STRATEGIC INDIVIDUAL DIFFERENCES IN

HIGH-SPEED HUMAN PERFORMANCE

by

Jonathon Joseph Kopecky

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
2008

Doctoral Committee:

Professor David E. Meyer, Co-Chair
Professor Richard L. Lewis, Co-Chair
Professor Shinobu Kitayama
Associate Professor Yili Liu

Acknowledgements

I'd like to thank my family. They may not really understand the appeal of a PhD, but they were still supportive, and that's most important.

I would also be remiss if I didn't thank the Brain, Cognition and Action (BCA) Lab. Four hours of lab meetings a week may sound like a lot, but a small price to pay to learn to think like a scientist and how to do research. Where else do you get a chance to see the connection between the Rolling Stones and articulatory suppression?

Thanks must go out to Dave Meyer and Rick Lewis, my advisors, for carrying out interesting and yet careful research, and for being great advisors. Thanks also to Shinobu Kitayama for getting me involved in my culture project, which I really enjoyed. Thanks also to Andrew Howes and Mason Smith, who were very instrumental in my computational modeling project. There is no way I would have finished my modeling without their patience and assistance.

Most of all I'd like to thank **Lydia L. Chen** for her support throughout graduate school, and for her helpful comments on an earlier draft. I don't know if I would have finished it without her. We all should be so lucky as to have a person like her in our life.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Multi-tasking and strategic human performance are everywhere. For example, imagine that as you approach a busy intersection, the traffic signal changes to yellow. Do you slow down and wait until the light turns green, or do you speed up and hope to beat the red light? Or imagine that you are working on an e-mail when the phone rings. Do you immediately answer the phone, or do you finish the sentence you are writing and then answer? Or imagine that you are driving your car when your cell phone rings. Should you answer? If so, how closely do you focus on this conversation, and how closely on your driving?

All of these scenarios involve deciding what strategy to perform when you have to make a decision quickly. The traffic-signal scenario is a classic example of the speed-accuracy tradeoff (SAT; Pachella, 1974). If you slow down and wait for a green light, you are sure not to break the law, but you have a delay before proceeding. If you accelerate, you do not have to wait as long, but you may run a red light and risk getting a ticket or into an accident. So what should people choose? What if you are late for an important job interview? What if another ticket would cause you to have your driver's license suspended? The present dissertation investigates how people may answer such questions.

The critical importance of strategies in cognitive psychology have been known since the study of perception via Signal-Detection Theory (SDT; Green & Swets, 1966; Sperling

& Dosher, 1986). In these studies, experimenters were often interested in how sensitive people were to a given stimulus or a given change in stimulus. But in order to understand perceptual limitations, these researchers realized they had to control for strategic differences. Two people with the same degree of sensitivity in a given stimulus modality may nonetheless produce different results, because each person would deal with uncertain stimuli depending on her bias toward a particular response. In order to isolate people's sensitivity, this strategic bias had to be separated from sensitivity, which could be done by manipulating bias via an external payoff. Sperling & Dosher (1986) had the insight that cognitive psychologists could and should generalize the techniques from SDT to other tasks, such as choice reaction-time tasks, to differentiate the effects of interest from the strategic bias of the participants.

One common strategic decision for participants in cognitive psychology experiments is where to settle along the speed-accuracy tradeoff. The difference between settling for 95% accuracy and 100% may result in a quite significant difference in reaction-time (RT) results. Pachella (1974) gives examples where researchers found a significant effect when they examined RT differences and discussed the exciting implications of this result, but when accuracy differences were also considered, the effect disappeared. This example demonstrates it is impossible for experimenters to properly interpret their results if they do not also consider the strategy of their participants.

The Speed-Accuracy Tradeoff can also have a profound effect in experiments beyond just the speed and accuracy for a given task. For example, in Psychological Refractory-Period (PRP) tasks (Welford, 1952; Pashler, 1994), a participant is given two tasks to perform on a given trial, separated by a short time known as the Stimulus-Onset Asynchrony (SOA). Participants have to respond to the first task before they respond to the second task. Within each task there is going to be a SAT point, such that faster performance on that task will result in lower accuracy, but there is also a SAT point in the difference between the first and second response, in that as a participant minimizes this difference she is more likely to

make the second response before the first one. The difference in time between the first and second response is known as the Inter-Response Interval (IRI). If a participant responds to Task 2 before they respond to Task 1, then they have made a mistake on that trial. Meyer et al. (1995) pointed out that instructions for the PRP task usually told participants to prioritize Task 1 over Task 2, and thus avoid making response-reversal errors[1]. These instructions strongly encourage participants to set a SAT point such that their accuracy for their IRI is very high, at the result of a time cost in doing Task 2. The realization that instructions were influencing participant's behavior, but not necessarily any structural cognitive limitation, prompted Meyer et al. to form a class of Adaptive Executive-Control (AEC) models that could be strategically varied. One such model, the Strategic Response-Deferment (SRD; see also Meyer & Kieras, 1997a) fit data from a variety of PRP experiments better than standard models that presupposed an immutable Response-Selection Bottleneck (RSB). Schumacher et al. (1999) found further evidence supporting the SRD model over a RSB model, when they found that across sessions participants first adopted a "cautious" strategy that resulted in very few response-reversal errors, but in later sessions adopted a more "daring" strategy that resulted in faster performance on Task 2 but at the cost of more response-reversal errors.

Howes, Lewis, and Vera (unpublished manuscript) expanded on these response-reversal errors in PRP tasks in their Variance-Bounded Response (VBR) theory of ordered responses. Their theory expands on the "daring" strategy discussed above, as they predict the IRI to be just enough time to separate the distributions of the RTs for Task 1 and for Task 2, given a utility function and certain architectural constraints, such as a measure of an individual's internal motor noise and modality-specific response preparation times. In other words, given the external payoff and estimates of internal motor noise, they are able to predict precisely what mean IRI an individual in the PRP task would set. The difference between cautious and daring strategies would thus be explained by differences in the

---

[1]Response-reversal errors are any trial where the response for Task 2 comes before Task 1, or, equivalently, where the IRI < 0.

perceived penalty for making response-reversal errors: early in the experiments individuals adopt a utility function such that the cost of errors is considerably greater than the actual cost due to the payoff, later in the experiment participants adopt a utility function which reduces the cost of such errors and hence result in a smaller IRI. Thus may strategic concerns guided by architectural limitations determine what the mean IRI is for an individual.

We thus see that strategic considerations are needed to answer fundamental questions in cognitive psychology. If we want to understand what constraints there are in the mind, and progress in understanding cognition, we have to follow the precedent of SDT and dissociate strategic choices from immutable cognitive limitations. Probably the most effective way to separate strategy from actual cognitive constraints is via cognitive modeling (Newell, 1973, 1990; Meyer & Kieras, 1999), because this allows one to test different strategies given certain architectural assumptions. The problem, of course, is that it is always possible to argue about the architectural assumptions, as the debate over whether or not a RSB exists shows (Pashler, 1994; Meyer & Kieras, 1997a), and the fact that different cognitive architectures all make some different assumptions, such as ACT-R (Anderson et al., 2005), SOAR (Laird & Rosenbloom, 1996), and EPIC (Meyer & Kieras, 1997a). To allow researchers to control for a number of architectural assumptions, Howes et al. (unpublished manuscript) used cognitively-bounded rational analysis via constraint-based optimizing reasoning engine (CORE) to help separate architectural assumptions from strategic choices (Howes et al., 2007; Vera et al., 2004) by allowing one to generate models that formally specify architectural constraints and strategy spaces.

In this dissertation, we expand on the previous research into strategies. In Chapter 2 we expand the VBR model, by designing experiments to test if people do adapt mathematically-optimal to an external payoff and their internal motor noise, and by generalizing it so that it does not just apply to the IRI in PRP tasks. By explicitly testing how individuals adapt to the payoff, we hope to narrow the strategy space of possible strategies people choose in cognitive psychology experiments. This also allows us not only to es-

tablish what participants in an experiment are adapting to maximize, but also to evaluate how well the participants have performed. In Chapter 2, we used a simple-RT experiment that allowed for a large range of strategic variation. We awarded points based on performance, to indicate to participants how they should perform. This also allowed us to compare human performance to the mathematically-optimal strategy that would have maximized points. We used two different payoffs, one emphasizing "speed" and another emphasizing "accuracy." We found that participants adapted to these payoffs, and that if there was a low enough level of random variation in the payoffs (i.e. internal motor noise plus external experimentally-controlled noise), participants adapted near-optimally to these payoffs.

After Chapter 2 established that participants were trying to maximize their points in a typical experiment, we then studied strategic performance in more complicated tasks. In Chapter 3, we used a task-interruption procedure, whereby people had to do two different tasks, but had to finish a task that starts later in time before they had finished a task that began earlier in time. For example, if Task 1 began at time 0, and Task 2 began at time 100, then Task 2 had to be completed before Task 1. This is the opposite of the standard PRP procedure, so it could be considered an "Anti-PRP" procedure.

This "Anti-PRP" procedure is complicated enough that there are many different strategies a person could adopt to perform under it, depending in part on cognitive limitations. For example, if a person is trying to do Task 1 as quickly as possible, such that as soon as she chooses a response she immediately begins making it, this may constrain how the task could be interrupted by Task 2. This way of performing, known as "immediate mode" (Meyer & Kieras, 1997a) can be contrasted with another way known as "deferred mode" whereby the response for Task 1 would first enter working-memory, only being further sent to the motor system after some control signal indicating processing of the response was allowed to proceed. We found some evidence indicating that the response mode used by a person did affect how she interrupted a task.

