but flipped in sign. Considering first compatible trials, the mean respective correct and incorrect voltages were as follows: *med* (correct = 3.5 μV ; incorrect = -3.0 μV), *fast* (correct = 2.3 μV ; incorrect = -2.4 μV). Similar patterns emerge for incompatible trials: *med* (correct = 3.3 μV ; incorrect = -3.2 μV), *fast* (correct = 2.4 μV ; incorrect = -2.9 μV). Consistent with these observations, all *t*-tests comparing correct and incorrect voltages were non-significant.

Missed Deadlines

To restate the present model, the onset of a critical stimulus begins an evidence accumulation process. This accumulation process continues until it reaches some critical threshold, at which point a response is emitted. As we have shown above, this threshold is related to SAT condition and hence, the amount of evidence accumulated before making a response. One critical prediction following this model concerns missed deadlines. Particularly, it should be the case that missed deadlines rise to the same

threshold as made deadlines, only later in time. Stated differently, the reason one misses a deadline is due to the fact that the information function reached threshold *after* the response deadline. This prediction can be addressed using a combination of R- and S-locked LRPs. Because subjects missed deadlines on only a fraction of trials (and this fraction had to be split between right and left hands to derive an LRP), the only conditions that could be analyzed were correct trials from the incompatible condition. Figures 25, 26, and 27 present these data for the *slow*, *med*, and *fast* conditions, respectively.

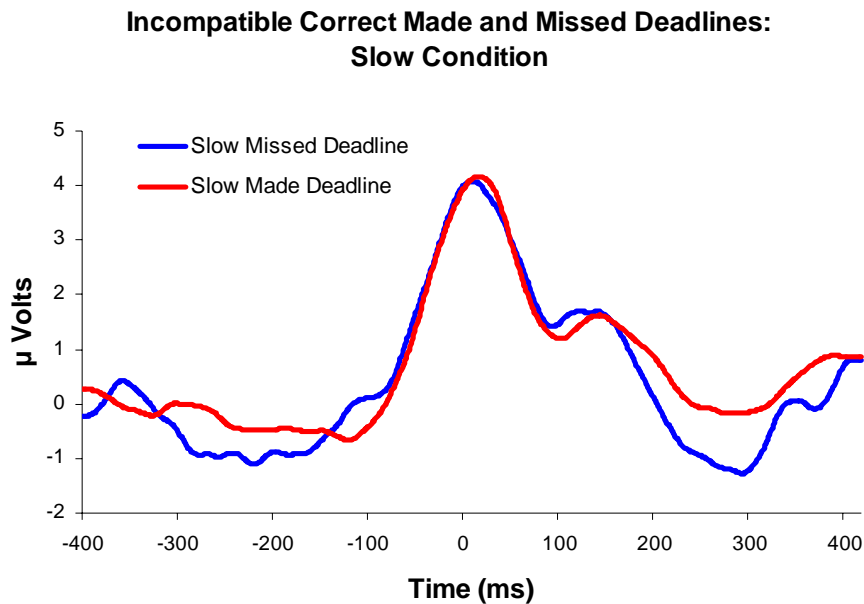


Figure 25. Grand average R-LRPs for correct incompatible trials in the *slow* condition.

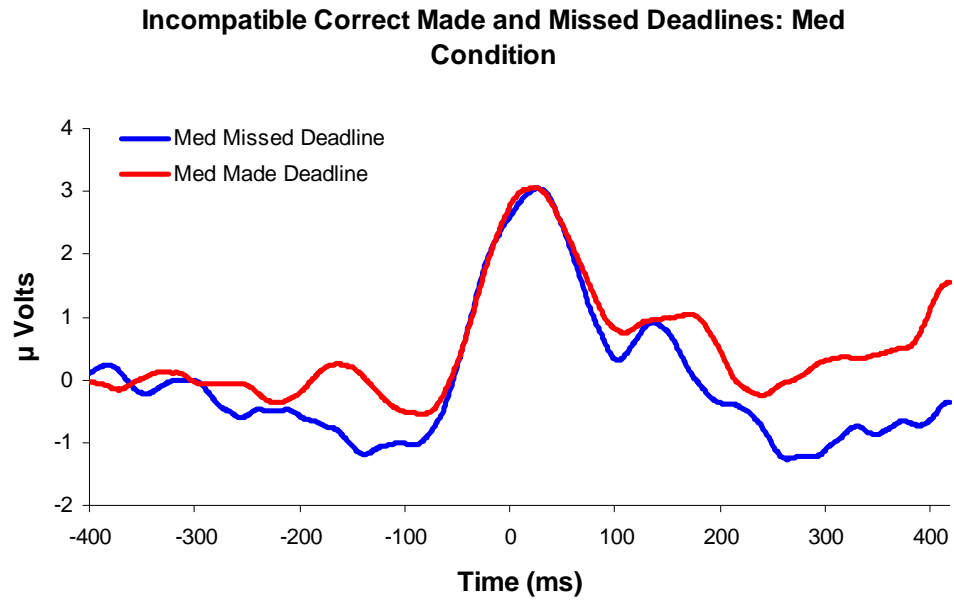


Figure 26. Grand average R-LRPs for correct incompatible trials in the *med* condition.

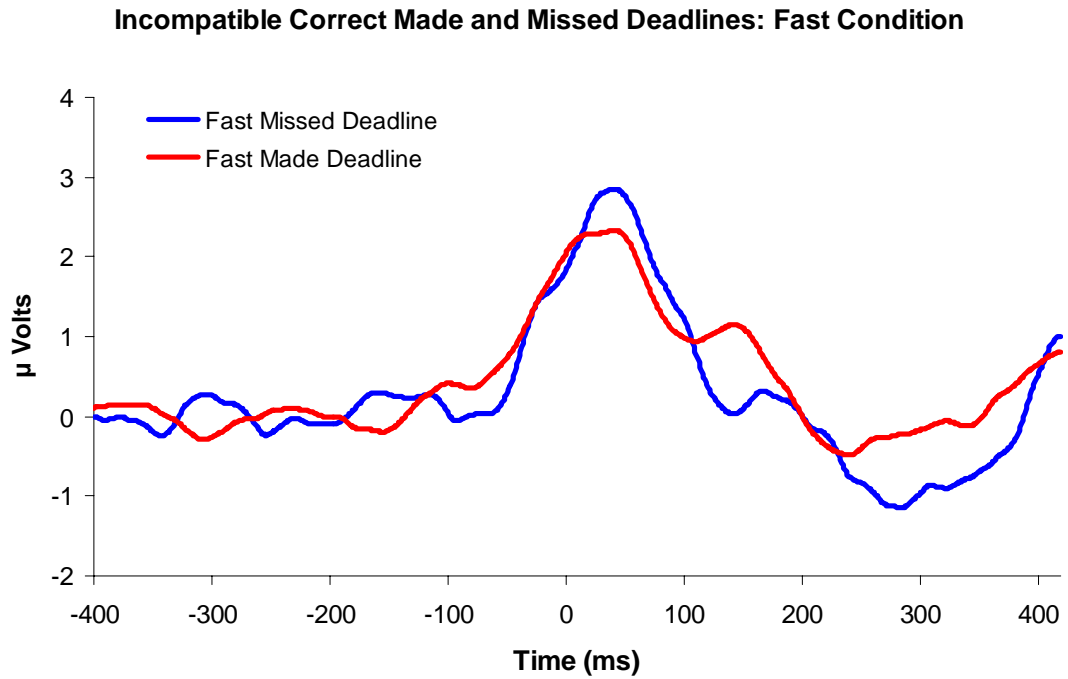


Figure 27. Grand average R-LRPs for correct incompatible trials in the *med* condition.

What should be evident in Figures 25-27 is that, consistent with predictions, both made and missed deadlines have identical thresholds for response. This was confirmed statistically in a 2 (2: deadline: made vs. missed) x 3 (SAT) repeated measures ANOVA. Though there was the expected decline in amplitude due to SAT condition, $F(1,10) = 25.4, p < .001$, partial $\eta^2 = .84$, no other effects attained significance (all F s < 1.0). These data suggest that subjects miss a deadline because evidence accumulation reaches threshold at a time point after the deadline has passed.

To further show how missed deadlines were delayed in time relative to made deadlines, Figure 28 presents S-LRPs for incompatible correct trials. S-LRPs are useful for assessing timing issues (i.e., at what point did the LRP begin to rise?), but are less amenable to hypotheses regarding voltage at response onset. The reason for this, to put it succinctly, is that the voltages associated with R-LRPs (which we have just dealt with at length above) become “smeared” due to averaging when they are locked to stimulus onset. Thus, the resulting S-LRP need not resemble the R-LRP for the same condition. Despite these inadequacies, S-LRPs presented in Figure 28 exhibit interesting characteristics. Two characteristics of these plots are noteworthy. First, the S-LRP for missed deadlines is much delayed with respect to made deadlines. Second, missed deadlines exhibit a prominent decrease in voltage prior to peak amplitude. This is often observed when subjects begin to prepare the incorrect hand, then recover, and produce the correct response (Gratton et al., 1988). This would be expected on incompatible trials, as the distractor letters may, on occasion, exert enough influence that the incorrect response channel is primed, causing a delay in response.

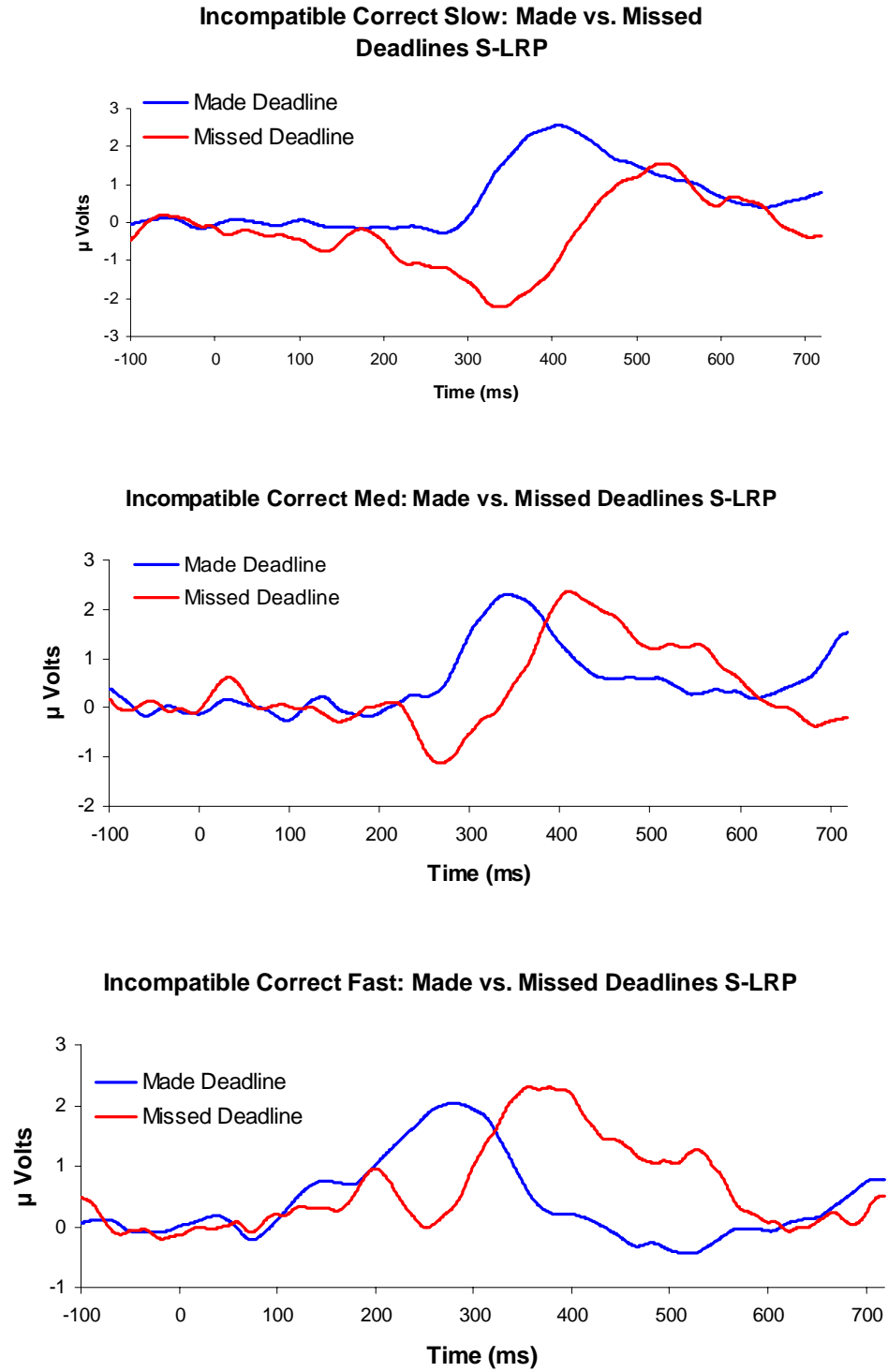


Figure 28. S-LRPs for made and missed deadlines in the *fast* (top panel), *med* (middlepanel), and *fast* (bottom panel) conditions.

Discussion

The rich dataset just described may be boiled down to a few critical points. First, the behavioral data are abundantly clear that the deadline manipulation did, in fact, affect subjects' response criterion. Not only did mean RT and accuracy rate decline as speed stress increased, but CAFs showed that accuracy rates, holding RT constant, were dependent on SAT condition. Furthermore, RT data were well described by a "swivel" operation of the recipit. In accord with the LATER model (Carpenter & Williams, 1995), this is representative of a change in response criteria.

Second, and most critically for the present work, the amplitude of the LRP at EMG onset was related to SAT condition. Consistent with predictions, the largest amplitude was observed for the *slow* condition, intermediate amplitude for the *med* condition, and the smallest amplitude for the *fast* condition. Not only did this pattern emerge for both compatible and incompatible trials, the threshold values were comparable across compatibility conditions. One possible alternative explanation for the lowering of LRP threshold with SAT, as mentioned, is the introduction of guess trials. To deal with this, the slow half of each median split was compared across the 3 SAT and 2 compatibility conditions with the assumption that when a guess occurs, it will be faster than non-guess responses (Ollman, 1966; see also Gratton et al., 1988). The overall pattern was unchanged.

As discussed in the introduction, these observations constitute an important step in linking psychophysiology with behavior, as well as establishing that sequential-sampling models are biologically plausible (Smith & Ratcliff, 2004; Bogacz, 2007; Lo & Wang, 2006). To reiterate, neural correlates have been found for drift rate (e.g., Roitman & Shadlen, 2002) and for prior probability (e.g., Basso & Wurtz, 1997, 1998; Dorris & Munoz, 1998), but until now, not for criterion placement and the role it plays in behavior. As predicted, responses were produced when the LRP reached a criterion voltage. Whether a response was made in error or not was determined by the valence of the LRP, but not the threshold (i.e., it was equivalent for correct and error trials). The deadline manipulation, known to affect subjects criterion placement, did in fact decrease criterion voltage needed to produce a response. This is in complete agreement with the computational models described earlier.