In order to further examine what strategies people employed to deal with task interruptions, for one of the experiments in Chapter 3, we modeled our results using CORE[2] (Howes et al., unpublished manuscript). We found evidence consistent with the Strategic Response Deferment theory of Meyer et al. (1995), such that participants initially locked up their motor processor for Task 1, and only unlocked it when given a go signal after certain stages in Task 2.

Once we have established in Chapter 3 some of the variety of strategies that could be used in a multitasking experiment, Chapter 4 began to characterize what factors encourage an individual to favor one strategy over another. We know from previous literature there are a variety of different strategies different individuals select to multi-task (Schumacher et al., 1999, 2001; Meyer & Kieras, 1997a; Dickman & Meyer, 1988). Are there any systematic differences that would lead us to predict a given individual would favor a particular strategy? In Chapter 4 we look at one such factor: the culture a person is raised in. Masuda & Nisbett (2001) found evidence that individuals from a collectivist society such as Japan tended to have more holistic attentional strategies which encouraged these individuals to process multiple objects simultaneously. However, they found individuals from individualistic societies such as America tended to have more analytic attentional strategies which encouraged these individuals to primarily focus on the focal object and then switch their attention to other objects when needed. Although Masuda & Nisbett did not test individuals in a multitasking context, holistic and analytic attentional strategies could also affect the strategies people choose to use when multitasking. Holistic attention would be consistent with a preference toward parallel strategies where multiple objects would be processed simultaneously. Analytic attention would favor switching between tasks. Indeed, we found evidence that in time-sharing tasks where people have to do two tasks simultaneously, Americans tended to choose a sequential strategy, while Japanese were more likely to choose a parallel strategy. Further, in a task-switching study, we found evidence

---

[2]For more details on how CORE modeling works, see Chapter 3

that Japanese were more likely to choose a parallel strategy whereby they had no costs, whereas Americans preferred sequential strategies, which they performed better than did Japanese who used those same strategies.

Chapter 5 is a brief chapter that took our findings from Chapter 4 about cultural differences in choosing a strategy, and applied that finding to the results from Chapter 2 about how people adapt to different payoff schemes. Because the experiments in Chapter 2 had tightly constrained strategy spaces, we did not think culture should affect the results there. As predicted, we found little evidence that culture had any effect on strategies under those conditions.

Combining our results from Chapters 4 and 5, we found that in tasks where different attentional strategies may be employed by participants, the culture of participants needs to be controlled for in order to ascertain meaningful results. However, in tasks where attentional strategies are constrained so that participants will all choose the same strategy, then cultural differences will not affect the results so participants from different cultures can safely be aggregated.

Chapter 2

Optimal Adaptations in Ordered Responses

2.1 Introduction

Many of our everyday tasks require extremely precise coordination of ordered motor actions, all performed in less than a second. For example, every time we type, we coordinate our key presses across multiple fingers and across our hands. People who know how to touch-type and want to type as quickly as possible, soon become aware of a speed-accuracy tradeoff in their typing: the faster they type, the sooner they finish, but the more mistakes they make, and these errors take time to correct. Maybe it would have been faster overall if they had typed slower and made fewer errors. Or maybe they should have typed quicker and made more errors, but they still would have finished typing and making their corrections sooner. Another example is a pianist playing "Flight of the Bumblebee" who may want to play it faster, but has to worry about making errors and taking away from the quality of the piece. What strategies are used in coordinating responses in tasks that require at least two motor responses, and how close do people get to performing these tasks optimally?

A crucial component in performing these tasks is the time between the first and second responses, known as the inter-response interval (IRI). For many tasks, the total amount of time to do the task could be minimized by minimizing this IRI. However, due to noise in the motor system (Meyer et al., 1988), if the average IRI is too small, this would result

in too many mistakes whereby the second response is made before the first response. If the average IRI is too large, although mistakes would be rare, the task would be done inefficiently as it could be finished earlier. So finding the optimal mean IRI to choose for a task is an interesting problem in decision-making, because optimal IRI has to combine both the speed in doing the task with accuracy in performing it to result in optimal behavior.

The study of optimality and optimal behavior dates back to signal-detection theory (SDT) in the 1950's (Green & Swets, 1966; Sperling & Dosher, 1986), which examined how a person's strategy should adapt to their payoff, regardless of their sensitivity to the stimuli in the actual task. There has been plenty of research within the perceptual domain comparing human perception to an "ideal observer." An ideal observer is a model of optimal performance on a given task, given known structural constraints. Ideal observer models have been used in vision research (e.g. Geisler, 1989; Knill, 1998), auditory perception (see Macmillan & Creelman, 2005, for a good overview), and even cognitive tasks such as classification tasks (Edwards & Metz, 2006). The general finding is that human observers differ from ideal observer models.

Sperling and Dosher (1986) discussed generalizing the findings of SDT to various other non-perceptual tasks, including choice-reaction time experiments and speed-accuracy tradeoffs. By showing that these other tasks are isomorphic to signal-detection tasks, we can apply their logic to other problems to discern optimal behavior. Because we are interested in a speed-accuracy tradeoff here, we will discuss below in more detail how the signal-detection approach applies to our experiments.

In our experiments, we are interested in the speed and accuracy of successive responses separated by an IRI. To determine how accurately participants determined their mean IRI, we examined the distribution of IRIs to assess what mean would maximize points. Although it is well-known that reaction-time distributions tend to follow a Gamma distribution (Luce, 1986), or a log-normal distribution, we only need an approximate distribution. Hence we used a Gaussian distribution as a convenient approximation.

9

If participants know the external payoff, as they do in our Experiment 1, they can use this as the basis of their utility payoff. For example, in the Accuracy-Payoff condition, participants earn 100 - RT/5 points if correct, and - 100 points if incorrect. Participants have to estimate what mean IRI they should use and what their level of internal motor noise is, if IRI comes from a Gaussian distribution: $\frac{1}{\sqrt{2\pi}\sigma_{Inoise}}e^{\frac{-(x-\mu_{IRI})^2}{2\sigma_{Inoise}^2}}$ where σ_{Inoise} is the level of internal noise, and μ_{IRI} is the mean IRI they should use. Participants are then trying to find the μ_{IRI} that will maximize their expected utility:

$$v = \int_0^\infty (100 - \frac{\frac{1}{\sqrt{2\pi}\sigma_{Inoise}}e^{\frac{-(x-\mu_{IRI})^2}{2\sigma_{Inoise}^2}}}{5}) + \int_{-\infty}^0 -100$$

Although this equation could be solved for a maximum, this is insufficient for our purposes, because there is no way to compute the variance of this optimal mean IRI. If, say, the participant chooses a mean IRI 10 msec off from optimal, would this result in a discernible difference in points earned? As Roberts and Pashler (2000) pointed out, it is not sufficient merely to fit the data; the confidence interval also has to be considered, i.e. the range of the prediction. This equation does not tell us how big a difference there is if a person is a few msec too slow or too fast. Hence we need to consider the confidence interval around the optimal IRI, in order to judge both the quality of our model and the closeness of our fit.

A better approach to analyze optimal performance on tasks is to compare human performance with mathematically optimal performance generated from Monte-Carlo models. Maloney, Trommershäuser, and Landy (2007) use this approach to understand performance on a timed pointing task. In their task, people had 700 msec to touch a computer screen within a particular circle, but would be penalized if they touched within a different circle that overlapped their desired circle, or would be both penalized and rewarded if they responded within the overlap. Participants were given a payoff function to motivate them, with more points earned in the task equal to more money earned in the experiment. The payoff was manipulated within-subject, such that a penalty (responding in the wrong circle)

would be 0, 100 or 500 points off; a correct response (responding within the proper circle) always earned 100 points.

Maloney et al. (2007) were then able to examine whether participants adapted to the payoff and to what extent they adapted. They found that participants clearly adapted, in that when the penalty was increased, participants made fewer errors where they responded within the wrong circle and more errors where they responded outside of both circles (for which they were not penalized). When the penalty was 0, however, participants almost always responded within the proper circle, even if that meant making more errors by pressing the screen where the two circles overlapped.

In order to assess how well participants adapted, Maloney et al. realized they needed a proxy to measure the participant's motor noise level. To do so, they averaged all of each participant's endpoints and used their average point as the goal each participant was aiming for. Given this point, they then measured to what degree participants differed from this point, and they found this followed a normal distribution, with a mean along each coordinate that was the average point, and a different standard deviation for each participant, σ. This standard deviation was then used as a measure of each participant's internal motor noise level. Given the objective payoff function and a person's motor-noise level, Monte Carlo simulations were then run to find the optimal point to aim for given those conditions. Participants could then be compared to these simulations to examine how close the point they aimed for was to the optimal point. They found that all but one participant did not significantly differ from optimal.

A similar line of research into optimality was conducted by Meyer et al. (1988), who focused their attention on Fitts' Law, which determines the time taken to reach something as a function of its distance and size. They found that a form of Fitts' Law may be derived from assuming the motor control system adapts optimally to its own noise.

Recently, Howes et al. (in press) posited a Variance-Bounded Response (VBR) theory for ordered responses. This theory posits that given cognitive and motor constraints, and

an individual's expected utility function, the IRI will be the duration that maximizes the payoff. If we use the actual objective payoff function as a proxy for the individual's expected utiltity function, we are able to test this zero-parameter theory in our experiments by comparing its predictions versus human behaviors.

Previous research also considered what constraints there are in a choice RT task with two key presses. One such constraint may be the goal of going as fast as possible, the soft-constraints hypothesis (Gray et al., 2006), which states that we try to use interactive routines that minimize the time it takes to complete the task, while achieving expected benefits. However, there are two important caveats in applying this hypothesis to a lot of tasks that focus on IRI. For one thing, the hypothesis does not clearly state how it combines its goal of minimizing reaction time with the risk of decreased accuracy. Another important caveat is that this hypothesis does not apply to tasks that take less than one-third of a second. In fact, it is not clear from this hypothesis that people can adapt to tasks that take less than 333 milliseconds to finish. So if there is a sub-task that takes this long but people cannot adapt to it, this may explain why people under this hypothesis do not perform optimally.

Another concern is that because IRI tasks have a very short RT, it is unclear whether participants would even be able to adapt for times so small. It is possible that the IRI could not average less than 50 msec. This result would be expected from a straightforward analysis of production rules used in cognitive architectures, which tend to have a 50 msec firing rate. This is the firing rate of production rules in many cognitive architectures, including ACT-R (Anderson et al., 2005), SOAR (Laird & Rosenbloom, 1996), and EPIC (Meyer & Kieras, 1997a). However, a participant could adopt a clever work-around to this structural limitation, if she was so inclined. Instead of trying to program two motor responses which would be separated by at least 50 msec, she could try to make one motor response, such as press down with both hands simultaneously. So long as one hand began at a slightly higher point than the other, then this one response could generate any mean IRI, depending on the height of this difference.

Inter-response interval is important to examine not just for its real-world use, but because it comes up in many complicated tasks within experimental psychology. For example, in Psychological Refractory Period (PRP) studies (Meyer & Kieras, 1997a,b), a person is given two tasks, A and B, where Task A is presented first and has to be completed first. Task B is presented following a stimulus onset-asynchrony (SOA) some time after Task A is first presented. After a participant responds to Task A, then they have to respond to Task B. Because participants who respond to Task B before Task A will be penalized for responding in the wrong order, it is very important for participants to consider how to schedule their responses to the two tasks such that Task A is completed first. We hypothesize that participants will schedule their responses as a function of the external payoff and their internal motor noise, although of course participants may internalize the payoff so that it doesn't perfectly correspond to the external payoff.

The IRI for the participant is of crucial theoretical importance in cases with a short SOA. If the penalty for responding to Task B before Task A is severe enough, or the reward for responding quickly to Task B is too small, a participant may choose to have a large IRI to minimize the possibility of making an order-error. This large IRI may even help explain the PRP effect (Pashler, 1994; Meyer & Kieras, 1997a), whereby Task B has a larger reaction time when the SOA is small than when the SOA is large. If IRI is controlled strategically, then to understand performance on any task that has two responses we need to consider how the participants are deciding what their mean IRI should be. Figure 2.1 shows a hypothetical example of two distributions, where the solid distribution corresponds to the RTs associated with Task A, and the dotted distribution corresponds to the RTs associated with Task B. The difference in means for each distribution is the mean IRI, here about 25 msec. The filled area corresponds to those times where the person responds to Task B before responding to Task A, i.e. makes a response-reversal or order-error. As mean IRI increased, the filled area will decrease, and the participant will make fewer errors.

In this dissertation we will present three experiments, all of which focus on analyz-
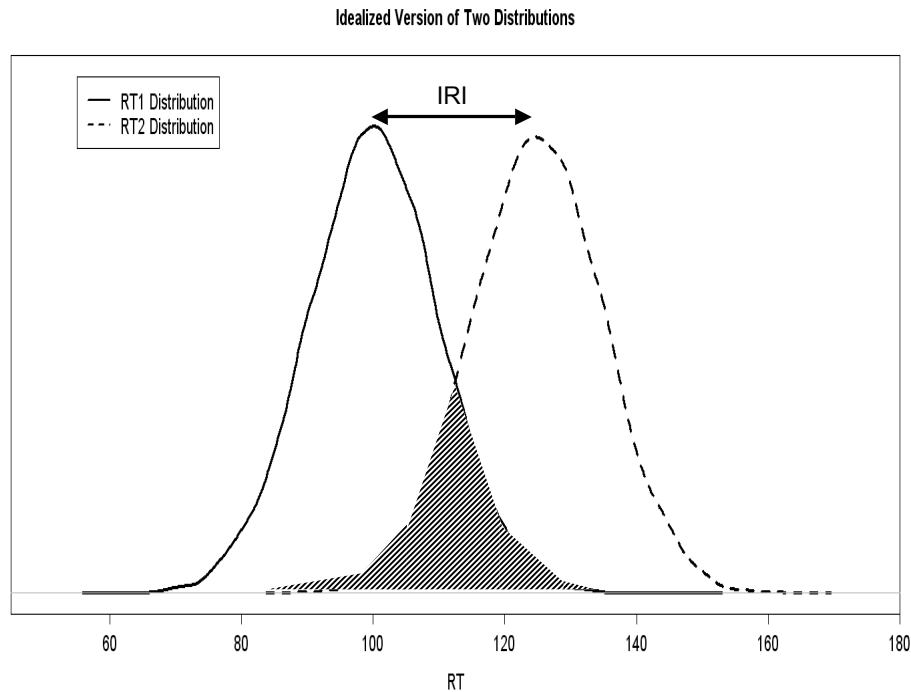
Figure 2.1. Idealized distributions for RT1 and RT2 in a task; filled region represents error.

ing inter-response intervals in ways similar to how Maloney et al. (2007) analyzed timed pointing tasks. Our experiments were designed to be as basic as possible, using a simple reaction-time (RT) task that only requires two key presses. The standard deviation of the IRI of these two key presses could be considered a measure of internal motor noise, once the participant had stabilized her strategy. We also had a payoff function during the experiment that translated into more money if participants earned more points, based on their performance. Although we did not make participants explicitly aware of the details of the payoff function, they were provided with enough feedback about their performance that they could learn the payoff function and earn the maximum number of points.

The previous research on optimality focused on people adapting to internal noise, but not an external, experimentally-controlled noise. Signal-detection theory did initially focus on how people respond to Signal compared to Signal and Noise, but the task was not really

to optimally adapt to noise but to detect the signal. It is quite possible that people are better able to adapt to internal noise than to external noise. With internal noise people may, unconsciously, have more information about the distribution of noise and so be better able to adapt. This internal noise may also be small, and people may only be able to adapt to small amounts of noise. However, by experimentally adding a random number to participants results that affects their performance, participants would have to learn to adapt not just to their internal noise but to this external noise.

In Experiment 1, we had a between-subject design that included three payoff conditions that were based on total reaction time from when a participant began until the second key press was made. After running a few participants, we realized the behavioral data was not what we expected. We then closely examined our payoff function itself, and discovered that optimal performance given our payoff functions was not differentiable across payoffs. We include this experiment here mostly as a warning of how crucial it is to derive the right payoff to get participants to perform in a meaningful way.

Experiment 2 had two payoffs, based solely on IRI and not total reaction time. These payoffs were first tested using Monte Carlo simulations to ensure that optimal performance was differentiable. Another change from Experiment 1 was we added external noise at different levels (a high and a low-level). We were interested to see if participants would adapt not just to internal motor noise but to external, experimentally controlled noise. We found that participants clearly adapted to their payoff, and adapted near-optimally when given little external noise. Participants in the external high-noise group, however, did not adapt particularly closely to optimal. Experiment 3 was basically a replication of Experiment 2, but with different levels of external noise (either no-noise or medium-noise). Here we found adaptation close to optimal in all four conditions. Participants adapted to the payoff, even when their mean IRI was only 15 msec.

In our Experiments 2 and 3, for the external-noise condition, we add a random number to a participant's IRI, drawn from a random, independent Gaussian distribution with

mean 0, but with a different standard deviation depending on the condition. Our payoff is more complicated than in Experiment 1, and participants are not explicitly told the payoff. However, if we assume that IRI is normally distributed, then the equation that participants adapt to is: $\frac{1}{\sqrt{2\pi(\sigma_{Inoise}^2+\sigma_{Enoise}^2)}}e^{\frac{-(x-\mu_{IRI})^2}{2(\sigma_{Inoise}^2+\sigma_{Enoise}^2)}}$, where σ_{Enoise} is the standard deviation of the external noise. Naturally this is a more complicated distribution to adapt to, as there are two parameters the participant must adapt to.

2.2 Experiment 1

2.2.1 Method

Participants.   Nine undergraduate students from the University of Michigan were run in return for monetary compensation. Participants received an $8 base payment, along with a bonus depending on performance (see below for details).

Apparatus.   Data collection was controlled by a computer with a WINDOWS™ operating system and E-Prime software. Stimuli were presented on a 17-inch Sony Trinitron video display, and responses were recorded via a customized manual-response panel with finger keys for each hand. The video display was located 80 cm from the participants, who sat at a table with their heads on a chin rest in a quiet lab room.

Experimental Design and Procedure.   Each participant was run for 1440 trials, divided into 12 blocks, each with 120 trials. Each block required a different set of key presses, but the key press was the same for every trial in a given block. Participants did not know which key press they would have to do until the first trial of a block, but they were informed that every trial would require the same key presses in a given block.