Third, the data call into question a strict version of the constant criterion hypothesis (Gratton et al., 1988; see also Mordkoff & Grosjean, 2001). Recall that these researchers have shown that the amplitude of the LRP at response onset is unrelated to response latency, indicative of a constant threshold. It is interesting to note that data supporting the constant criterion hypothesis come from experiments where RT was not specifically controlled. To the contrary, subjects in these tasks likely used a “be fast/be accurate” criterion setting, usually producing intermediate RTs and off-ceiling accuracy rates. Compare this situation to the *slow* condition presented here. It is possible that behavior in this block of trials is roughly comparable to “be fast/be accurate” instructions. Like such previous research, the present data also showed a constant criterion for the first

and second halves of a median split, but only for the *slow* condition. In contrast, R-LRP amplitude tended to be smaller for the fast half of trials in the *med* and *fast* conditions. One explanation for this, touched upon earlier, is the “micro-SAT”. According to some (e.g., Osman et al., 2000; Rinke­nauer et al., 2004), response criteria are not fixed in a strong sense; rather, they have an average value. On trials where the criterion happens to be a bit lower than average, faster responses result. Conversely, when the criterion happens to be above average, slower trials result. While this would explain the grand averages presented in Figures 19 and 20, it does not explain why the effect is idiosyncratic, appearing in some subjects/conditions, reversing in others. One contributing factor could be the reduction in the signal-to-noise ratio (SNR) incurred when performing a median split. Because the critical data concern only correct trials, it was generally the case that the *slow* condition provided many more observations than the *med* condition, and certainly more than the *fast* condition. Thus, the SNR tends to decrease with SAT condition. This is not problematic for the grand average waveforms, given that LRPs are known to be stable at about 60 trials/hand, and most conditions approached if not exceeded this number. In computing the median split, however, it was often the case that the number of contributing observations was as low as 25. The fact that the first half was smaller than the second half sometimes, but not always, could be due to this noise, although this is highly speculative.

Another important outcome of this experiment concerns electrophysiological activity on trials where the deadline was missed. Under the present model, a response is produced when the LRP reaches a criterion voltage. Thus, it must be true that missed

deadlines, when they occur, are marked by an LRP that reaches this criterion voltage at a later time than made deadline trials. Though it is not surprising that missed deadline R-LRPs are later than made deadlines (they occur later by definition), it is important to note that the criterion voltage *is* the same for both. Comparable data is evident in other work. Shin, Fabiani, & Gratton (2004) observed subthreshold LRP activity even on trials where a response was not produced. This suggests that responses will only occur when critical threshold is reached. In other work, Hanes and Schall (1996) recorded FEF neurons during a stop-signal task. They estimated response threshold during no-signal trials (in which subjects simply make an eye movement toward a target). Then, they examined neural activity to stop-signal trials. They found that neural activity during successfully inhibited trials did not reach this critical threshold, while unsuccessfully inhibited trials did cross this threshold. This is interesting in light of the S-LRP data presented in Figure 28. These waveforms suggest that missed deadline trials are marked by an initial lateralization towards the incorrect response. The time required for subsequent recovery from this produces a delay in reaching criterion threshold.

Baseline Versus Criterion

The LATER model, and other sequential sampling models, posit that SAT settings either lengthen or shorten the distance that an information function must travel before reaching threshold; hence the variability in RT and accuracy rate. It is interesting to note that, at least in the LATER model, these data are equally well accommodated by a model in which the *baseline* shifts. In other words, the LATER model cares only about the distance from baseline to threshold. Whether the effect of SAT is to reduce threshold

or increase baseline is not separable (it is, however, separable in more sophisticated models).

At least one study has made the argument that the effect of SAT is to change baseline, not threshold. Although this is contrary to many current and past models, it cannot be discounted as a viable alternative. Van Veen (2006) measured neural activity using fMRI while subjects performed a modified Simon task (Simon, 1969). Subjects were told that upon onset of a red square, they should press a key with their left hand; upon onset of a green square, they were to push a button with their right hand. Critically, these squares could appear in two locations: either to the left or to the right of center. The typical Simon effect is exhibited by a slowing of responses to “respond left” when presented on the right, and to “response right” stimuli, when presented on the left, than when the response mapping and screen locations coincide. Subjects performed this task under two levels of speed stress (be fast vs. be accurate), cued prior to each trial (i.e., trials were not blocked according to SAT condition). Van Veen found areas of premotor cortex (PMC), supplementary motor area (SMA), and lateral prefrontal cortex (PFC) that were more active in the high speed stress condition. He argued, based on this that subjects attempt to reduce the “distance” between baseline and threshold by modifying the former rather than the latter. There are a number of reasons to question Van Veen’s conclusion. First, due to its poor temporal resolution, fMRI can not be used to measure neural activity at the precise moment a response is produced; if SAT instructions did interact with criterion, it could not be observed. Second, it is possible that in response to speed instructions, subjects prime *both* hands. This could lead to greater neural

activation as evidenced by fMRI; however, the present model (and others) argues that it is the difference in activity that constitutes criterion. Thus, priming both response channels should not matter (i.e., one channel would still have to beat out the other channel by a criterion amount). Still, this is not incompatible with so-called *race* models, which posit two independent counters racing towards independent criteria. In this case, priming both response channels would have a facilitating effect. Further work will be required to assess the viability of a changing baseline model.

One potential alternative explanation to Experiment 1 concerns the confounding of block order with practice. It is not altogether clear what effect practice has on the LRP, and particularly, the amplitude of the LRP at response onset. Though some researchers claim that the amplitude of the LRP is resistant to practice effects even after significant practice (Gratton, personal communication, 6 February, 2007), this claim has not been put to systematic test. The potential problem this raises for the current work is straightforward. Recall that the order of presentation was, for all subjects, *slow*, *med*, then *fast*. Hence, if practice serves to reduce the peak amplitude of the LRP, then the pattern observed in Experiment 1 could be artifactual. As well, it is possible that some other aspect of block order led to the observations detailed above. Perhaps subjects exert less effort, pay less attention, or even give up later in the task. Each of these could lead to a decrease in R-LRP amplitude. To deal with these possibilities, Experiment 2 presented the blocks in reverse order (i.e., *fast*, *med*, then *slow*). Also, to reduce potential practice effects, Experiment 2 employed the same subjects from Experiment 1, who at this point had 3 hours of training on this task. Additionally, subjects were provided with

3 practice blocks at each level of speed stress prior to beginning the experiment proper.

If the results of Experiment 1 were due solely to block order, then Experiment 2 should exhibit a reversal in the R-LRP orderings. In other words, the largest to smallest amplitude should be, respectively, *fast, med, slow*. Alternatively, if the results of Experiment 1 were solely a practice effect, then one would expect to see no difference in R-LRP amplitude at response onset across the 3 SAT conditions. To put it simply, they will all rise to the same point.

The basis for this confound in Experiment 1 was the possibility that following a *fast* set, subjects would be less able to revert to a slower SAT setting. If this occurs in Experiment 2, it will be readily observable in the behavioral data: all RTs and accuracy rates should be similar to the *fast* condition (i.e., there will be no speed-accuracy tradeoff).

CHAPTER 3

EXPERIMENT 2

Method

Participants

The 6 individuals who participated in Experiment 1 returned. As before, they were paid \$10/hour in a single 3-hour session. The latency between sessions was variable, but less than 4 weeks for all subjects.

Flanker Task

The task was identical to that used in Experiment 1 with the following exceptions. First, the order of blocks was reversed. Subjects began with the *fast* block, then performed the *med* and *slow* blocks. Prior to beginning the experiment, subjects also completed 3 practice blocks: one at each deadline. Each practice block consisted of 40 trials, half of which were compatible and half of which were incompatible.

Electrophysiological Recording

All instruments and recording parameters were identical to Experiment 1.

Results

Behavioral Data

Data were first structured to support a time-course analysis. 10 Vincentized *ntile* bins were created on each subjects' RTs separately for compatible and incompatible trials. The analyses to be reported below were run in the context of a 2 (compatibility: compatible vs. incompatible) x 3 (SAT: fast, medium, slow) x 10 (RT bin) repeated-measures ANOVA. Huynh-Feldt corrections were employed for violations of sphericity.

The speed-accuracy tradeoff was again apparent in the behavioral data (i.e., RT and accuracy rate) as reflected in Table 4. Subject-level data are presented in Tables A3 and A4. For RT, there was a main effect of SAT condition, $F(2,10) = 61.36, p < .001$,

Table 4.

Mean RT (ms) and Accuracy Rate by Compatibility and SAT				
Condition		Slow	Med	Fast
Compatible	RT	427	381	329
	ACC	.96	.89	.77
Incompatible	RT	461	400	336
	ACC	.86	.66	.54

partial $\eta^2 = .93$, and a main effect of compatibility (i.e., the “flanker” effect), $F(1,10) = 68.91, p < .001$, partial $\eta^2 = .93$. Repeated contrasts revealed that the *fast* condition was significantly faster than the *med* condition, $F(1,5) = 44.43, p < .01$, partial $\eta^2 = .90$, and the *med* condition was significantly faster than the *slow* condition, $F(1,5) = 35.98, p < .001$, partial $\eta^2 = .88$. As Table 4 suggests, the “flanker” effect tended to diminish as speed stress increased. This is supported by a compatibility x SAT interaction, $F(2,10) = 4.81, p < .05$, partial $\eta^2 = .49$.

Because each subject in Experiment 2 was the same as that for Experiment 1, an additional within-subjects factor of “session” was included in a 2 (session) x 2 (compatibility) x 3 (SAT) x 10 (latency bin) ANOVA. There was no main effect of session, but there did emerge a significant session x compatibility x bin interaction, $F(9,45) = 5.47, p < .05$, partial $\eta^2 = .52$. RTs for incompatible trials tended to be lower in the second session, but only at later latency bins. Stated differently, the flanker effect in RT was decreased in session 2 when considering only slower, overall, responses.

Similar patterns were evident for accuracy rate. A main effect of SAT condition

indicated poorer performance as speed stress increased, $F(2,10) = 26.34, p < .001$, partial $\eta^2 = .84$. The flanker effect, reflected by higher accuracy rate to compatible than incompatible trials, was supported by a main effect of compatibility, $F(1,10) = 20.42, p < .01$, partial $\eta^2 = .80$. As shown in Table 4, the decrease in accuracy rate was stronger for incompatible trials than compatible trials. Accordingly, the strength of the flanker effect increased with speed stress, as indicated by a compatibility x SAT interaction, $F(2,10) = 5.00, p < .05$, partial $\eta^2 = .50$. Note that the flanker effect results are in opposite for RT and accuracy rate: with increasing speed stress, the flanker effect diminished for RT and increased for accuracy rate.

As with RT, session was added as an additional within-subjects factor. This analysis revealed that the flanker effect on accuracy rate was somewhat lower in session 2, again only considering slower overall responses. This was supported by a session x compatibility x bin interaction, $F(9,45) = 2.37, p < .05$, partial $\eta^2 = .32$.

Latency Distributions

Figures 29 and 30 present latency distributions for compatible and incompatible trials, respectively, as a function of SAT condition. Observations < 150 ms were discarded from analyses as anticipations. Doing so eliminated less than .002% of the data. The upper panels depict correct trial latencies, the bottom panels error latencies. As well, Table 5 presents the proportion of missed deadlines for each compatibility and SAT condition along with median RT.

Table 5.

% Missed Deadlines and Median RT (ms) for Made and Missed Trials: Experiment 2						
Condition	Slow (dead = 550)		Med (dead = 450)		Fast (dead = 375)	
Compatible	3.6%		11.2%		19.7%	
	418	586	368	475	314	404
Incompatible	8.3%		19.7%		25.5%	
	450	586	381	482	313	409

Conditional Accuracy Functions

CAFs for Experiment 2 are depicted in Figures 31 and 32. These CAFs indicate that the SAT manipulation did alter subjects' response criteria: accuracy rates are dependent on SAT condition, even when holding RT constant. Because the latency bin means were not statistically equivalent between the SAT conditions, ANOVA is again inappropriate. To verify that the CAFs vary by SAT condition, the Friedman nonparametric test was used as a surrogate for a one-way repeated-measures ANOVA. For compatible trials, the first test considered *slow* bin 7, *med* bin 9, and *fast* bin 10 (latencies between 440 and 442 ms). This test was significant, $\chi^2(2) = 10.00, p < .01$. A second test considered RTs between 374-388 ms (*slow* bin 2, *med* bin 6, and *fast* bin 9). This test, too, was significant $\chi^2(2) = 7.64, p < .05$. Similar tests were run on the incompatible CAFs. The first test on incompatible trials considered *slow* bin 6, *med* bin 9, and *fast* bin 10, while a second test considered *slow* bin 2, *med* bin 5, and *fast* bin 9. The first test was significant, $\chi^2(2) = 6.87, p < .05$, but the second was not, $\chi^2(2) = 3.00, ns$. In comparing Figure 32 with Figure 12, it appears that the *med* and *fast* conditions are somewhat more similar in Experiment 2 than in Experiment 1, particularly

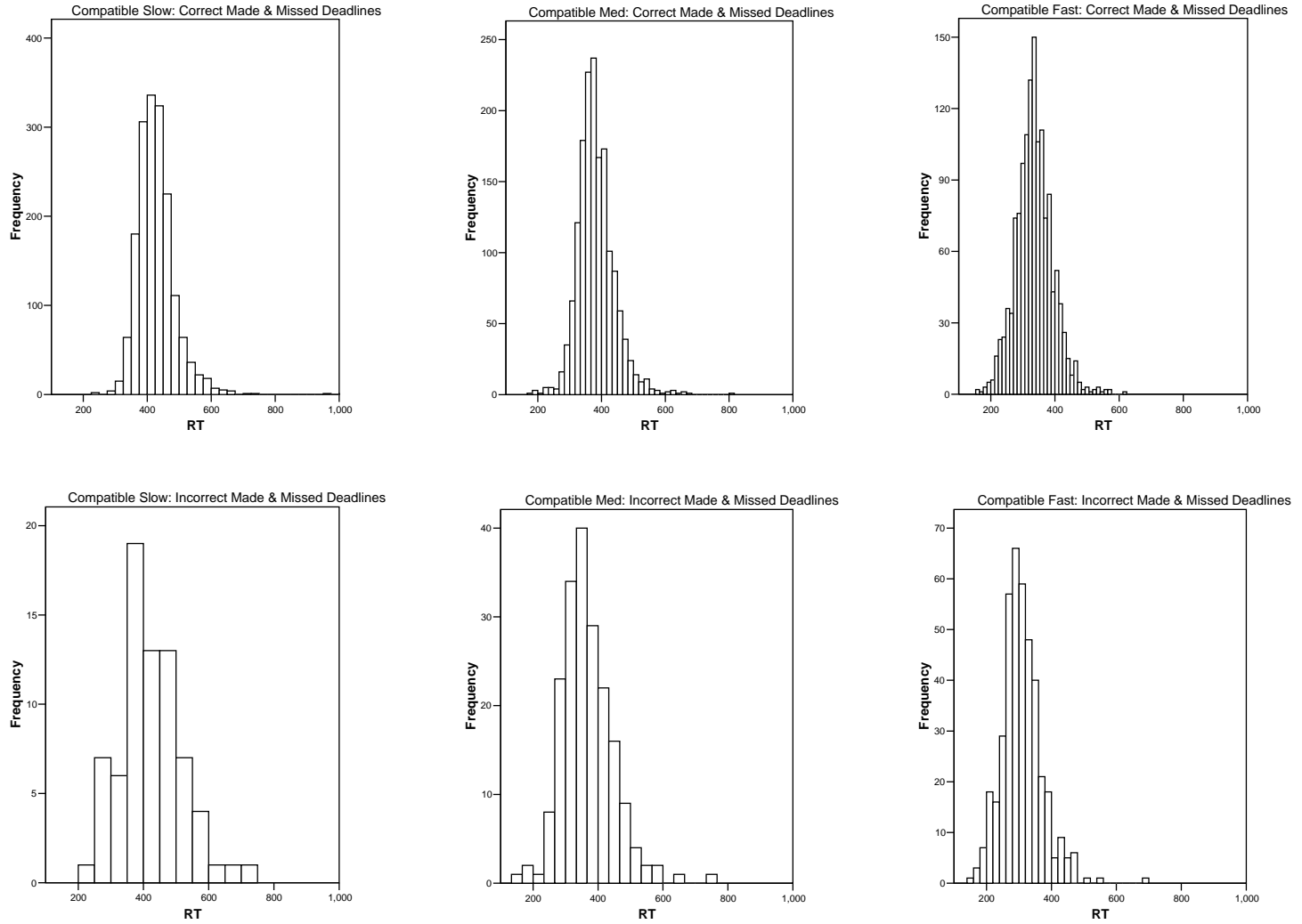


Figure 29. Latency distributions for compatible trials. Upper panel: correct trial latencies. Lower panel: incorrect trial latencies.

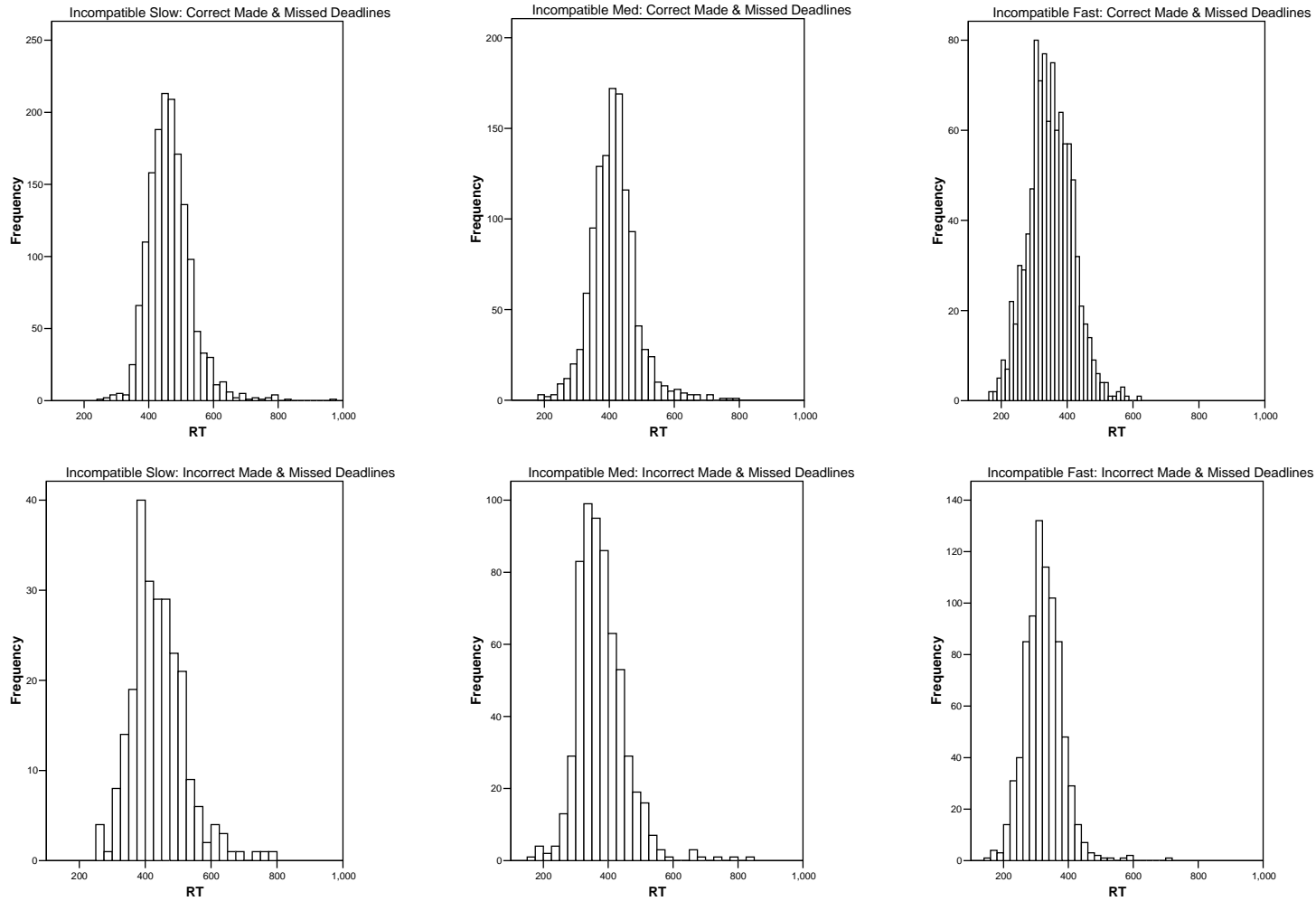
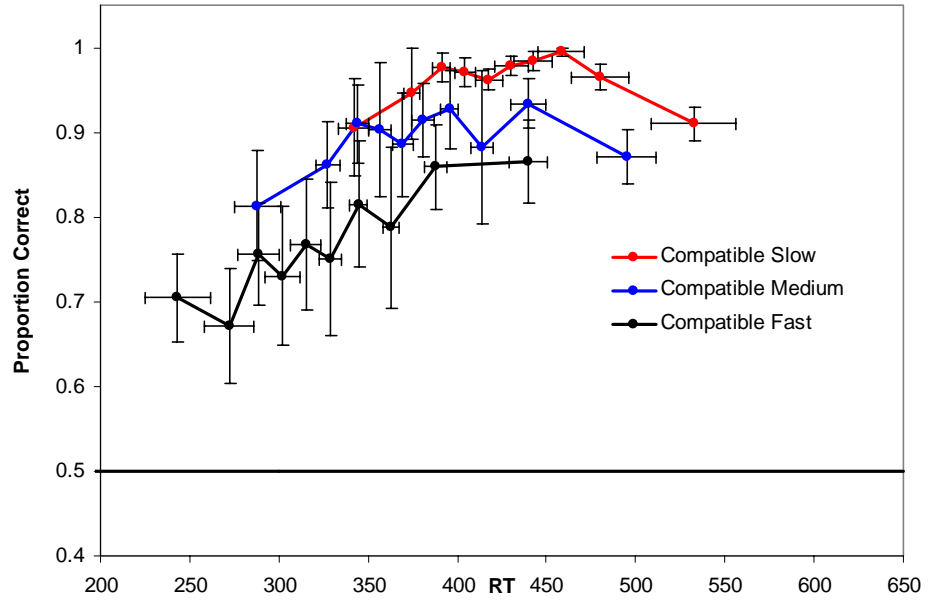


Figure 30. Latency distributions for incompatible trials. Upper panel: correct trial latencies. Lower panel: incorrect trial latencies.

Compatible: SAT x RT



Compatible: Observed RT by Vincitized RT Bin

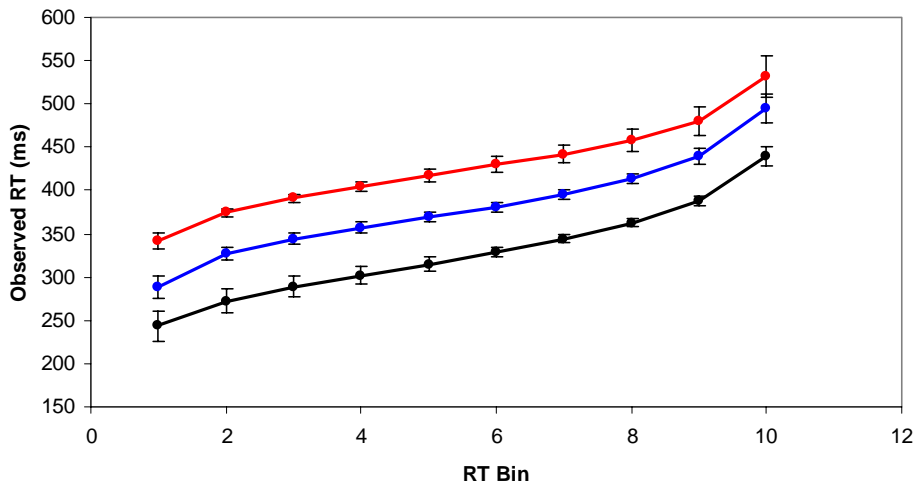


Figure 31. Upper panel: Compatible CAFs for Experiment 2. Horizontal bars depict ± 1 s.e.m for RT; vertical bars depict the same for accuracy rate. Lower panel: observed RTs by *n*tile bin. Vertical bars represent ± 1 s.e.m.

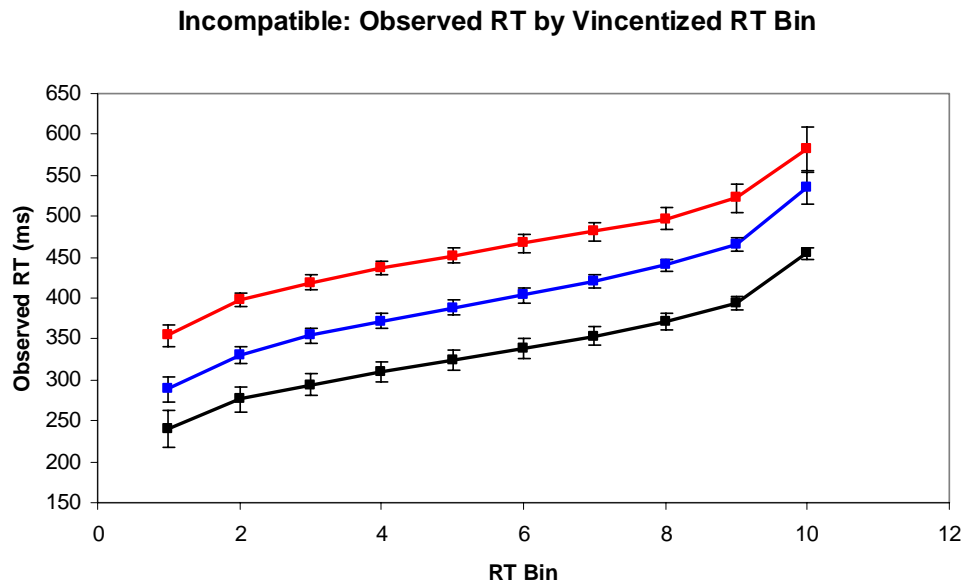
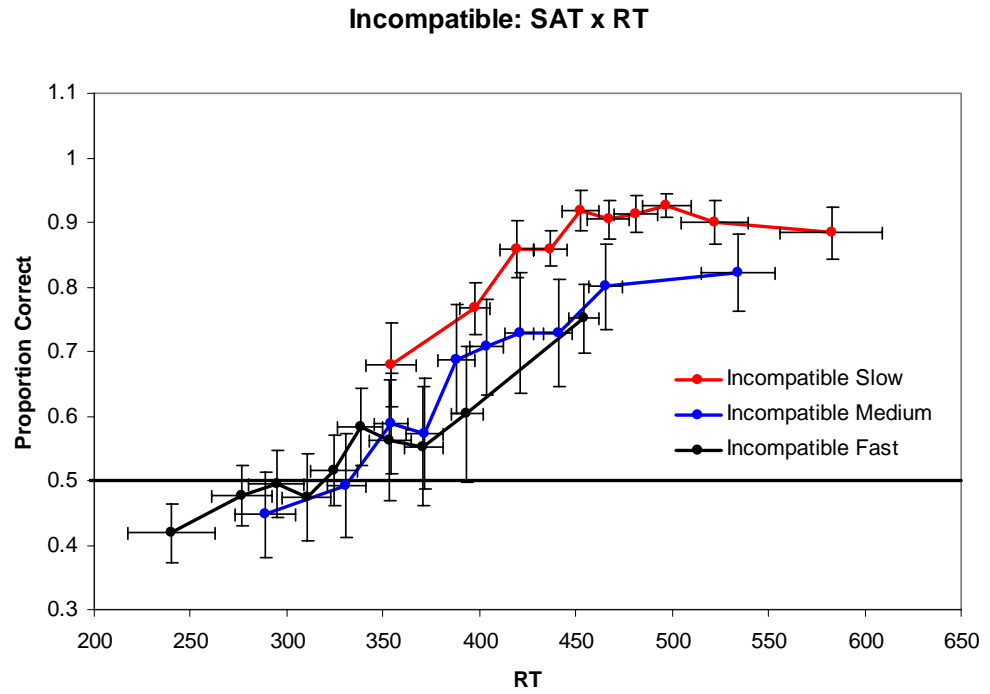


Figure 32. Upper panel: Incompatible CAFs for Experiment 2. Horizontal bars depict ± 1 s.e.m for RT; vertical bars depict the same for accuracy rate. Lower panel: observed RTs by *ntile* bin. Vertical bars represent ± 1 s.e.m.

at earlier time points. Another difference between Experiments 1 and 2, concerning incompatible trials, is the “dip” below chance performance. Despite the fact that mean accuracy was below .50 for the first bin in the *med* and *fast* conditions, this could not be confirmed statistically.

To summarize thus far, mean comparisons and time-course analyses confirmed that subjects’ response criteria were changing due to the deadline manipulation. It must be noted, however, that accuracy rates for the incompatible *med* and *fast* conditions were more similar in Experiment 2 than in Experiment 1 at early time points. We can conclude that subjects did not get “stuck” in a *fast* set, as mean RT and accuracy rates moved in the same direction as in Experiment 1.

LATER Model

Behavioral data for correct trials were fit to the LATER sequential sampling model using SPIC software (Carpenter, 2007). Figures 33 and 34 present these fits. Tested specifically was a “swivel” versus a “shift” model using a log likelihood ratio test. Recall that in the LATER model, manipulations that affect criterion will lead to a “swiveling” of lines about a common intercept when plotted on reciprobbit. At the same time, manipulations that affect drift rate lead to a horizontal shifting of lines. The log likelihood ratio test favored a “swivel” model for both compatible, $LLR = 15131, p < .01$ and incompatible trials, $LLR = 12703, p < .01$. Subject level tests are provided in Figure A3 and A4. All were best described by a “swivel” model with one exception: a “shift” model was favored for Subject 2 in the compatible condition.

As a whole the behavioral data are consistent with a shifting criterion due to the

enforced response deadlines. First, mean RT and accuracy rate decreased as speed-stress increased. Second, CAFs showed that accuracy rates were dependent on SAT condition, holding RT constant. Finally, fits to the LATER sequential-sampling model were consistent with a shifting criterion between conditions. We can thus be confident that the deadline manipulation had the expected behavioral effects.

Electrophysiological Data

LRPs were derived in the same manner as in Experiment 1. Again, certain conditions did not yield enough observations to support stable LRPs. These mainly concerned incorrect compatible trials and compatible trials with missed deadlines. Data for these conditions will not be reported. All waveforms depicted below include only *made* deadlines, unless otherwise noted. LRPs were computed so that correct activation led to a positive deflection, while incorrect activity produced a negative deflection.