The task was a simple reaction-time experiment, in which on every trial as soon as they saw the word "GO!" on screen, participants had to press two of their finger keys in

a particular order. Only the right-middle, right-index, left-index, and left-middle fingers were used. To allow the possibility of pressing the keys in the wrong order and thus make an order error, the combination involving the same finger pressed twice was not used. All twelve possible combinations of two key presses were performed. For example, in Block 1 on every trial participants may have had to first press their right-middle finger and then their left-middle finger. Block 2 may have required them to press their right-index finger and then their right-middle finger on every trial. Blocks were randomized for each participant.
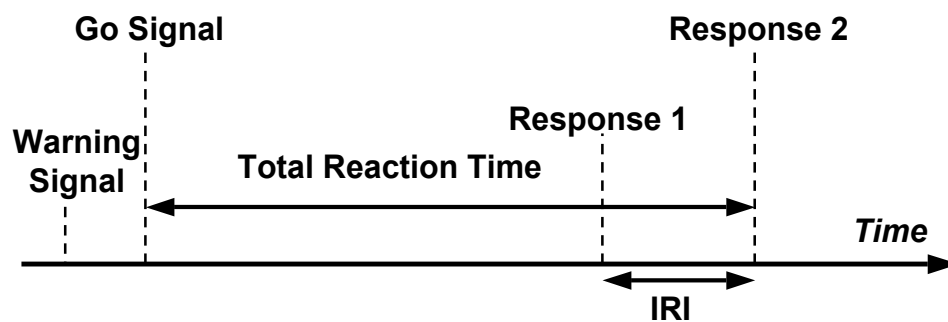
Figure 2.2. Timeline for a typical trial in Experiment 1.

We used a between-subject design, manipulating payoff across participants. All participants were told that their goal was to earn as many points as possible. They were also given monetary reward depending on the number of points they earned, so that the more points they earned, the more money they earned. Participants earned points the same way across conditions: on every trial, a participant began with 100 trial points, and then lost a point

17

for every 5 msec it took to respond. So if a participant correctly responded in 200 msec, they would earn 100 - 200/5 = 60 points on that trial. The three payoffs differed in how errors were treated. In the accuracy condition, participants lost 100 points if they made an error on a trial. In the speed condition, participants lost 0 points if they made an error on a trial. In the tradeoff condition, participants lost their Reaction-Time/5 points if they made an error on that trial, so that participants lost more points if they took longer to respond and made an error

Between blocks, participants were allowed to rest as long as they wanted. When they were ready, the experimenter pressed a key to begin the block. On screen, each participant saw four finger keys that resembled the finger keys in front of the participant. A "1" appeared above one of the keys on screen, and a "2" appeared above another. Participants had to press the keys in front of them in the order that they appeared on screen, i.e. press the "1" key first and the "2" key second.

These instruction keys appeared on screen for 1000 msec (for the first 10 trials of each block), then were replaced by a fixation that appeared in the center of the screen. After the fixation was there for 700 msec, it disappeared and that same location was filled with the word "GO!" and participants could respond. Participants who responded before the word "GO!" appeared on screen got feedback with an "Early Error" and received 0 points for that trial in all payoff conditions.

Participants were given feedback after every trial, telling them how many points they earned on that trial, and how many points they had earned for that block so far. They also saw a visual image of the keys that corresponded to the type of error they made. If participants responded correctly, they saw the instruction keys.

At the end of a block, participants were told their average reaction time on the previous block, their average points per trial, and their total points.

18

Data Analysis. We examined each participant's mean IRI and mean RT. We then also ran Monte Carlo simulations to compare each participant's performance versus optimal performance given her standard deviation of reaction time. Although it did not change our results, we also removed any outliers that were more than three standard deviations away from the mean.

2.2.2 Results

In examining overall main effects, we only used data from the second half of the experiment (Block 7 on) because by then performance across blocks had stabilized and participants seemingly settled upon a strategy. However, because each block had a unique key press, and we had to use data from all second-half blocks for these analyses, we aggregated over individual types of key presses.

Payoff	Total RT (msec)	IRI (msec)	sd.IRI (msec)	Accuracy (%)
Speed	124	37.24	22.43	95.67
Accuracy	143	46.38	24.24	97.67
Middle	165	47.34	27.82	98.33

Table 2.1. Comparison of mean RTs across payoffs.

Given we ran only nine subjects across three conditions, we did not find any significant effect of payoff on Total RT ($F(2,8) < 1$, ns) or on IRI ($F(2,8) < 1$). The fact that IRI did not reliably differ across payoffs indicated we were unlikely to find any tradeoffs, although there was a slight trend in the direction we predicted. We also examined accuracy (by doing an arc-sin transform) and again found no difference, $F(2,8) < 1$, ns. Table 2.1 presents the mean scores for the three participants in each payoff condition.

Because the trend in our results did not conform to what was expected, we made formal mathematical models using Monte Carlo simulations to compare our data to predicted results. In these models, we programmed the actual external payoff, and used each participant's standard deviation of IRI as the basis of internal motor noise. Hence a participant

Optimal IRI as a function of SD of IRI

sd of IRI
Trials=500, Opt. Trials=5, RT1.mean=100, RT1.sd=45, IRI.mean=44, IRI.sd=25
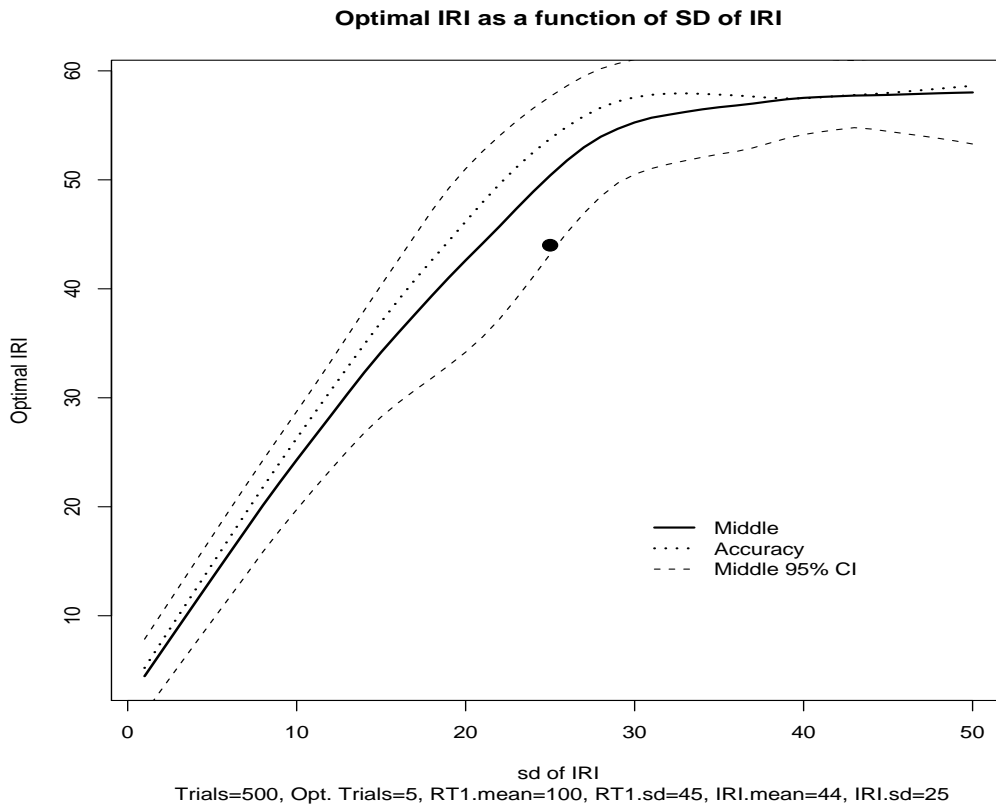
Figure 2.3. Optimal mean IRI as a function of standard deviation of IRI; the solid dot is one participant's observed mean IRI and s.d. of IRI.

with more internal motor noise should have a larger IRI than a participant in the same payoff condition with less motor noise.

Figure 2.3 shows a plot of optimal IRI as a function of the standard deviation of IRI. This plot was generated via Monte Carlo simulations, where 500 points were randomly generated from a Normal Distribution with a mean of 100 and a standard deviation of 50, corresponding to taking 100 msec to make the first key press and a standard deviation of that first key press of 50 msec (a reasonable approximation to the human data). These points were added to 500 points generated from another Normal Distribution used to represent the IRI, with mean 44 and standard deviation 25 (estimated from the human data). We find that the optimal IRI is very similar for the middle and accuracy-payoffs, regardless of internal noise level. As can clearly be seen, the optimal IRI for accuracy falls within the 95% confidence interval for the middle-payoff. The optimal IRI for the Speed Payoff also