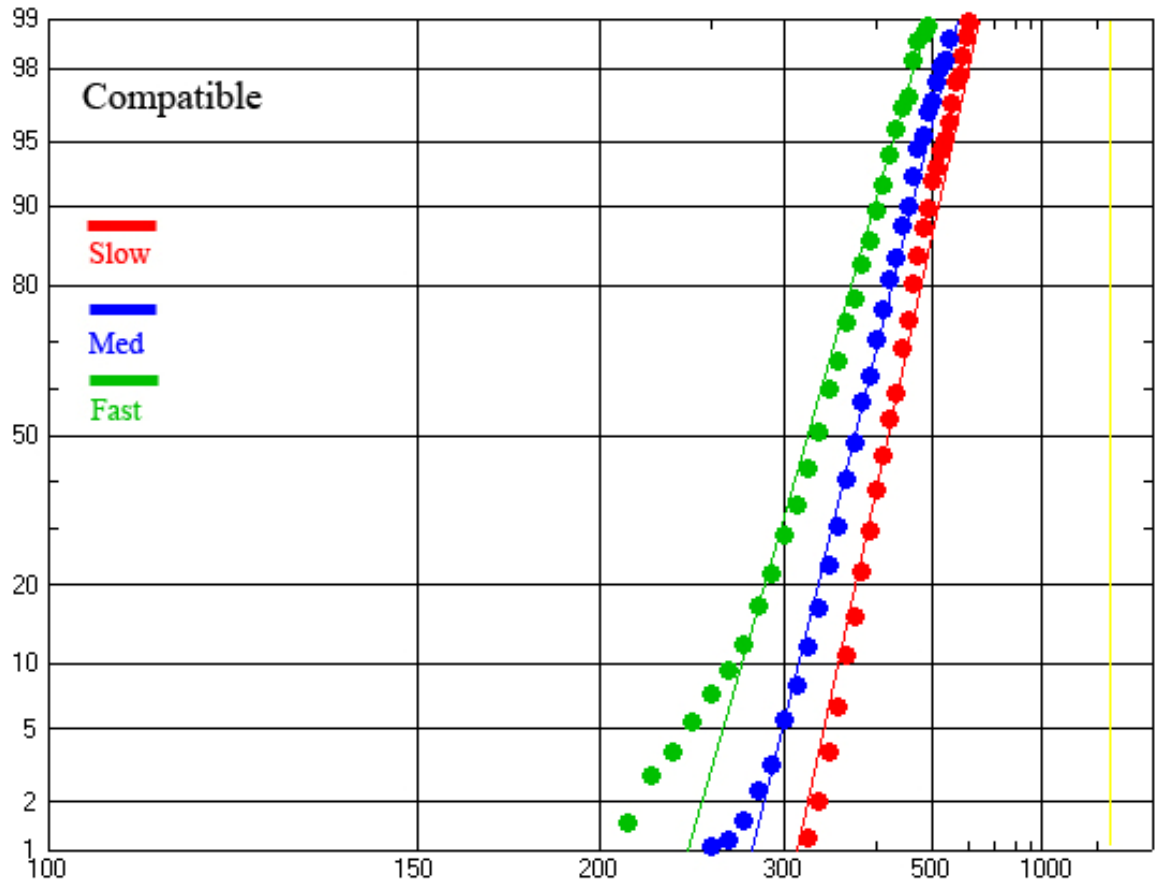


Figure 33. Reciprobit plot for compatible trials. The swiveling of the lines about a common intercept is consistent with a change in criterion/threshold.

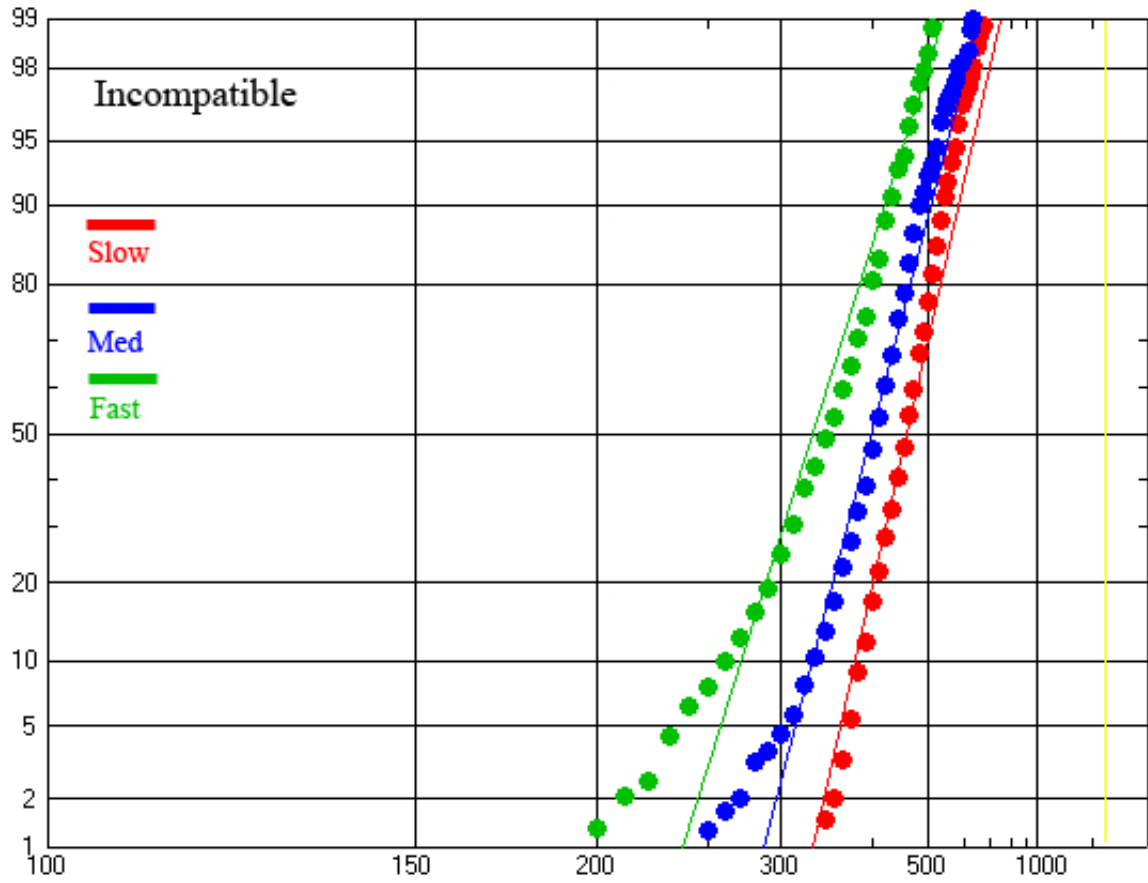


Figure 34. Reciprobit plot for incompatible trials. The swiveling of the lines about a common intercept is consistent with a change in criterion/threshold.

EMG Onset Detection

EMG was detected as in Experiment 1. For each subject and each hand, a criterion voltage was established that reliably differentiated muscle potentials from background noise. On average, EMG medians were 76.0 ms faster than RTs. Table 6 presents EMG and RT medians as a function of SAT condition and compatibility (for correct trials on which the deadline was met).

Table 6.

Median EMG versus RT and (standard deviation): Correct and Deadline Met						
	Slow (dead = 550)		Med (dead = 450)		Fast (dead = 375)	
	EMG	RT	EMG	RT	EMG	RT
Compatible	337 (91)	418 (47)	292 (69)	370 (42)	250 (58)	321 (42)
Incompatible	369 (97)	453 (50)	316 (86)	394 (46)	254 (56)	318 (45)

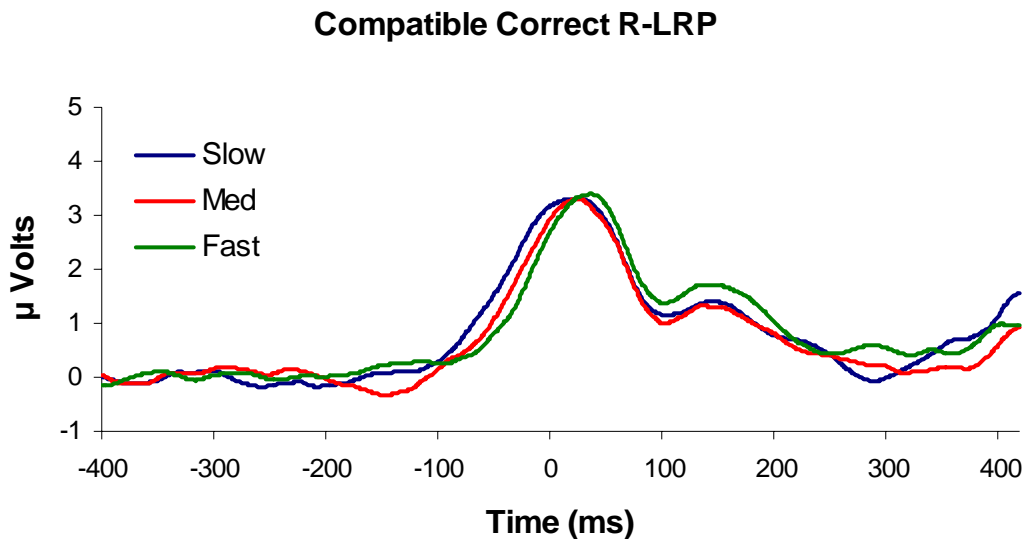


Figure 35. Grand-average R-LRP for correct, compatible trials.

Incompatible Correct R-LRP

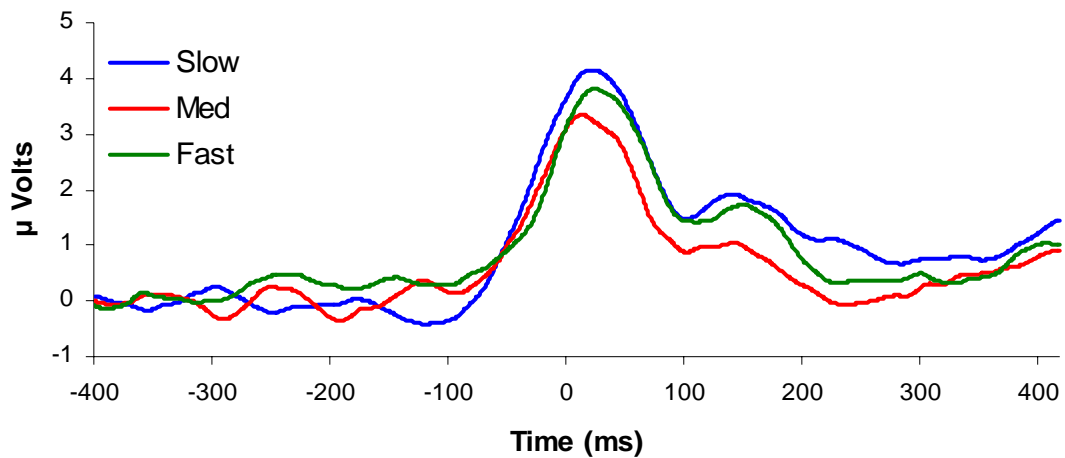


Figure 36. Grand-average R-LRPs for correct, incompatible trials.

Compatible and Incompatible Correct R-LRPs

Figures 35 and 36 present grand average R-LRP waveforms for compatible and incompatible trials, respectively. The difference between Experiment 1 and Experiment 2 is readily apparent. Whereas Experiment 1 showed R-LRP threshold to be a decreasing function of SAT, Experiment 2 suggests that there may be no relationship. Figure 35, for compatible trials, clearly shows that the R-LRP voltage at EMG onset is nearly identical for the *slow*, *med*, and *fast* conditions. For incompatible trials, although the *slow* condition has a somewhat larger amplitude, the *med* and *fast* conditions are identical at response onset (time 0). A 2 (compatibility) x 3 (SAT) repeated-measures ANOVA confirmed that there was no main effect of SAT condition, although there was a main effect of compatibility, $F(1,5) = 7.03, p < .05$, partial $\eta^2 = .58$ and an SAT x compatibility interaction, $F(2,10) = 3.61, p = .066$, partial $\eta^2 = .42$. Simple effects analysis revealed

that this interaction was due to increased incompatible R-LRP amplitudes (as compared to compatible R-LRP amplitudes) in the *slow* condition, $t(5) = -4.99, p < .01$.

Compatible and Incompatible Median Splits

As in Experiment 1, a median split was performed on each subjects' RTs for each compatibility and SAT condition. The purpose of this analysis is to evaluate the extent to which criterion remains constant within a given deadline block. Prior research has shown that the amplitude of the LRP at response does not vary with a trial's latency within a condition, giving rise to the so-called "constant criterion" hypothesis (Gratton et al., 1988). Recall that data presented in Experiment 1 was not altogether consistent with this hypothesis. Namely, a constant criterion was observed for the *slow* condition, but the fast half tended to be smaller for the *med* and *fast* conditions. Experiment 2 plots are presented in Figure 37 and 38. It is clear that there is no consistent relationship between R-LRP amplitude and latency. For instance, the first half actually has higher amplitude in the incompatible *med* condition, while it is smaller in the incompatible *fast* condition. A 2 (compatibility) x 2 (median: fast vs. slow) x 3 (SAT) repeated-measures ANOVA yielded no significant interactions regarding compatibility or SAT, although a main effect of median did obtain, $F(1,5) = 6.73, p < .05$, partial $\eta^2 = .57$. Collapsing across compatibility and SAT condition, the mean amplitude for the fast half was $M = 3.82 \mu\text{V}$ and $M = 3.77 \mu\text{V}$ for the slow half. This is opposite to that observed in Experiment 1.

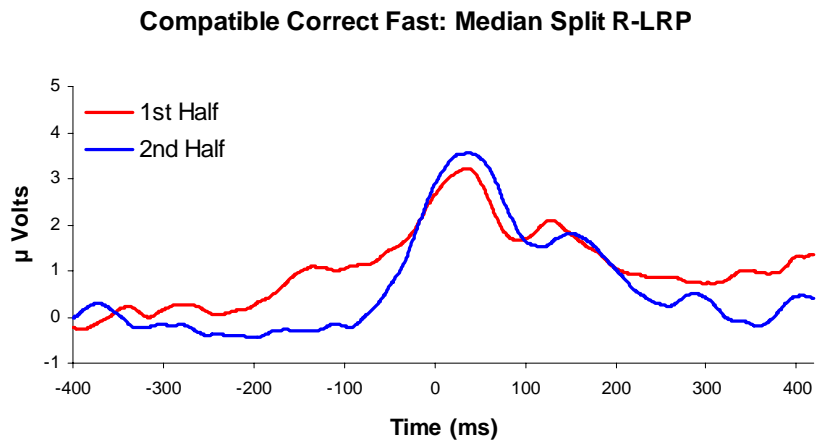
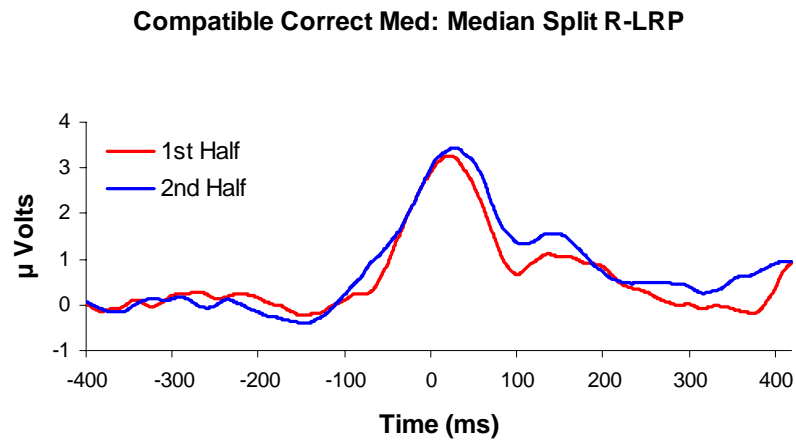
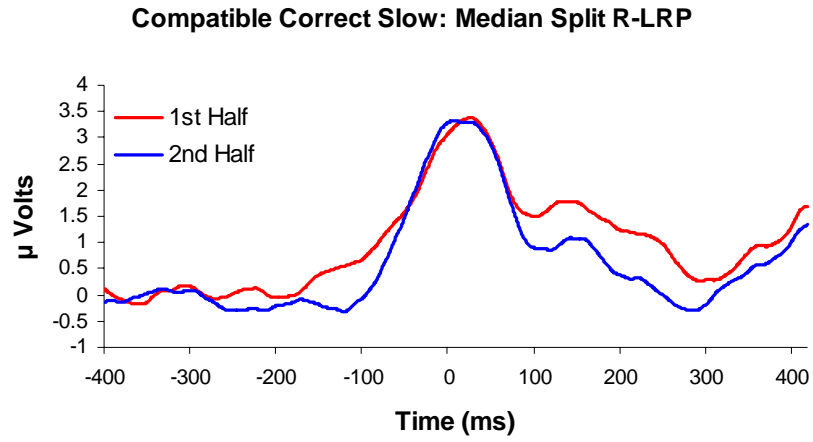


Figure 37. Grand average R-LRP waveforms after median split for compatible trials.

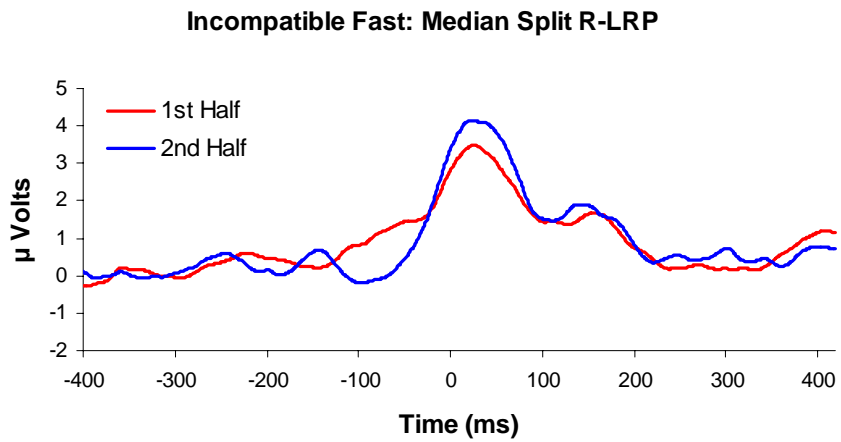
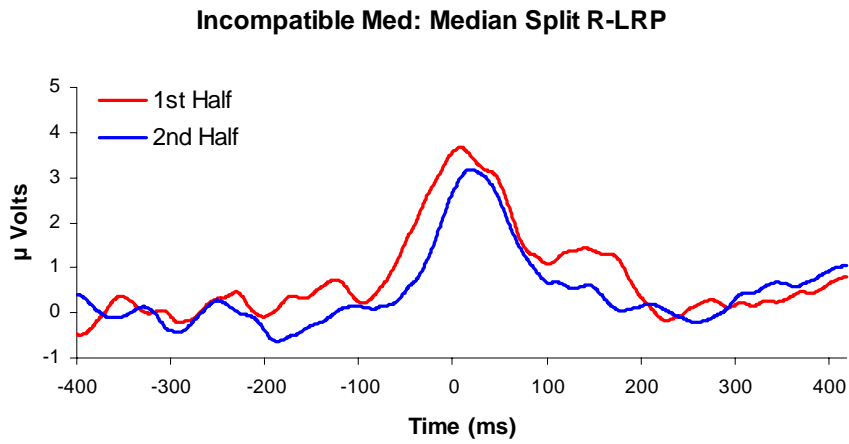
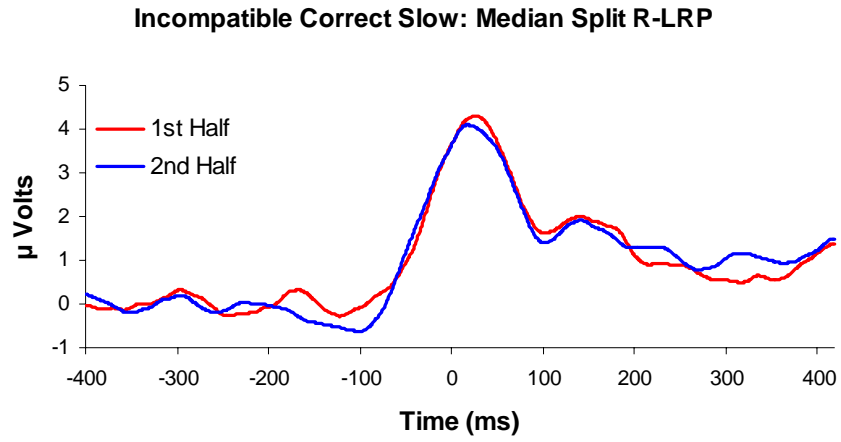


Figure 38. Grand average R-LRP waveforms after median split for incompatible trials.

Incorrect Responses

In Experiment 1, the argument was made that subjects commit an error when the information function reaches the wrong boundary, although the criterion placement is roughly the same. Stated differently, errors should be marked by an R-LRP that looks just like a correct trial, but flipped in sign. Plots for compatible (*fast*) and incompatible (*med* and *fast*) are presented in Figures 39 and 40 (there were not enough observations to derive any other conditions). The mean amplitude at response were as follows: compatible *fast* (correct: 3.29 μV ; incorrect: -3.32 μV), incompatible *med* (correct: 3.69 μV ; incorrect: -2.64 μV), incompatible *fast* (correct: 3.81 μV ; incorrect: -3.60). Although the mean amplitude between correct and incorrect trials was quite similar for compatible *fast* and incompatible *fast*, there was a marked departure for the incompatible *med* condition. Still, all paired *t*-tests comparing correct versus incorrect trials were non-significant.

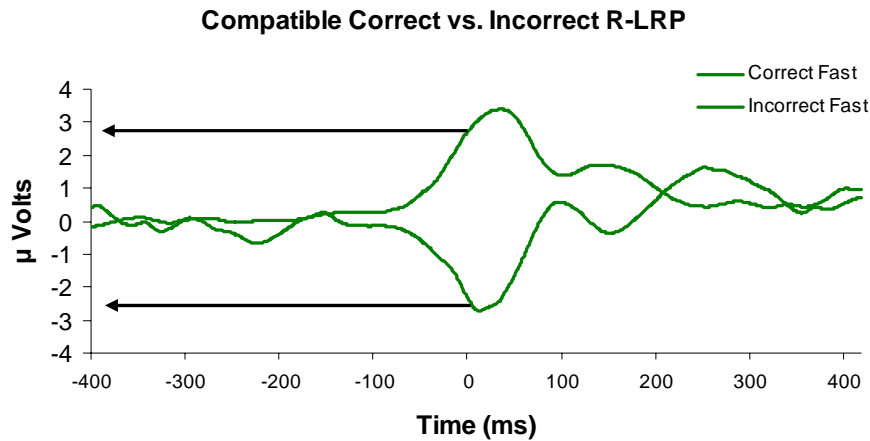


Figure 39. Grand-average R-LRP for correct and incorrect compatible *fast* trials.

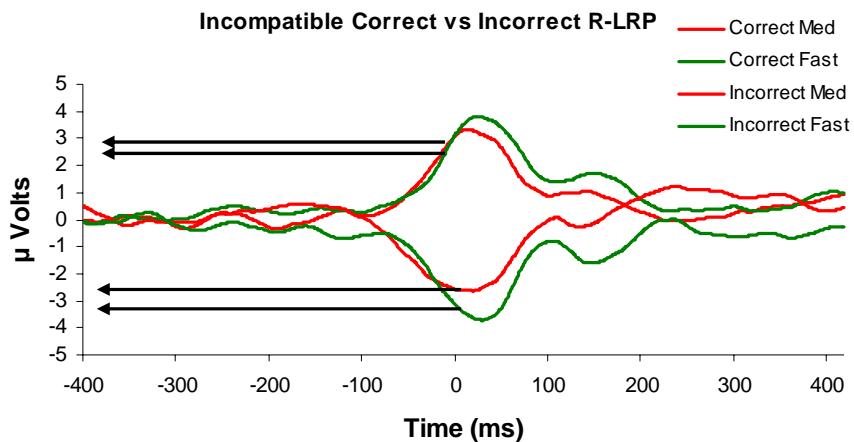


Figure 40. Grand-average R-LRP for correct and incorrect incompatible *fast* and *med* trials.

Missed Deadlines

In Experiment 1, the argument was made that the reason subjects miss a deadline is because they did not reach criterion voltage before the response deadline was enforced. This was supported in two ways. First, the amplitude of *made* and *missed* deadlines was the same for each SAT condition. Also, S-LRPs showed a pattern whereby *missed* deadlines began to rise later and were encumbered by activity that appeared to go in the incorrect direction (i.e., subjects began to prepare the incorrect response). These patterns are evident in Experiment 2 as well. Figure 41, 42 and 43 below depict grand-average R-LRP waveforms for incompatible correct *made* and *missed* deadlines. It is clear that the amplitude of these waveforms at response onset is quite similar, with the exception of the *fast* condition, where *made* deadlines elicit a somewhat larger peak. A 2 (compatibility) x 2 (deadline: made vs. missed) x 3 (SAT) repeated-measures ANOVA, however, did not

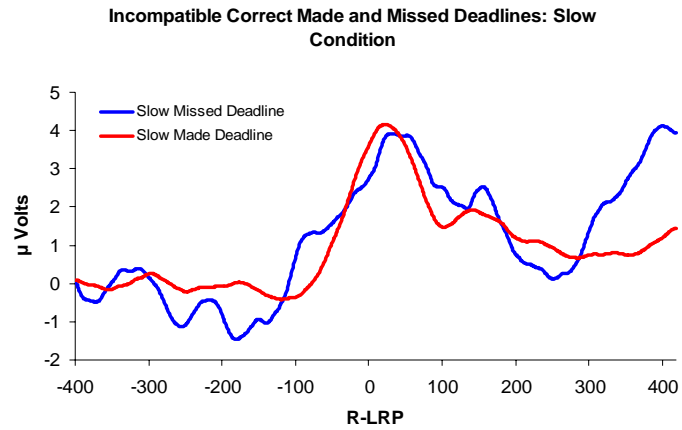


Figure 41. Grand-average R-LRPs for *made* and *missed* deadlines: *slow* condition.

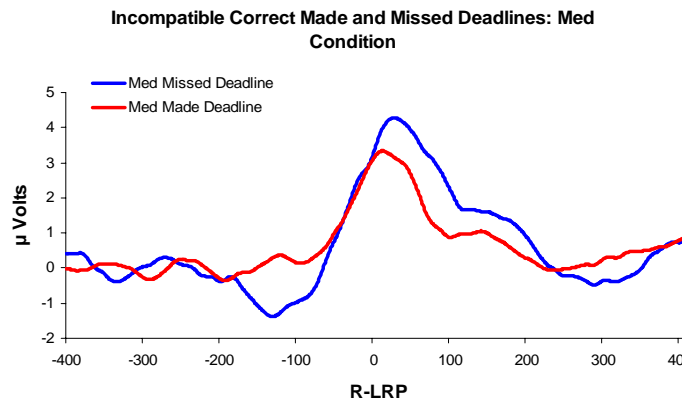


Figure 42. Grand-average R-LRPs for *made* and *missed* deadlines: *med* condition.

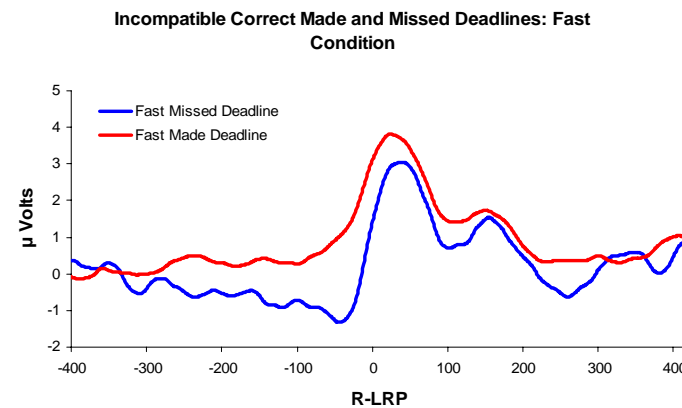
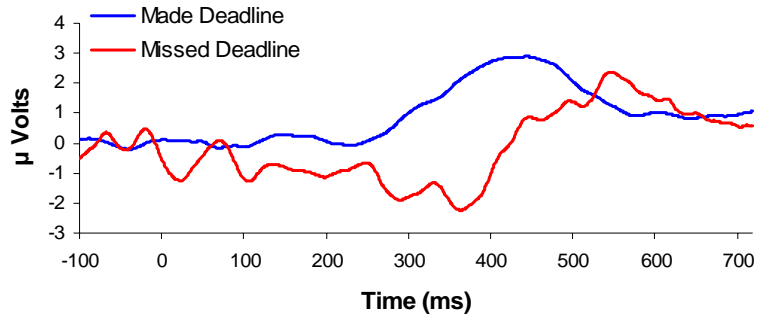
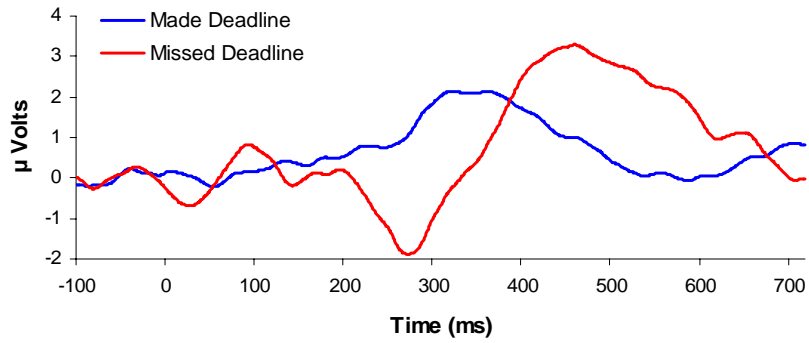


Figure 43. Grand-average R-LRPs for *made* and *missed* deadlines: *fast* condition.

Incompatible Correct Slow: Made vs. Missed Deadlines S-LRP



Incompatible Correct Med: Made vs. Missed Deadlines S-LRP



Incompatible Correct Fast: Made vs. Missed Deadlines S-LRP

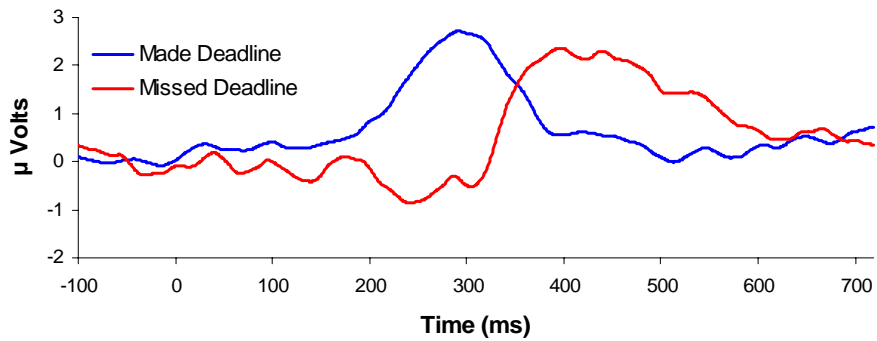


Figure 44. Grand-average S-LRPs for *made* and *missed* deadlines in the incompatible condition (correct trials only).

indicate any main effects or interactions. Looking next at S-LRPs for the same conditions, it is not difficult to see why subjects' responses would have been delayed. Clearly, the lateralization of the LRP indicates priming of the incorrect hand (downward deflections) early in the trial, followed later by a recovery (see Figure 44). It is likely that this is the basis for delayed responses.

Discussion

Unlike Experiment 1, amplitude of the R-LRP at response onset was unrelated to SAT condition. Rather than declining with speed stress, all waveforms tended to reach the same amplitude at response. Also unlike Experiment 1, median splits revealed that amplitude of the LRP is unrelated to response latency, being equivalent for fast trials and slow trials.