20

Expected payoff as function of IRI at SD=25
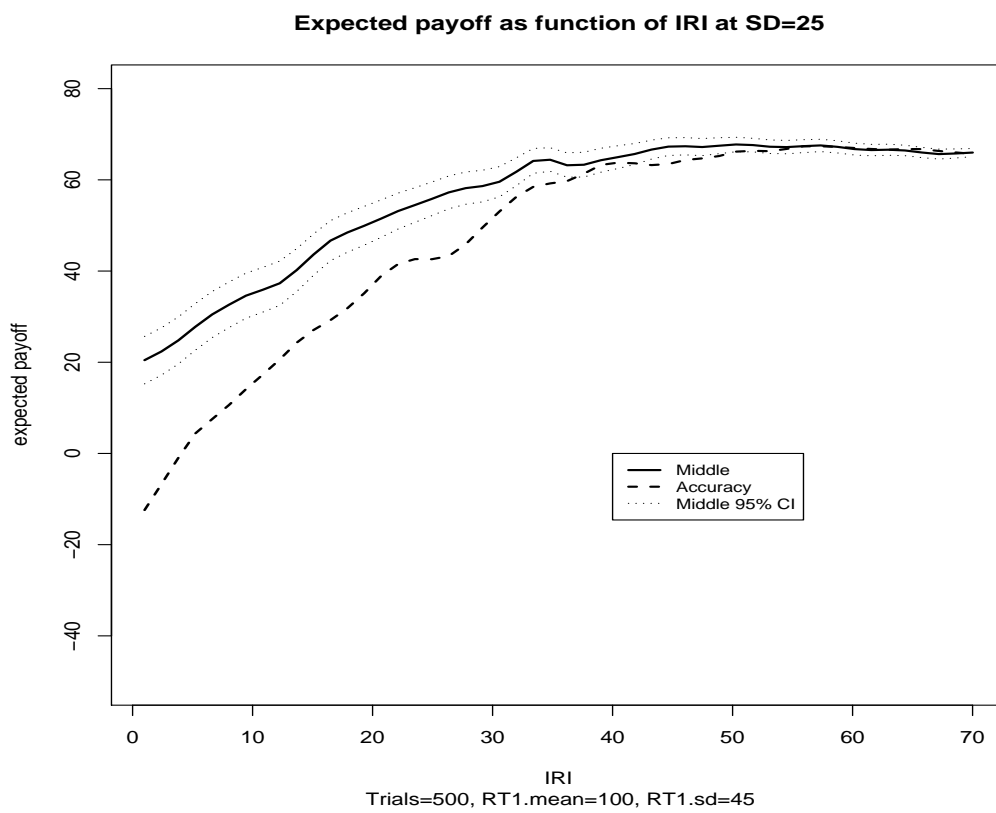
Figure 2.4. Expected payoff as a function of IRI: Middle and Accuracy Payoff.

fell within the 95% confidence interval. The large dot in the figure is a typical participant's observed mean IRI and s.d. of IRI. Because this participant's results fall within the confidence interval of optimal performance for different payoffs, the different payoffs did not sufficiently separate different optimal performance.

In Figure 2.4 we next examined the degree that our payoffs were separated. From Table 2.1 we can see that mean IRI of participants in the accuracy and middle-payoff conditions are virtually identical, so for ease of readability we plotted the expected payoffs for those two conditions. We plotted this again via Monte Carlo simulation, assuming that RT1 comes from a Normal Distribution with mean of 100 and standard deviation of 50. We then generated 500 points from that distribution, and assumed IRI comes from a normal distribution with an unknown mean and a standard deviation of 24 msec (which we estimated from the data). We then generated 500 points at each different mean IRI, and calculated how many points participants would earn on average per trial. This allowed us to figure out expected payoff as a function of mean IRI, across different payoffs. From Figure 2.4 it is apparent that after an IRI of about 20 msec, there is not sufficient separation between the two payoffs. In fact, the expected payoff for accuracy falls within the 95% confidence interval of the payoff for the middle-payoff condition.

2.2.3 Discussion

Although Experiment 1 did not work out as we originally planned, it was illuminating in that it allowed us to learn from our mistakes. We initially did not fully consider some of the issues involved in a payoff scheme, such as examining separation of different payoffs before running participants. Without doing so, even if participants did perform optimally, we may not be able to differentiate their adaptations to different payoffs, and so may incorrectly conclude participants did not adapt properly.

Another problem with our payoff scheme was that it allowed for a larger range of strategies than we expected. For example, some participants tried to anticipate the "go" signal

and so made more early errors than other participants who waited until they perceived the signal. Our models that factored in what would be the optimal time for RT1 and an optimal IRI quickly got so complicated that we realized we needed to simplify our experiment to get better data.

2.3 Experiment 2

Experiment 1 did not have a sufficiently precise payoff scheme to differentiate optimal performance in adapting to those payoff schemes, so participants did not adapt the way we expected, due perhaps to there being too many strategies that would result in optimal or near-optimal performance. In order to better constrain the number of possible strategies participants should choose for optimal performance, the strategy space, we made the payoff based solely on the IRI instead of total RT. Finally, we were worried that participants would not adapt optimally because they initially learned a sub-optimal strategy and decided it was an acceptably good strategy. To deal with this issue, we had practice blocks through the session allowing participants to try new strategies with no consequences, and the experimenter encouraged participants to do so. In Experiment 1, we had used all possible combinations of two key presses because we expected that the amount of motor noise for some of these combinations would be greater than for other combinations of key presses, so we had wanted to see a participant adapt to different levels of noise. However, there was no systematic difference in motor-noise level across these different combination key-presses, so we collapsed these different key presses into just one combination that participants do throughout the session.

2.3.1 Method

Participants. Twenty-four undergraduate students from the University of Michigan were run in return for monetary compensation. No participant was also in Experiment 1. Par-

ticipants received an $8 base payment, along with a bonus depending on performance (see below for details). Participants were aware of how points translated into extra money.

Apparatus.   The apparatus was the same as in Experiment 1.

Experimental Design.   Each participant was run for 1080 trials, divided into 18 blocks, each with 60 trials. These 18 blocks were organized into 6 superblocks made up of 3 blocks each, with the first block being a practice block where points did not count, and the next two being real blocks where points did count. So, in total, there were six practice blocks interspersed throughout the session, and twelve test blocks. Participants were told beforehand whether the block was practice or not, and given oral encouragement from the experimenter to try out different strategies before practice blocks.
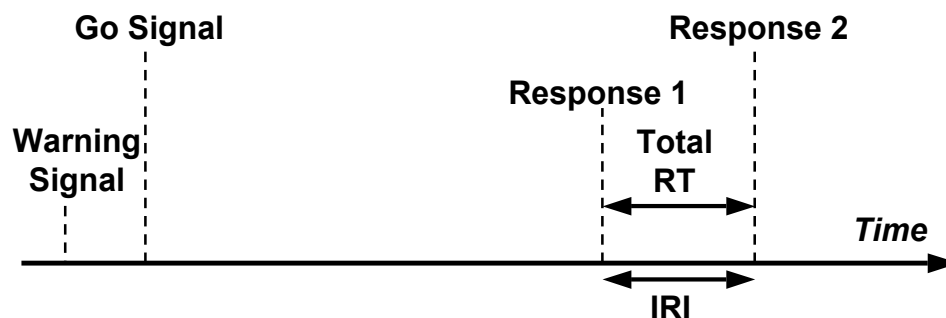
Figure 2.5. Timeline for a typical trial in Experiment 2.

The task, as in Experiment 1, was a simple reaction-time task. On every trial as soon as

24

Optimal IRI as a function of SD of IRI
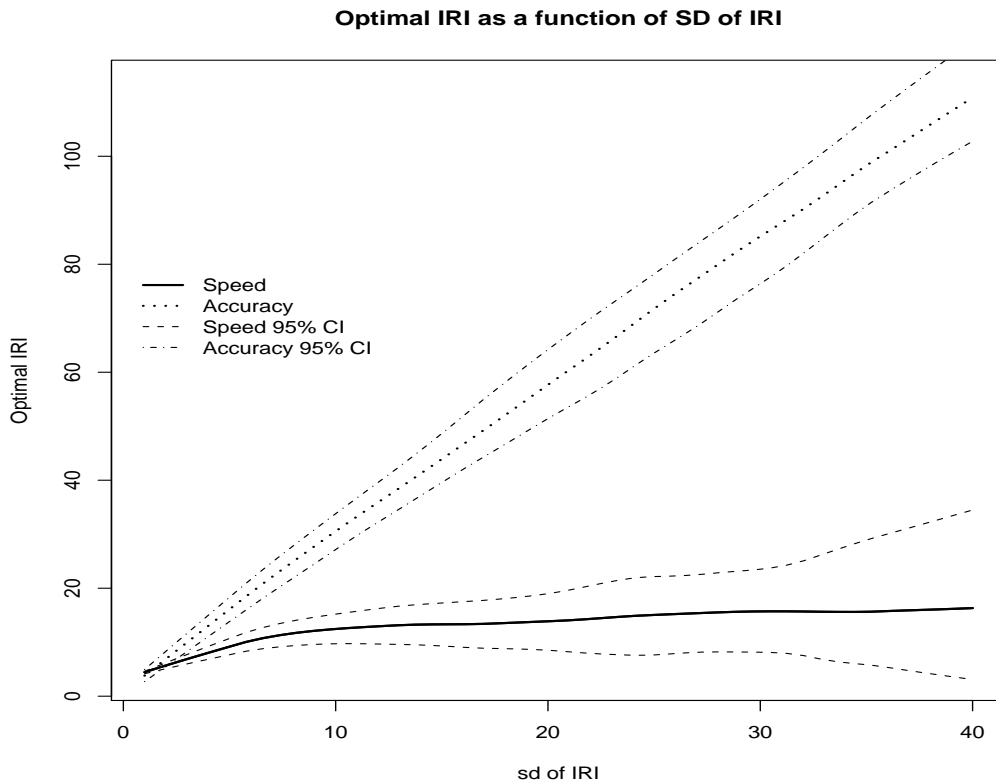
Figure 2.6. Optimal mean IRI as a function of standard deviation of IRI.

they saw the word "GO!" on screen, participants had to first press their right middle finger and then press their left index finger. Participants were informed in written instructions and by the experimenter that performance was based solely on the time and accuracy between their first and second key press. Hence, as in Figure 2.5, Total RT and IRI are the exact same, as the time before the key press does not affect the points earned.

We used a 2x2 between-subject design, manipulating payoff and external noise across participants, and there were 6 participants in each condition. All participants were given the same instructions and told that their goal was to earn as many points as possible, along with a monetary reward depending on how many points they earned.