There were, however, notable similarities between Experiments 1 and 2. Behaviorally, there was very little difference between the two datasets. Mean RT and accuracy rates were remarkably similar, and time-course decomposition (CAF functions) indicated the same. As well, LATER model fits indicated that a change in response criteria was occurring in roughly the same manner as in Experiment 1.

Electrophysiologically, incorrect and correct trials tended to have the same threshold for response, but opposite in sign. Likewise, *made* and *missed* deadline trials had the same criterion, although S-LRPs showed that *missed* deadlines were much delayed. It was speculated that one reason for this delay is the priming of the incorrect response; the fact that recovery must occur before the waveform rises towards the correct boundary may be the basis for the delay.

CHAPTER 3

GENERAL DISCUSSION

The results of Experiments 1 and 2 paint an interesting, if not altogether consistent, picture of human decision making. The central thesis, that the amplitude of the LRP is related to response criteria as predicted by sequential-sampling models and single-unit recordings, was strongly supported in Experiment 1. Experiment 2 data pose problems for this conclusion, as no relationship between R-LRP amplitude and SAT condition was observed. It is likely that further work will be required to fully understand the nature of the relationship, though some speculation is warranted. In the following sections, each piece of the theory will be detailed along with the conclusions from Experiments 1 and 2.

Main Thesis

Two alternative explanations were initially examined to account for the predicted pattern of decreasing R-LRP amplitude with increasing speed stress. First was the possibility that an increasing proportion of “guess” trials contaminated the data. As mentioned, it is unclear what effect random guesses would have on the LRP, given that a response is still produced. It is entirely possible that such trials would elicit the same LRP, but earlier in time. Also, there is a large difference between a “true guess” (i.e., the response is predetermined before the stimulus onset) and responses that merely *feel* like guesses due to low criterion placement. Subjects were instructed to *never* randomly guess – all trials were to be responded to meaningfully, even if those responses resulted in a high proportion of errors. During debriefing, subjects were probed as to the

likelihood that some of their responses consisted of random guesses. All indicated that random guesses played no role. Of course this is speculative, and requires systematic test. Unfortunately, it is impossible to determine which trials are truly “guesses” and which are simply inaccurate. However, it may be possible to attenuate their influence, and this is the approach taken in Experiment 1. Research suggests that guess trials, if they occur, generally occur at very fast latencies. For instance, CAFs presented here and elsewhere (Gratton et al., 1988; Heitz & Engle, in press) show that true chance performance occurs at the fastest overall latencies. Hence, if one simply considers the slow half of a median split, the influence of such guesses should be much attenuated, if not eliminated altogether. Doing so led to the same overall pattern of R-LRPs, suggesting that guess trials were not a confounding factor.

Second was the possibility that as speed stress increased, subjects began pre-lateralizing in preparation of a response. When this occurs, the computation of an R-LRP uses a baseline correction that is artificially high. By subtracting out an average activity that is too high, these LRPs would become too small. The pattern observed in Experiment 1 could thus be due to changing proportions of “pre-lateralization” trials. To deal with this, R-LRPs for the *fast* condition were re-computed using a baseline period centered on the warning tone, eliminating the problem described above. This, too, left the pattern unchanged. Why, then, did subjects in Experiment 2 show equivalent R-RLP thresholds at each SAT condition?

On the one hand, behavioral data suggest that subjects performed the task quite similarly across Experiments 1 and 2. Not only were mean RT and accuracy rate

remarkably similar, but CAF functions exhibited little change. Notable also is the fact that these data (including LATER model fits) indicate that subjects were, in fact, altering response criteria in roughly the same manner across Experiments 1 and 2. Also, it is clear that subjects did not get “stuck” in a *fast* cognitive set, ruling out the possibility that all conditions were functionally *fast*. On the other hand, electrophysiological data are discrepant across the two studies. In other words, Experiment 2 dissociated behavior from electrophysiology.

The basis for Experiment 2 concerned the confounding of SAT condition order with practice. For that reason, the same subjects were brought back for Experiment 2. The prediction was made that if Experiment 1 was due to a practice effect, then Experiment 2 should yield identical R-LRPs in each SAT condition – exactly what was found. Yet, if a practice effect is indeed to blame, it apparently only affects electrophysiological responses, as behavioral data showed little to no effect of session (i.e., non-practiced vs. practiced). Perhaps it is the case that subjects respond in a qualitatively different way in Experiment 2. Suppose that subjects in Experiment 2, who received *fast*, *med*, then *slow*, actually do get “stuck” in a *fast* cognitive set. Or, to use less loaded terminology, perhaps subjects became accustomed to the timing of the *fast* condition. At the same time, subjects knew that they were supposed to show increasing RTs with decreasing speed stress. It is possible that subjects determined their response in the *slow* block in the same way that they did in the *fast* block, but withheld their response so as to produce a longer RT. This could lead to identical electrophysiological data while leaving RTs unaffected. Unfortunately, this does not explain the SAT in accuracy rate,

which should be absent had subjects selected their response early.

Another possibility concerns the median split data across the two experiments. In Experiment 1, the fast half of each split yielded a lower R-LRP amplitude than the slow half, at least for the *med* and *fast* conditions. This pattern was not evident in Experiment 2. If Experiment 1 *med* and *fast* conditions contained some trials with overall lower R-LRP amplitudes, then averaging across these would serve to lower R-LRP amplitude. This would not have been an issue in Experiment 2, where R-LRP amplitude was unrelated to median split. This concern is somewhat alleviated by the observation that the same pattern emerged in Experiment 1 when considering only the slow halves of each SAT condition. For this to be viable, however, one must make the assumption that the slow halves were not contaminated by whatever it was that caused the fast halves to elicit smaller R-LRPs. Whether or not this assumption is tenable cannot be determined.

While the data thus far are consistent with a practice effect explanation, this author is reluctant to concede this without further test. For instance, a simple practice effect would predict declining R-LRP amplitude if subjects performed *slow-slow-slow* or *med-med-med*. As well, a direct replication of Experiment 1 with a new set of subjects would be desirable to establish that the effect is real.

Errors and Missed Deadlines

One interesting pattern emerging from both Experiments 1 and 2 concern the ability of the R-LRP to predict errors and missed deadlines. It is quite astonishing how similar the waveforms are for correct and error trials, particularly the consistency of the threshold value between the two. That the threshold remained constant reinforces the

idea that the amplitude of the R-LRP *is* in some way related to response production. To reiterate, all responses are produced when brain asymmetry reaches a threshold value. For a right hand response to occur, the left hemisphere must “beat out” the right hemisphere by a set amount. This set amount is unchanged even when an error is to be produced.

Similar observations are made regarding *made* and *missed* deadline trials. Like correct and error responses, *made* and *missed* deadline responses are produced at a set LRP threshold. *Missed* deadlines, when they occur, simply rise to that threshold value later in time. Unfortunately for the subject, this time point was after the enforced response deadline. This is not altogether surprising in and of itself (*missed* deadlines by definition must occur later than *made* trials), though the accompanying S-LRP data is somewhat revealing. Not only are *missed* deadline trials delayed, but the LRP tends to drift towards the incorrect boundary. This provides additional evidence supporting the efficacy of R-LRP amplitude in predicting response onset. It is worthy of note that identical patterns were observed in Experiments 1 and 2.

Conclusion

The data presented in this work speak to the use of the LRP in examining the link between behavior and physiology. The amplitude of the LRP is unquestionably related to some aspect of response production, and in particular, there is good evidence that it represents a threshold. Whether or not this threshold is equivalent to the idea of “criterion” emerging from computational models remains to be conclusively determined. Determining the reliability of data presented in Experiment 1 will be key in this regard.

APPENDIX A
INDIVIDUAL SUBJECT DATA

Experiment 1

Table A1.

E1: COMPATIBLE TRIALS				
Subject		Slow	Med	Fast
1	RT	425	368	285
	ACC	.96	.83	.65
2	RT	440	389	348
	ACC	.97	.96	.90
3	RT	463	399	336
	ACC	.95	.95	.83
4	RT	425	381	330
	ACC	.97	.93	.73
5	RT	453	365	325
	ACC	.91	.68	.56
6	RT	403	372	332
	ACC	.99	.99	.84

Table A2.

E1: INCOMPATIBLE TRIALS				
Subject		Slow	Med	Fast
1	RT	471	379	287
	ACC	.85	.60	.48
2	RT	480	420	367
	ACC	.92	.77	.62
3	RT	512	422	348
	ACC	.88	.74	.46
4	RT	481	411	331
	ACC	.75	.53	.46
5	RT	470	376	314
	ACC	.63	.56	.46
6	RT	455	412	343
	ACC	.93	.80	.57

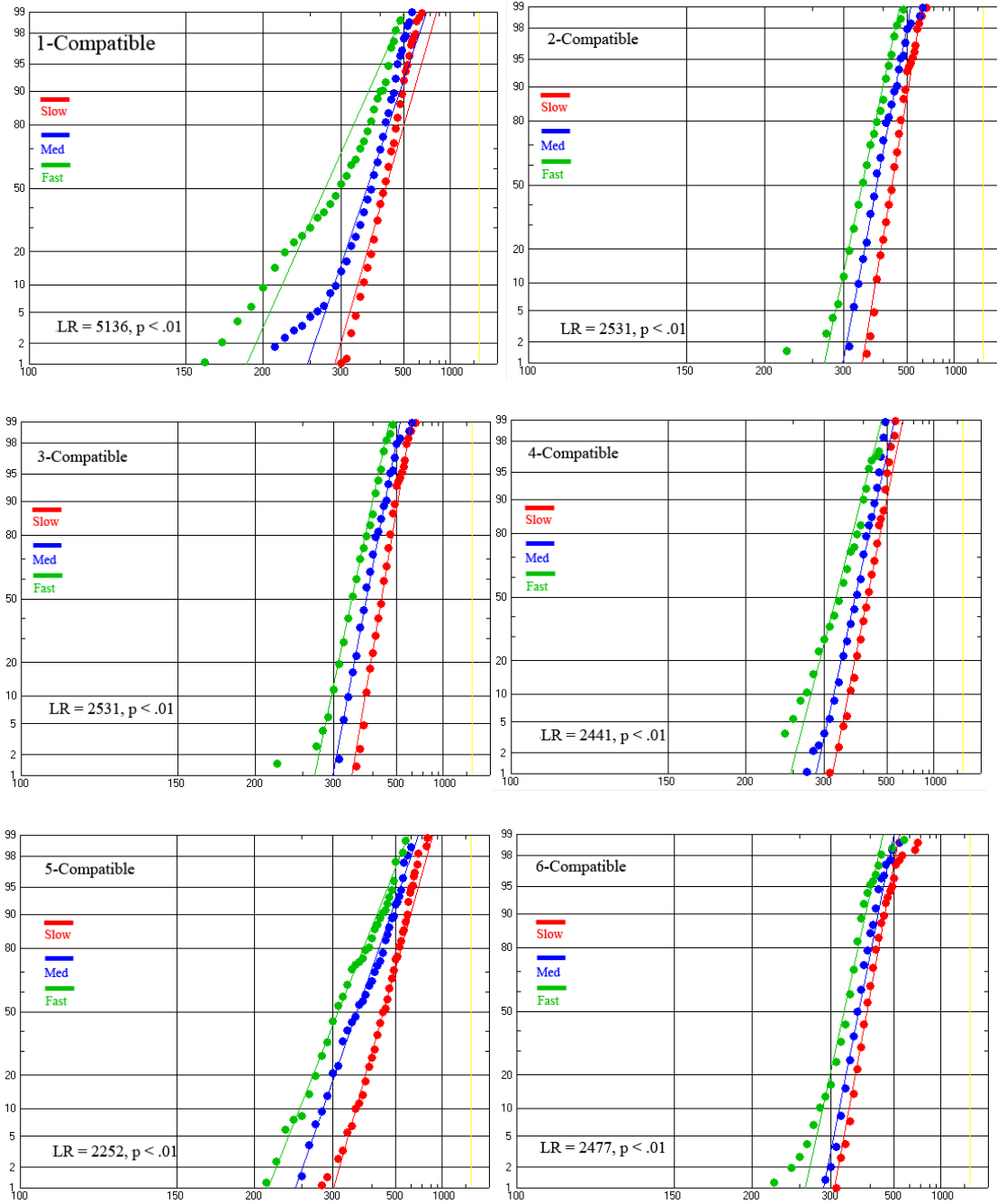


Figure A1. Subject-level likelihood ratio tests for a “swivel” versus a “shift” LATER model for compatible trials, Experiment 1.

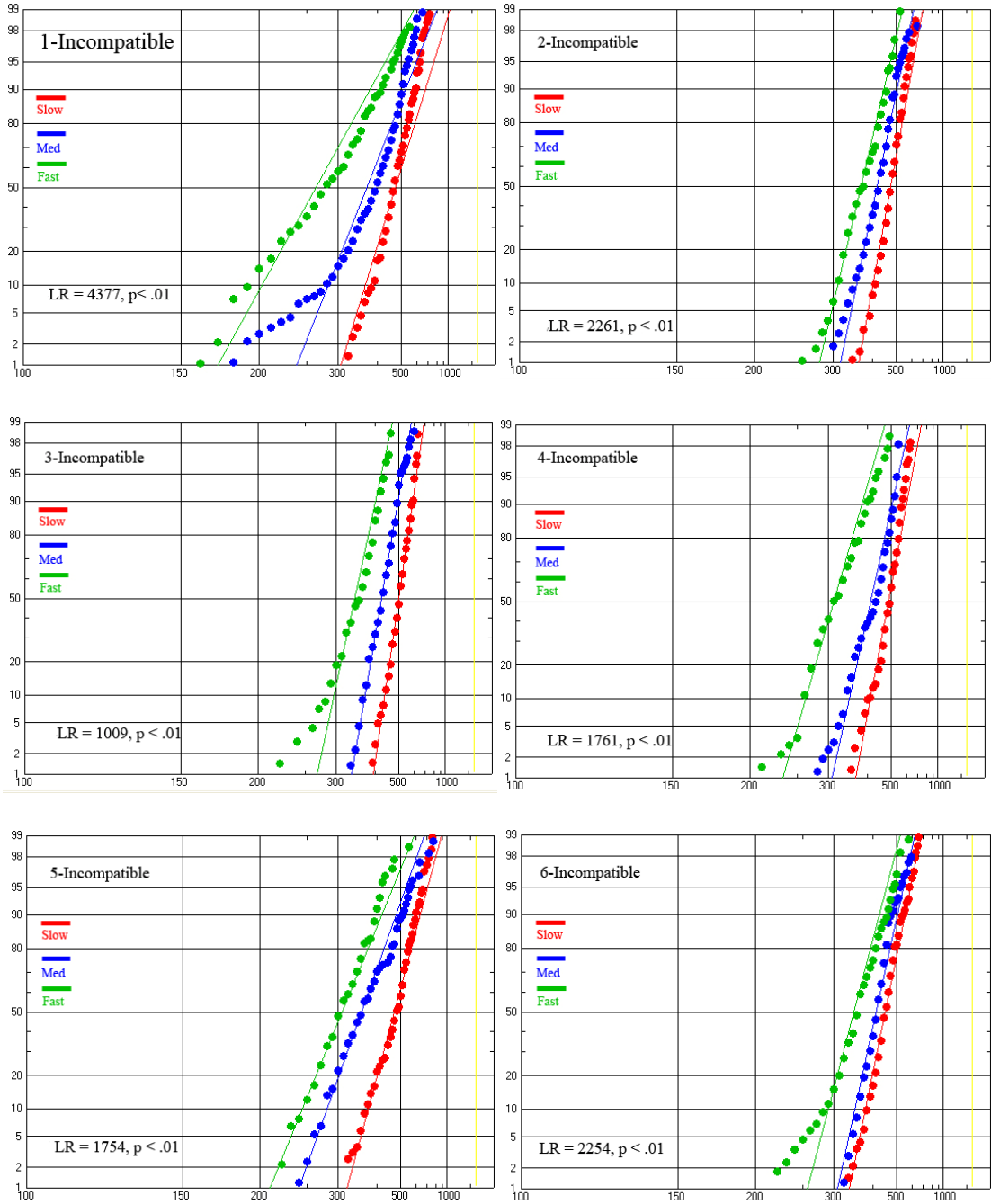


Figure A2. Subject-level likelihood ratio tests for a “swivel” versus a “shift” LATER model for incompatible trials, Experiment 1.

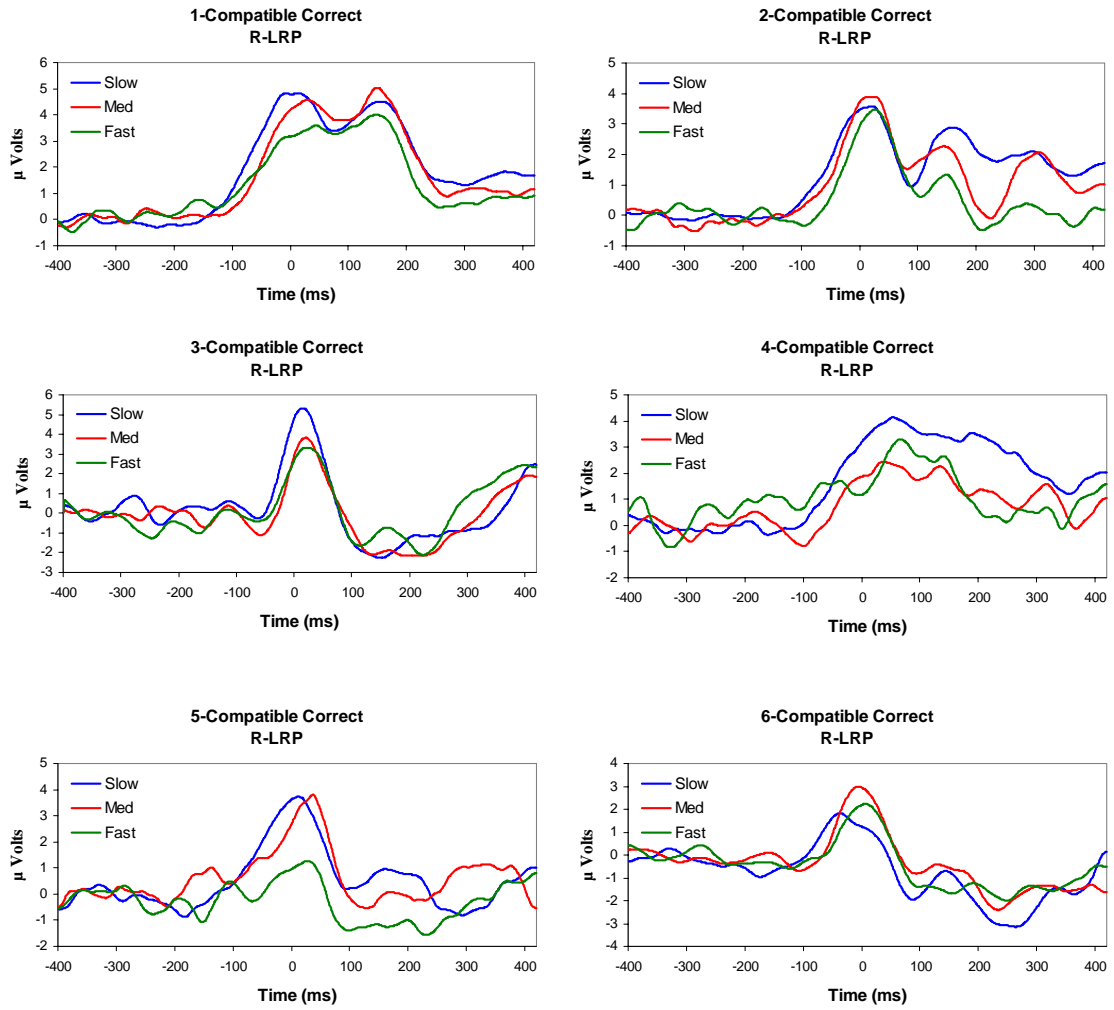


Figure A3. Subject-level R-LRPs for the compatible condition, Experiment 1.

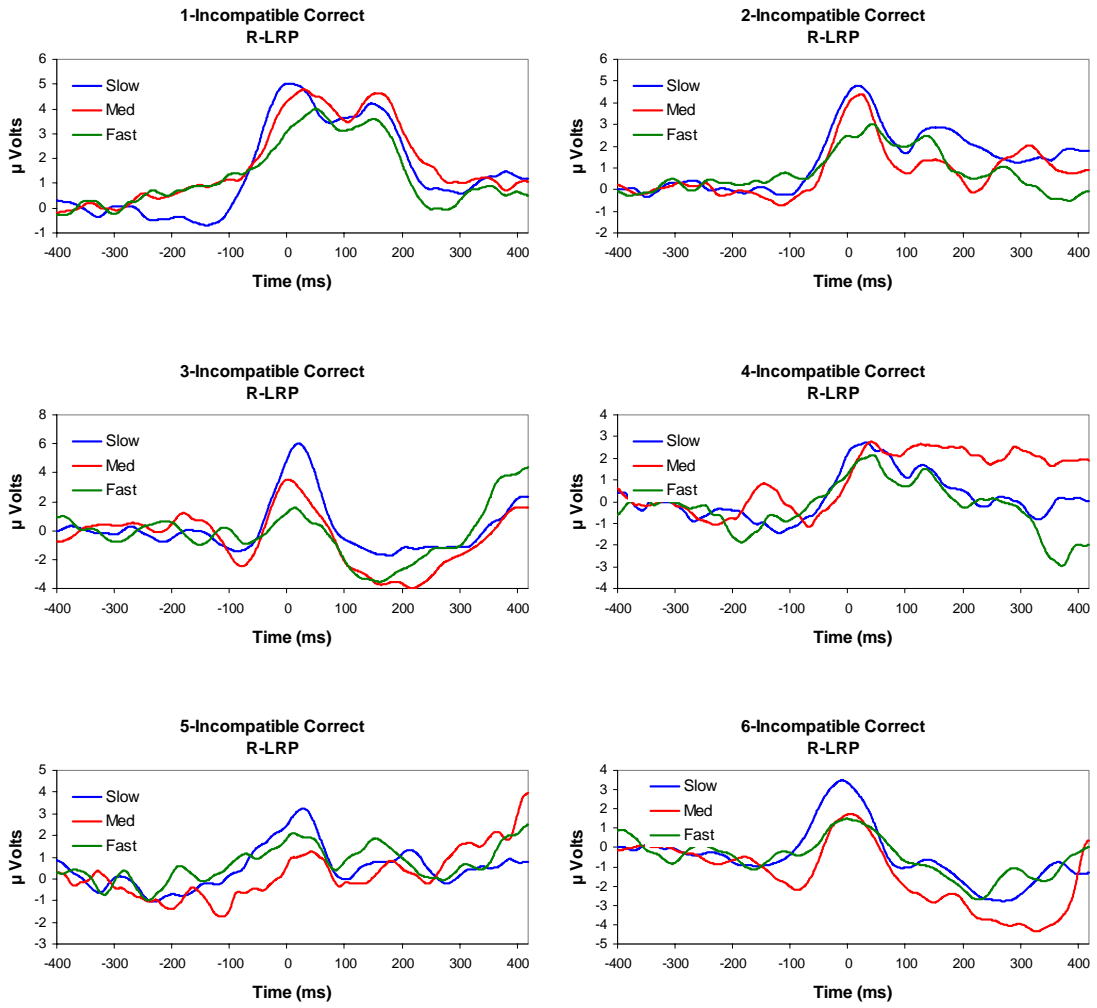


Figure A4. Subject-level R-LRPs for the incompatible condition, Experiment 1.

Experiment 2

Table A3.

E2: COMPATIBLE TRIALS				
Subject		Slow	Med	Fast
1	RT	433	390	311
	ACC	.97	.90	.68
2	RT	430	389	355
	ACC	.98	.95	.95
3	RT	429	384	335
	ACC	.97	.94	.86
4	RT	424	381	329
	ACC	.99	.94	.80
5	RT	455	375	319
	ACC	.87	.65	.55
6	RT	392	368	321
	ACC	.99	.96	.78

Table A4.

E2: INCOMPATIBLE TRIALS				
Subject		Slow	Med	Fast
1	RT	464	417	308
	ACC	.84	.63	.53
2	RT	450	407	381
	ACC	.94	.82	.77
3	RT	454	399	358
	ACC	.85	.84	.61
4	RT	492	400	331
	ACC	.85	.45	.34
5	RT	481	382	323
	ACC	.80	.51	.52
6	RT	424	395	312
	ACC	.87	.70	.49

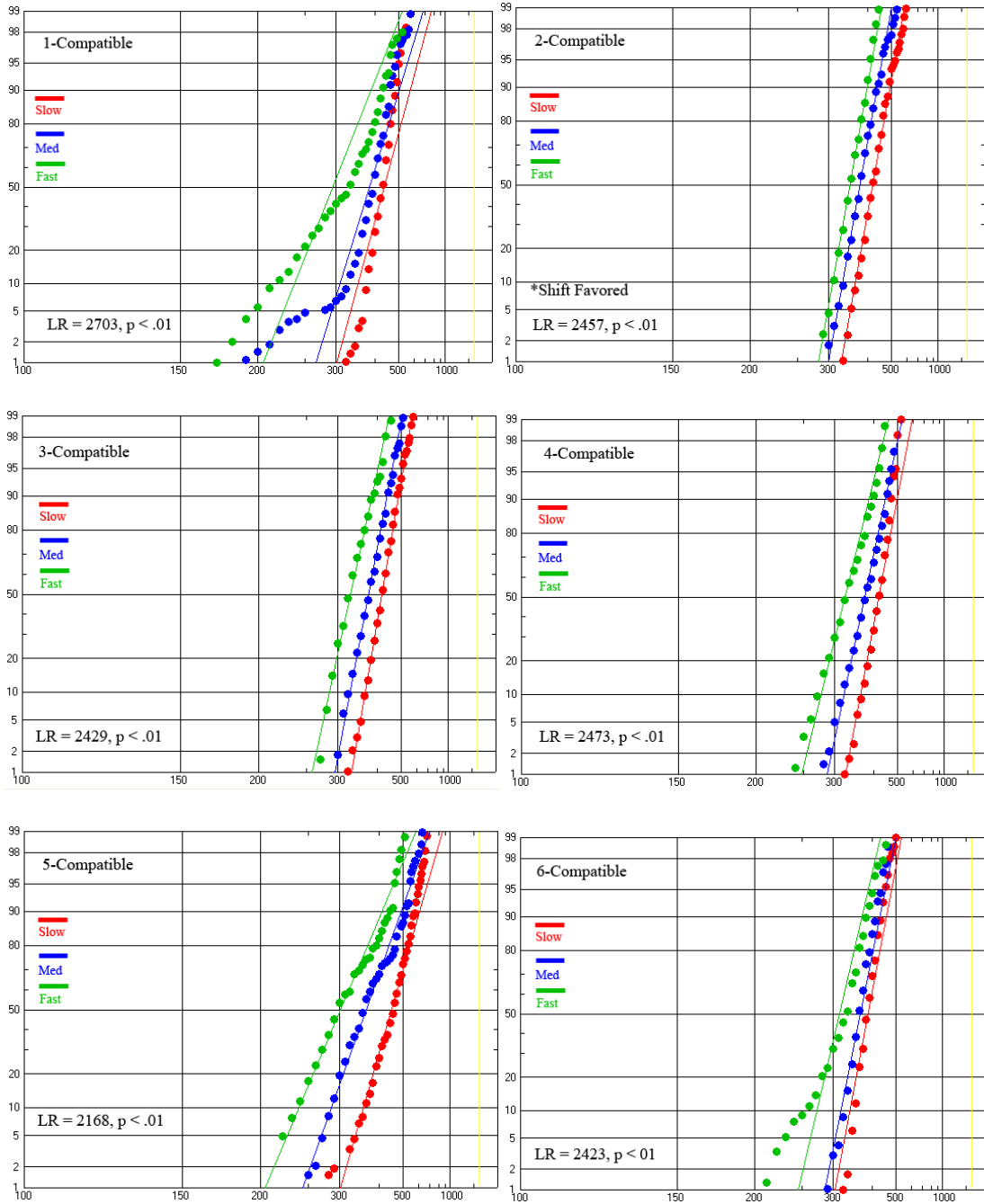


Figure A5. Subject-level likelihood ratio tests for a “swivel” versus a “shift” LATER model for compatible trials, Experiment 2.

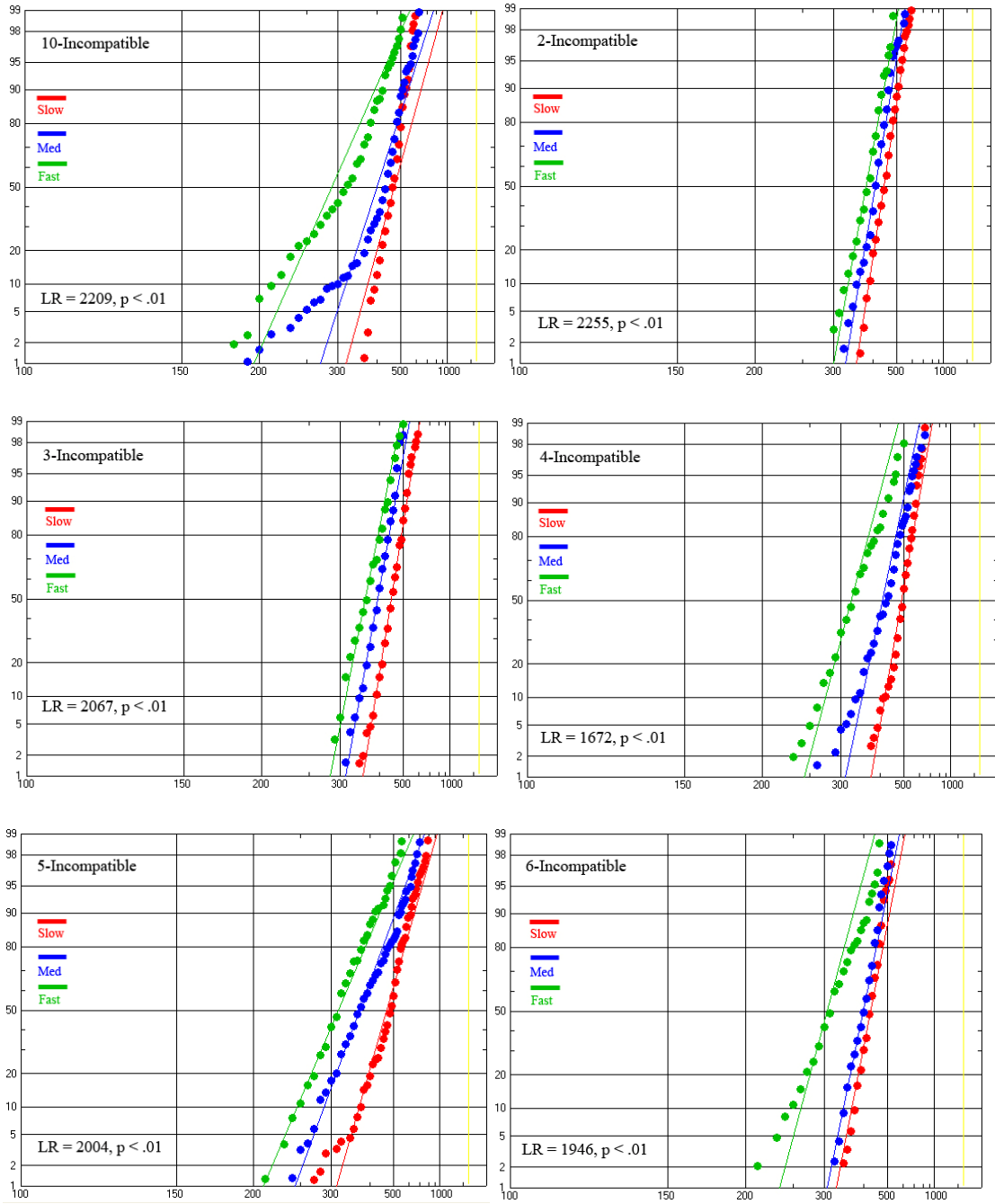


Figure A6. Subject-level likelihood ratio tests for a “swivel” versus a “shift” LATER model for incompatible trials, Experiment 2.