Payoff: We used two payoff schemes, which we refer to as a Speed and Accuracy Payoff.

25

$$(2.1) \quad \underline{\text{Speed payoff:}} \ Points(IRI) = \begin{cases} (100 - IRI/5)^4/1000000 & 0 \leq IRI < 25 \\ (100 - IRI/5)^2/1000 & IRI \geq 25 \\ 0 & IRI < 0 \end{cases}$$

In words, the Speed Payoff was a step function such that participants averaged around 80 points if their IRI was less than 25 msec, and they averaged around 8 points if their IRI was greater than 25 msec. If they made a response-reversal error, they received 0 points. We used a step function because this meant a steep drop in points if participants took longer than the deadline between responses.

$$(2.2) \quad \underline{\text{Accuracy payoff:}} \ Points(IRI) = \begin{cases} (200 - IRI/10) & IRI \geq 0 \\ -600 & IRI < 0 \end{cases}$$

In words, the Accuracy Payoff was a linear function with a steep punishment for a response-reversal error. We used a linear function because this meant that participants only lost one point for every 10 msec of IRI, so they were not heavily punished for taking a little longer between responses.

External Noise: In order for us to manipulate noise levels to see how well participants can adapt to different amounts of noise, we added a random number to participants' observed IRI, generated from a Gaussian distribution with mean 0 and a pre-determined standard deviation depending on whether participants were in the high or low-noise condition. Participants received feedback and points based on this adapted, noisy IRI. For example, if a participant correctly responded with an IRI of 25 msec, then we would take that 25 and add a random number generated from a Gaussian distribution. The new sum would be used for feedback and points purposes. In the low-noise condition, the random number

was generated from a Gaussian distribution with mean 0 and standard deviation 10. In the high-noise condition, the random number was generated from a Gaussian distribution with mean 0 and standard deviation 60. All participants knew that there might be some external noise, but participants did not know how the noise was generated or which condition they were in.

Procedure.   The procedure was the same as in Experiment 1, except that on every trial, participants had to first press their right middle finger and then their left index finger. Also, participants first had one block of 60 trials that was practice, so that the points did not count toward their final total. This was followed by two blocks each of 60 trials where the points counted, then another practice block, then two real blocks, etc.

All participants were given the same instructions, and were never told what condition they were in. Before every practice block (except the first one), participants were encouraged to try different strategies. The experimenter told participants they could try shorter or longer IRIs to see if their averaged points increased or not. The experimenter never told participants the details of the payoff function.

Feedback was the same as in Experiment 1, except that at the end of a block, participants were told their average IRI (not their average total RT) on the previous block, their average points per trial, and their total points.

2.3.2   Results

For all results, we only used the final 6 test blocks of data. This way participants not only had time to get used to the task, but had four practice blocks to settle on a strategy they liked. This was important for data analyses, because if participants were changing strategies, this would result in a seemingly large standard deviation of IRI, which is our proxy statistic for internal motor noise. We also removed any trials where the Inter-Response Interval (IRI) was more than 3 standard deviations away from the mean IRI. These trials were less than

27

1% of all trials. All subsequent analyses were done on mean IRI.

Group	Mean-IRI (msec)	sd.IRI (msec)	Accuracy - Internal Noise Only (%)	Accuracy - With External Noise (%)
Accuracy-Payoff High-Noise	106	21.14	100	95.17
Accuracy-Payoff Low-Noise	45	15.30	99.83	99.00
Speed-Payoff High-Noise	63	19.07	99.00	82.67
Speed-Payoff Low-Noise	20	14.88	88.17	78.67
Overall	58	17.60	96.75	88.88

Table 2.2. Comparison of mean IRIs across payoffs.

The overall omnibus F-test revealed a significant difference in IRIs across our four groups, $F(3, 20) = 25.96$, $p < .001$. We then computed contrasts across all four groups and found that the mean IRI for all participants in the Accuracy Payoff, 75.47 msec, was significantly greater than the mean IRI for participants in the Speed Payoff, 41.31 msec, $t(20) = 4.778$, $p < .001$. The fact that accuracy-payoff IRI was longer than in the speed-payoff IRI was as we predicted. See Table 2.2 for each group's mean IRI.

We also computed a contrast to examine whether participants adapted to the level of external noise. We found that the mean IRI of the high-noise participants was 84.49 msec, significantly larger than that for the low-noise participants, 32.29 msec, $t(20) = 7.30$, $p < .001$. For participants in both the speed and accuracy payoffs, we also computed contrasts showing that high-noise participants differed from low-noise participants, for speed $t(20) = 4.22$, $p < .001$, and for accuracy $t(20) = 6.104$, $p < .001$.

We also computed contrasts that compared accuracy-payoff low-noise participants with speed-payoff low-noise participants, to make sure that when controlling for external noise,

we still found differences in payoffs. For high-noise, we found that accuracy mean IRI was significantly greater than speed mean IRI, $t(20) = 4.32$, $p < .001$, and for low-noise, we also found this difference, although not as strongly $t(20) = 2.44$, $p < .05$.

We also tested whether error rate differed across conditions, by doing an arcsin transformation of the square root of the accuracy percentage. We first analyzed those errors that were caused only by internal noise, i.e. those cases where the keys were actually pressed in the wrong order. Not surprisingly, we found that the accuracy-payoff participants were more accurate than the speed-payoff participants, $t(20) = 12.46$, $p < .001$. However, we did not find a difference in the high-noise participants (88.92%) compared to the low-noise participants (88.83%), $t(20) = -.921$, ns. Nor was the accuracy different for participants in the Speed Payoff at the different noise levels.

We also examined accuracy after factoring in external noise, i.e. cases where the feedback told the participant he made an error, regardless of whether the participant actually pressed the keys in the wrong order or not. Once again, we found that accuracy participants (97.08%) were more accurate than speed participants (80.67%), $t(20) = 9.414$, $p < .001$.

We calculated optimal IRI for each participant. To calculate optimal IRI, we generated 500 points from a Normal Distribution with mean 1 and a standard deviation that was the same as each participant's standard deviation of IRI. To factor in the external noise, we added to this distribution 500 points generated from another Normal Distribution with mean 0 and standard deviation of 10 or 60, depending on the condition that each participant was in. Summing these distribution gave us a distribution of IRIs, and we could then use our external payoff to figure out how many points on average the model would earn with that distribution of IRIs. We then repeated this process by using 500 points generated from a Normal Distribution with mean 2, then a Normal Distribution with mean 3, etc. up to about mean 200. Whichever mean resulted in the most points on average was considered "optimal."

Along with finding the differences we predicted in mean IRI, we again computed Opti-

Actual IRI – Optimal IRI, By Condition
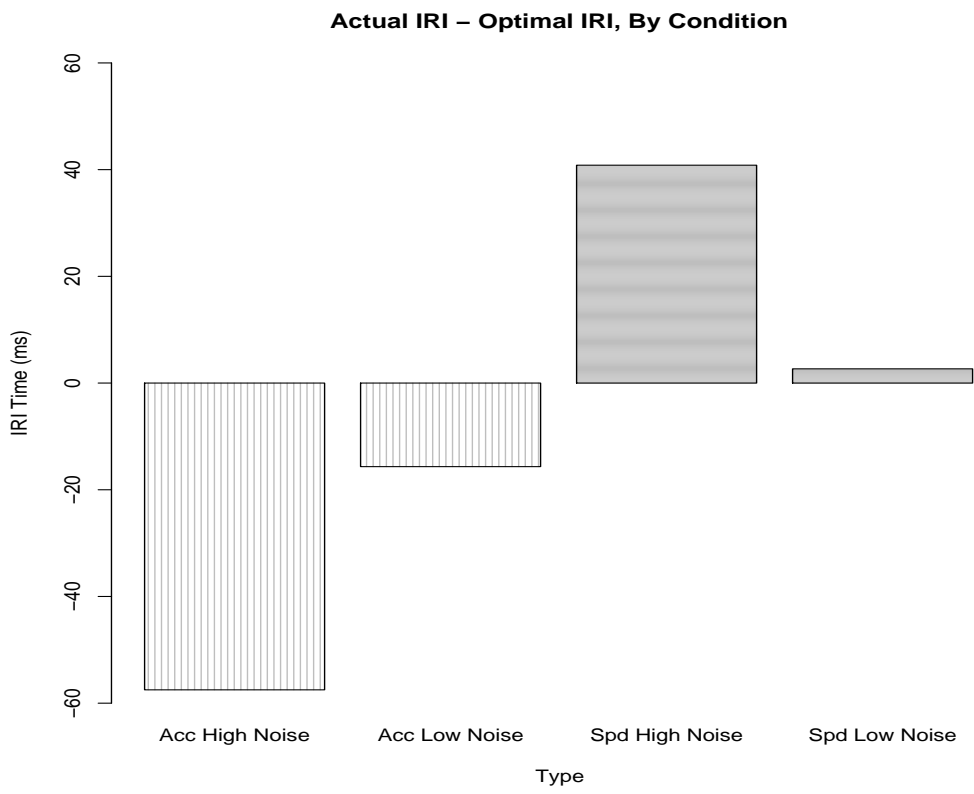
Figure 2.7. Relative difference from optimal IRI by condition.
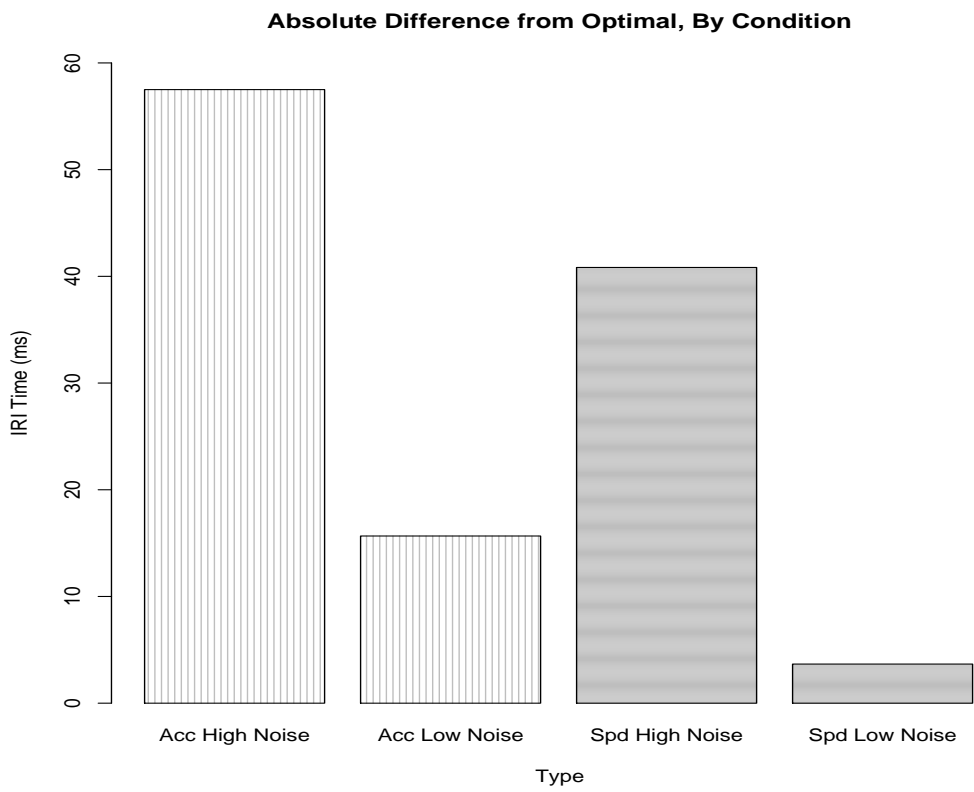
Figure 2.8. Absolute difference from optimal IRI by condition.

mal IRI via Monte-Carlo models, because we have a well-defined payoff function that we can use to compute optimal IRI, we have a proxy measure of a participant's internal noise (their standard deviation of IRI), and we know their external noise. Figure 2.7 has a barplot averaged over participants in each group showing how much each group differs from optimal. We can see that participants in the low-noise condition were closer to optimal than participants in the high-noise condition. Also, we see the effect of payoff. Accuracy-payoff participants tended to be too fast compared to optimal, whereas speed participants tended to be too slow (especially in the high-noise group).

2.3.3 Discussion

We found that participants did adapt to the payoff, as we expected, and that adding external noise had an effect on performance. There were two major drawbacks in Experiment 2, though. For one thing, it is unclear how total noise level affected how well participants could find an optimal strategy. At 60 msec of external noise, participants could not find an optimal IRI, at 10 msec they came close. It is unclear if total noise works like a staircase function, whereby below a certain level of noise people can find optimal or near-optimal mean IRI, or it works more like a monotonic function, whereby as the level of total noise increases the difference from optimality also increases. In order to help elucidate this, we need to test people at a level of noise in-between the two extremes we have.

The other drawback is that even our low-noise condition includes some external noise. Although participants in this condition did adapt quite well, they significantly differed from optimality in the accuracy-payoff condition. It is possible that giving no external noise would allow participants to adapt perfectly. People have had countless hours to get used to their own level of internal motor noise, so may be able to adapt extraordinarily well to that. But most people have very little experience dealing with external noise, so are not as adroit at adapting to that. Further, most cognitive experiments do not require participants to adapt to any sort of external noise, only their own cognitive, perceptual and motor pro-

cesses. Hence we thought it important to establish how well participants could adapt in a no-external noise condition, as this most closely mirrors both their own experiences and how well they could perform in typical experiments.

2.4   Experiment 3

In Experiment 3 we replicated our results in Experiment 2, but with different levels of external noise. Because we felt participants have most experience adapting just to their own internal motor noise, we included a condition with no external noise. We also included a medium-noise condition because we were worried that the high-noise condition in Experiment 2 had so much noise, participants could not discern what was optimal in the limited number of trials they had.

2.4.1   Method

Participants.   Twenty-five undergraduate students from the University of Michigan were run in return for monetary compensation, none of whom were run in the previous experiments. One participant was removed for failing to obey instructions early in the experiment (by using the wrong fingers on the finger keys).

Apparatus.   The apparatus was the same as in Experiments 1 and 2.

Experimental Design.   The design was the same as in Experiment 2, except for two different levels of external noise. In the no-noise condition, participants did not have a random number added to their IRI. In the medium-noise condition, participants IRI included a random number generated from a Gaussian Distribution with mean 0 and standard deviation of 25.

Procedure. The procedure was the same as in Experiment 2. Note that even in the no-noise condition, participants were still instructed that there would be external noise affecting their performance, so that we could give them the same instructions we gave all other participants.

2.4.2 Results

All results, as in Experiment 2, came from only the final six real blocks.

The overall omnibus F-test revealed a significant difference in IRIs across our conditions, $F(3, 20) = 25.274$, $p < .001$. We then computed contrasts across all four groups and found that the mean Inter-Response Interval (IRI) for all participants in the Accuracy Payoff, 51.55 msec, was significantly greater than the mean IRI for participants in the Speed Payoff, 22.12 msec, $t(20) = 6.37$, $p < .001$. Accuracy-payoff IRI was longer than that of participants in the speed-payoff IRI, as predicted. See Table 2.3 for each group's mean IRI.

Group	Mean-IRI (msec)	sd.IRI (msec)	Accuracy - Internal Noise Only (%)	Accuracy - With External Noise (%)
Accuracy-Payoff Medium-Noise	69	17.39	100	98.50
Accuracy-Payoff No-Noise	34	16.12	97.67	97.67
Speed-Payoff Medium-Noise	30	17.58	92.00	77.67
Speed-Payoff No-Noise	14	10.60	79.17	79.17
Overall	34.50	15.42	92.21	88.25

Table 2.3. Comparison of mean IRIs across payoffs.

We also computed a contrast to examine whether participants again adapted to the level

of external noise. We found that the mean IRI of the medium-noise participants was 49.61 msec, significantly longer than that for the no-noise participants, 24.06 msec, t(20) = 5.53, p < .001.

For groups with the same level of external noise, but different payoffs, we computed contrasts that showed within either payoff, medium-noise participants had longer IRIs than no-noise participants. For participants in the speed condition, medium-noise mean IRI of 30 was significantly slower than the no-noise mean IRI of 14, t(20) = 2.32, p < .005, for accuracy, medium-noise mean IRI of 69 was greater than no-noise mean of 34 msec, t(20) = 5.59, p < .001.

We also computed contrasts that compared accuracy-payoff no-noise participants with speed-payoff no-noise participants, to make sure that when controlling for external noise, we still found differences in payoffs. For medium-noise, we found that accuracy-payoff mean IRI of 69 was significantly longer than speed-payoff mean IRI of 30 msec, t(20) = 6.03, p < .001, and for no-noise, we also found that accuracy-payoff no-noise mean IRI of 34 msec was significantly longer than the mean speed-payoff no-noise mean IRI of 14 msec, t(20) = 2.98, p < .01.

Once again, accuracy based solely on internal motor noise differed across payoffs, after doing the arc-sine transform, with accuracy-payoff participants being more accurate than speed-payoff participants, t(20) = 8.12, p < . 001.

We also got similar results as in Experiment 2 for accuracy after factoring in external noise. Participants who had the Accuracy Payoff were more accurate (98.08%) than participants in the Speed Payoff (78.42%), t(20) = 9.726, p < .001. Once again, the effect of noise was not significant, as medium-noise participants had accuracy of 88.08%, compared to 88.42% in the no-noise condition, t(20) = .238, ns. Accuracy did not differ across noise level for the two speed-payoff groups or for the two accuracy-payoff groups.

Optimal IRIs were calculated as in Experiment 2, except that the external noise was now set to be 0 or 25. Figure 2.9 shows the difference between actual IRI and optimal

Actual IRI – Optimal IRI, By Condition
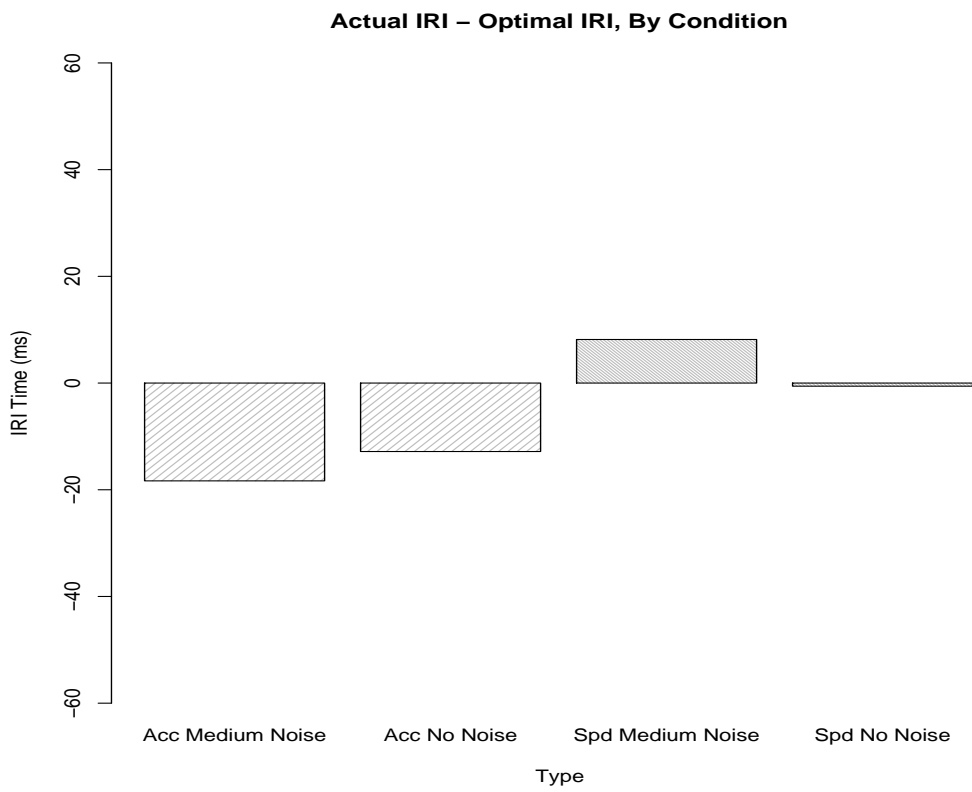
Figure 2.9. Relative difference from optimal IRI by condition.