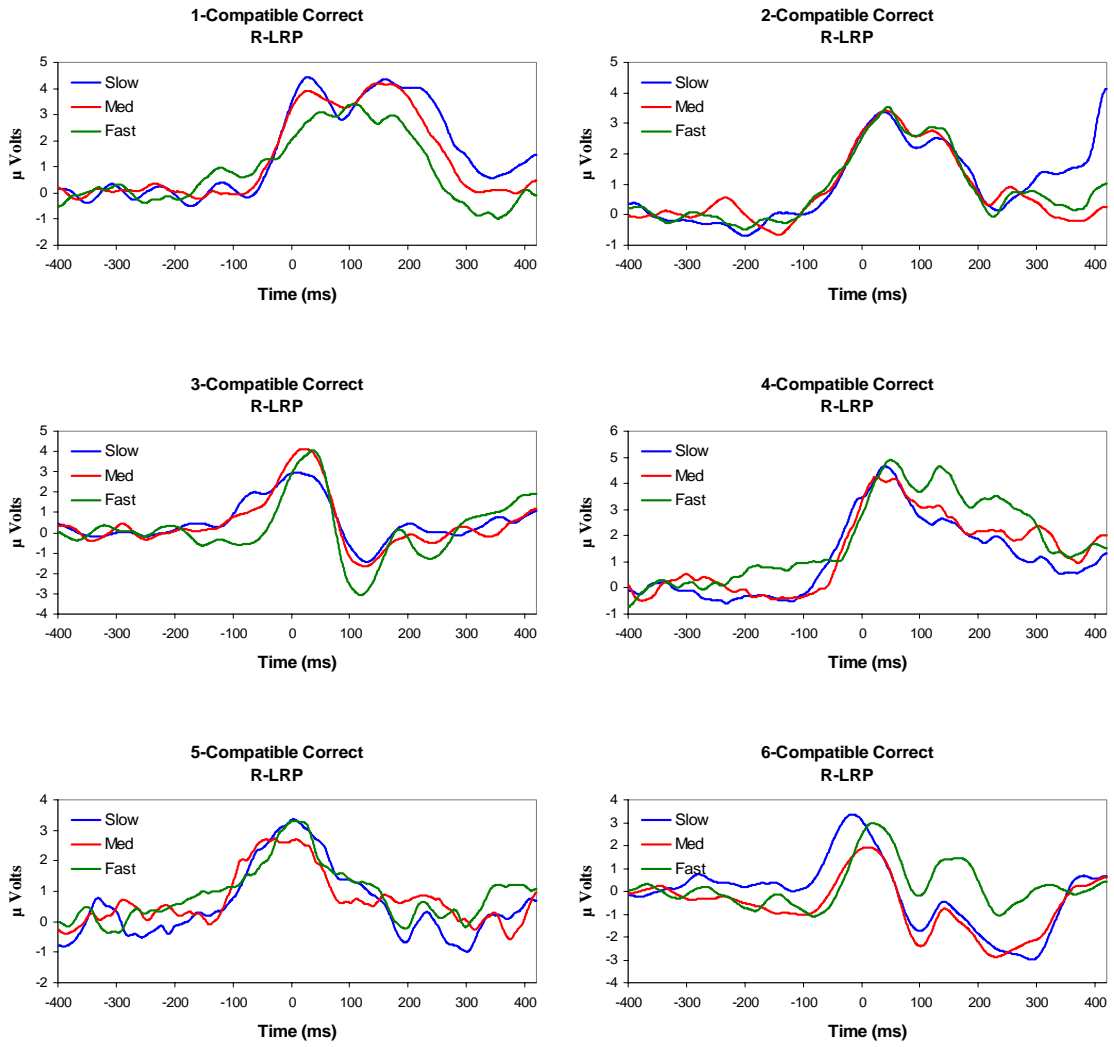


Figure A7. Subject-level R-LRPs for the compatible condition, Experiment 2.

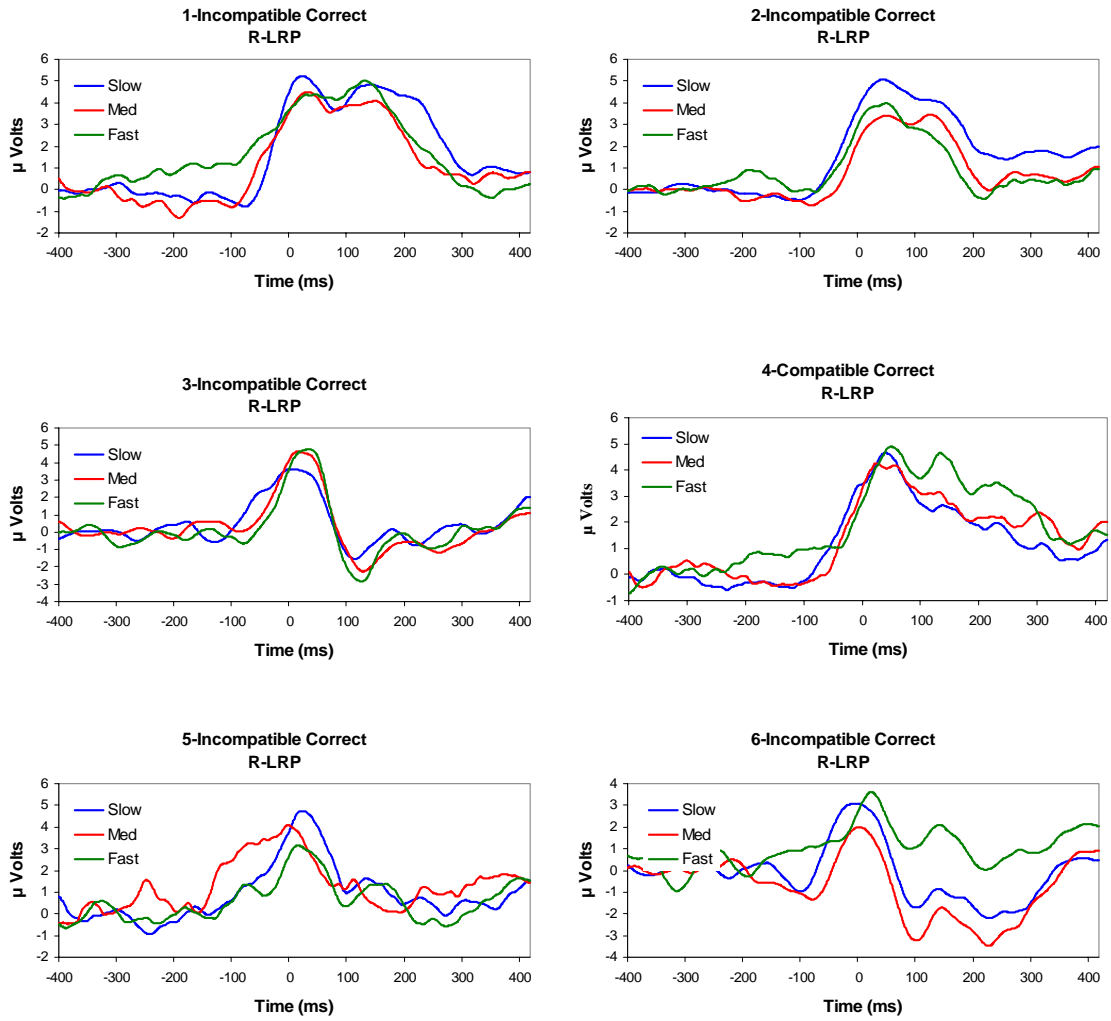


Figure A8. Subject-level R-LRPs for the incompatible condition, Experiment 2.

Footnotes

¹There are important differences between signal detection theory (SDT) and sequential-sampling models, sometimes known as “dynamic signal detection theory” (DST; Balakrishnan, Busemeyer, MacDonald, & Lin, 2006; see also Smith & Ratcliff, 2004). Essentially, SDT assumes that decisions are based on a single sample of information, while sequential-sampling models allow a decision process to evolve, over time. While some attempts have been made to extend SDT in this regard (see Wickens, 2002), it still fails as a complete model of decision making, as it makes only accuracy-rate predictions. Other models, such as stage theory (e.g., Sternberg, 1969) similarly only make RT predictions. Some sequential-sampling models have a considerable advantage over both, as they simultaneously account for both accuracy rate and RT without over-parameterization.

²The author will occasionally use the term “accumulator models” synonymously with “sequential-sampling” models. The latter term is used to emphasize the accumulating nature of the process, *not* to endorse any particular model, as “accumulator models” proper, which consist of a race between two independent processes.

REFERENCES

- Balakrishnan, J. D., Busemeyer, J. R., MacDonald, J. A., & Lin, A. (2003). Dynamic signal detection theory: The next logical step in the evolution of signal detection analysis. Unpublished manuscript.
- Basso, M. A., & Wurtz, R. H. (1997). Modulation of neuronal activity by target uncertainty. *Nature*, *389*, 66-69.
- Basso, M. A., & Wurtz, R. H. (1998). Modulation of neuronal activity in superior colliculus by changes in target probability. *Journal of Neuroscience*, *15*, 7519-7534.
- Bogatz, R. (2007). Optimal decision-making theories: Linking neurobiology with behavior. *Trends in Cognitive Sciences*, *11*, 118-125.
- Bogatz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*, 700-765.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, *10*, 1157-1169.
- Carpenter, R. H. S. (2007). SPIC: A PC-based system for rapid measurement of saccadic responses. Computer Software.
- Carpenter, R. H. S., & Williams, M. L. L. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*, *377*, 59-62.
- Coles, M. G. H. (1989). Modern mind-brain reading: Psychophysiology, physiology, and cognition [SPR Presidential Address, 1988]. *Psychophysiology*, *26*, 251-269.
- Coles, M. G. H., Smid, H. G. O. M., Scheffers, M. K., & Otten, L. J. (1995). Mental chronometry and the study of human information processing. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 86-131). New York: Oxford University Press.
- Ditterich, J., Mazurek, M. E., & Shadlen, M. N. (2003). Microstimulation of visual cortex affects the speed of perceptual decisions. *Nature Neuroscience*, *6*, 891-898.
- Dorris, M. C., & Munoz, D. P. (1998). Saccadic probability influences motor preparation signals and time to saccadic initiation. *Journal of Neuroscience*, *7015-7026*.

- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143-149.
- Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, *25*, 249-263.
- Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (2000). Visuomotor neurons: ambiguity of the discharge or 'motor' perception? *International Journal of Psychophysiology*, *35*, 165-177.
- Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, *173*, 652-654.
- Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*, 390-394.
- Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, *36*, 299-308.
- Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neuroscience*, *55*, 468-484.
- Gratton, G., Coles, M. G. H., Sirevaag, E., J., Eriksen, C. W., Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 331-344.
- Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*, 427-430.
- Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*, 859-862.
- Heitz, R. P., & Engle, R. W. (in press). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*.
- Hernández, A., Zainos, A., & Romo, R. (2002). Temporal evolution of a decision-making process in medial premotor cortex. *Neuron*, *33*, 959-972.
- Kim, J. & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral

- prefrontal cortex of the macaque. *Nature Neuroscience*, 2, 176-185.
- Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40, 77-105.
- Lohman, D. F. (1989). Individual differences in errors and latencies on cognitive tasks. *Learning and Individual Differences*, 1, 179-202.
- Lo, C., & Wang, X. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature Neuroscience*, 9, 956-963.
- Luce, R. D. (1986). *Response times*. New York: Oxford University Press.
- Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, 13, 1257-1269.
- Miller, J. (1983). Can response preparation begin before stimulus recognition finishes? *Journal of Experimental Psychology: Human Perception and Performance*, 9, 161-182.
- Miller, J. (1998). Effects of stimulus-response probability on choice reaction time: Evidence from the lateralized readiness potential. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1521-1534.
- Mordkoff, J. T., & Grosjean, M. (2001). The lateralized readiness potential and response kinetics in response-time tasks. *Psychophysiology*, 38, 777-786.
- Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341, 52-54.
- Nuwer, M. C., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guérit, J, Hinrichs, H., Ikeda, A., Luccas, F. J. C., & Rappelsburger, P. (1998). IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology*, 106, 259-261.
- Ollman, R. T. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, 6, 155-156.
- Osman, A., Lou, L., Muller-Gethmann, H., Rinkenauer, G., Mattes, S., & Ulrich, R. (2000). Mechanisms of speed-accuracy tradeoff: Evidence from covert motor processes. *Biological Psychology*, 51, 173-199.
- Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal

- cortex. *Nature*, 400, 233-238.
- Ratcliff, R. & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347-356.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111, 333-367.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Reddi, B. A. J., & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*, 3, 827-830.
- Riehle, A., & Requin, J. (1989). Monkey primary motor and premotor cortex: Single-cell activity related to prior information about direction and extent of an intended movement. *Journal of Neurophysiology*, 61, 534-549.
- Rinkenauer, G., Osman, A., & Ulrich, R., Müller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General*, 133, 261-282.
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the Lateral Intraparietal Area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22, 9475-9489.
- Salinas E., & Romo, R. (1998). Conversion of sensory signals into motor commands in primary motor cortex. *Journal of Neuroscience*, 18, 499-511.
- Schall, J. (2003). Neural correlates of decision processes: Neural and mental chronometry. *Current Opinion in Neurobiology*, 13, 182-186.
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86, 1916-1936.
- Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience*, 16, 1486-1510.
- Shin, E., Fabiani, M., & Gratton, G. (2004). Evidence of partial response activation in a memory-search task. *Cognitive Brain Research*, 20, 281-293.

- Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, *81*, 174-176.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*, 161-168.
- Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analysis of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 506-526.
- Stern, R.M., Ray, W.J., & Quigley, K.S. (2001). *Psychophysiological Recording* (2nd Ed.). New York, NY: Oxford University Press.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and Performance II* (pp. 276-315). Amsterdam: North-Holland.
- Tehovnik, E. J., Sommer, M. A., Chou, I., Slocum, W. M., & Schiller, P. H. (2000). Eye fields in the frontal lobes of primates. *Brain Research Reviews*, *32*, 413-448.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550-592.
- Van Veen, V. (2006). A neuroimaging approach to the relationship between attention and speed-accuracy tradeoff. Unpublished doctoral dissertation, University of Pittsburgh.
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*, 67-85.
- Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.
- Wood, C. C., & Jennings, J. R. (1976). Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. *Perception & Psychophysics*, *19*, 92-101.
- Zhang, J., Riehle, A., Requin, J., & Kornblum, S. (1997). Dynamics of single-neuron activity in monkey primary motor cortex related to sensorimotor transformation. *Journal of Neuroscience*, *17*, 2227-2246.