Absolute Difference from Optimal, By Condition
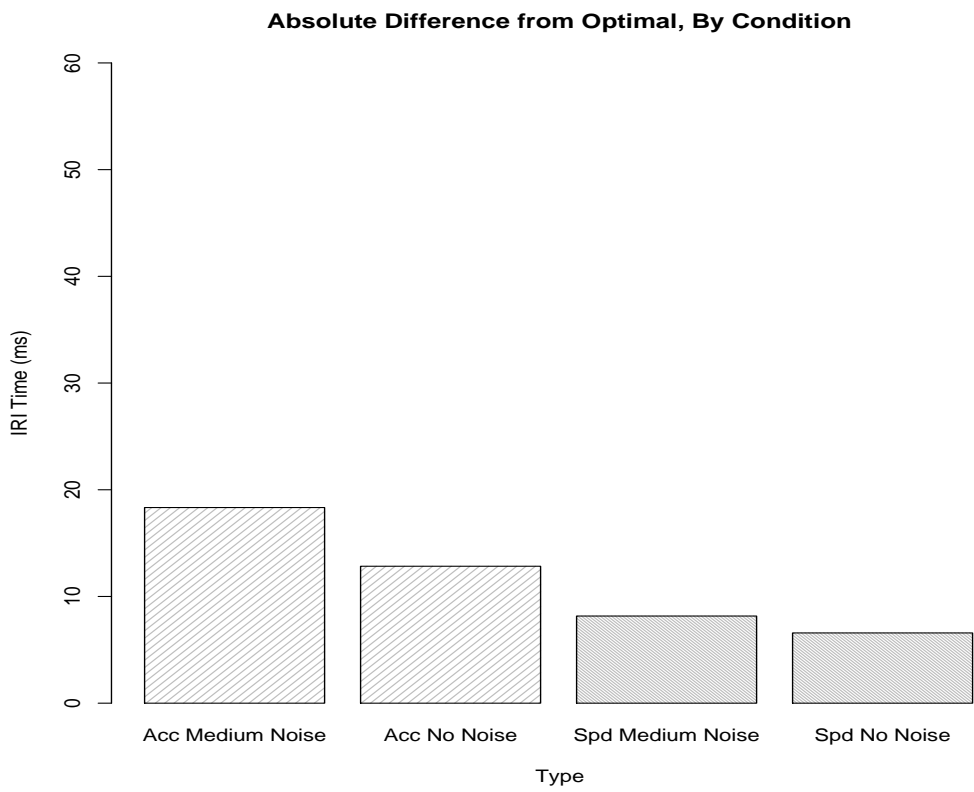
Figure 2.10. Absolute difference from optimal IRI by condition.

IRI in a barplot averaged over participants in each group. We can see that participants in the no-noise condition were again closer to optimal than participants in the medium-noise condition, although they both did very well. Generally all participants regardless of the group were within 20 msec of optimal IRI. Accuracy participants again had too short IRIs, and speed medium-noise participants had too long ones.

2.5   Discussion and Results Across Experiments 2 and 3

Across these two experiments, we hypothesized that participants could adapt optimally. Although evidence supporting this was reported above, we decided to do some further tests to see just how close to optimal participants were. The most straightforward test was just to see if the mean of the difference from optimal differed from 0. We found that in the Speed No-Noise condition, the difference of 0.58 msec from optimality was not reliable ($t(5) = .15$, $p > .8$), nor was there a reliable difference in the speed low-noise condition ($M = -2.67$, $t(5) = 1.44$, $p > .2$). In all other conditions, this mean did significantly differ from 0.

Although this may seem disappointing, before blindly defining significance based on p-values, we have to take into account effect sizes. For example, in the speed medium-noise condition, we found a reliable difference from 0 ($M = -8.17$, $t(5) = -2.81$, $p < .05$). However, if this mean only differed by one msec, so that the mean difference from optimal IRI was -7.17, this difference would no longer be significant $t(5) = -2.47$, $p = .06$. We have to be careful of over interpreting a null effect, but we clearly did not find strong evidence supporting the claim that the mean IRIs of these participants significantly differed from 0, except in the high-noise condition ($M = -40.83$.

For the accuracy payoffs, we initially found all conditions significantly differed from 0. But again, the effect sizes were usually very small. In the no-noise condition, the mean difference from optimality of 12.83 was highly significant ($t(5) = 4.49$, $p < .01$). However, if we added 6 msec to the mean IRI for each participant, this was no longer significantly

Figure 2.11. Individual participant's IRI as a function of standard deviation of IRI in the low-external-noise, accuracy-payoff condition. The vertical black line is the participant's 95% CI for optimal IRI

different. In the low-noise condition, we need only have added 4 msec to the mean IRI. In the medium-noise condition, we would have to add 13 msec to no longer significantly differ; in the high-noise condition 30 msec.

Another concern for us was to ensure that it was non-trivial to perform optimally. As Roberts & Pashler (2000) point out, a good fit is not enough, if the range that qualifies as a good fit is quite broad. Figure 2.11 shows an individual participant in the low-noise condition from the Accuracy Payoff. The vertical arrows are his actual IRI with his 95% confidence interval. The dashed lines represent the 95% confidence band to perform optimally. One can see that if he performed only 10 msec faster or slower, he would no longer be optimal. Yet given his internal noise level (on the x-axis), he essential achieved the optimal IRI. Given how narrow the band is, this is impressive performance and not likely to

Opt IRI as a function of IRI.SD , Monte Carlo Trials=500, IRI sd=20

Figure 2.12. Individual participant's IRI as a function of standard deviation of IRI in the high-external-noise, accuracy-payoff condition. The vertical black line is the participant's 95% CI for optimal IRI.

have resulted just by chance.

Participants in the high-noise condition had mean IRIs that were too short, as can seen by the typical participant in Figure 2.12. This participant had an actual IRI of about 120 msec, but optimal would have been 30 msec or so longer than that. Still, we notice that he was at least heading in the right direction, with a longer IRI than the typical participant in the low-noise condition. Accuracy participants should be cautious about making errors, because they are so severely penalized. Yet participants in the accuracy high-noise condition did not have 100% accuracy once we factor in the external noise, but only around 95%. Optimally, then, these participants would have longer IRIs and make fewer errors. However, it does seem natural that participants, over the course of hundreds of trials, would wonder if they could have shorter IRIs and maintain their accuracy. Of course, shorter IRIs

result in more points until they make an error. This phenomenon is well known in the reinforcement-learning literature as the difference in exploration versus exploitation (Sutton & Barto, 1998). Participants who explore their strategy space at all can only discover they are going too fast by making an error, which will result in sub-optimal performance. Hence participant desire to explore the strategy space may be the reason we never got accuracy of 100% factoring in external noise.

For the speed-payoff, we found that in the high-noise condition, participants had IRIs that were too long, despite a fairly wide band around optimal. In the low-noise condition, like in Figure 2.13, there is a very narrow band around optimal IRI, but four of the six participants fell right within that narrow band[1].

Figure 2.14 examines the relationship between optimality and total noise people must adapt to. Total noise was simply the summation of people's internal motor noise and the external noise we added. The data comes from individual subjects in Experiments 2 and 3.

The first thing to notice in this figure is how a horizontal line right around 0 was the best fit (technically, the best-fitting line had an intercept of 5.4 and a slope of 0.02, but neither parameter significantly differed from 0, $pr > .5$ for both parameters); hence on average people were optimal. Of course, closer inspection reveals that people in the Accuracy Payoff tend to have a positive difference between optimal IRI and actual IRI, indicating their actual IRI is too short. This can be seen in the behavioral data as well.

Conversely, participants in the speed-condition tend to be negative (or around 0) in Figure 2.14. This indicates their actual IRI is too long. Generally, participants in the speed-condition were close to mathematically optimal, as their difference from optimal IRI did not differ from 0 in the no and low-noise condition, and only marginally so in the medium-noise condition. In the high-noise condition for the speed-participants, the noise was too much for participants to adapt optimally. All of these participants noticed that sometimes they earned a lot of points, but no participant was able to figure out the pattern that would

---

[1]Of the other two participants, one was slightly too fast to fall within the band; the other participant was about 20 msec too fast

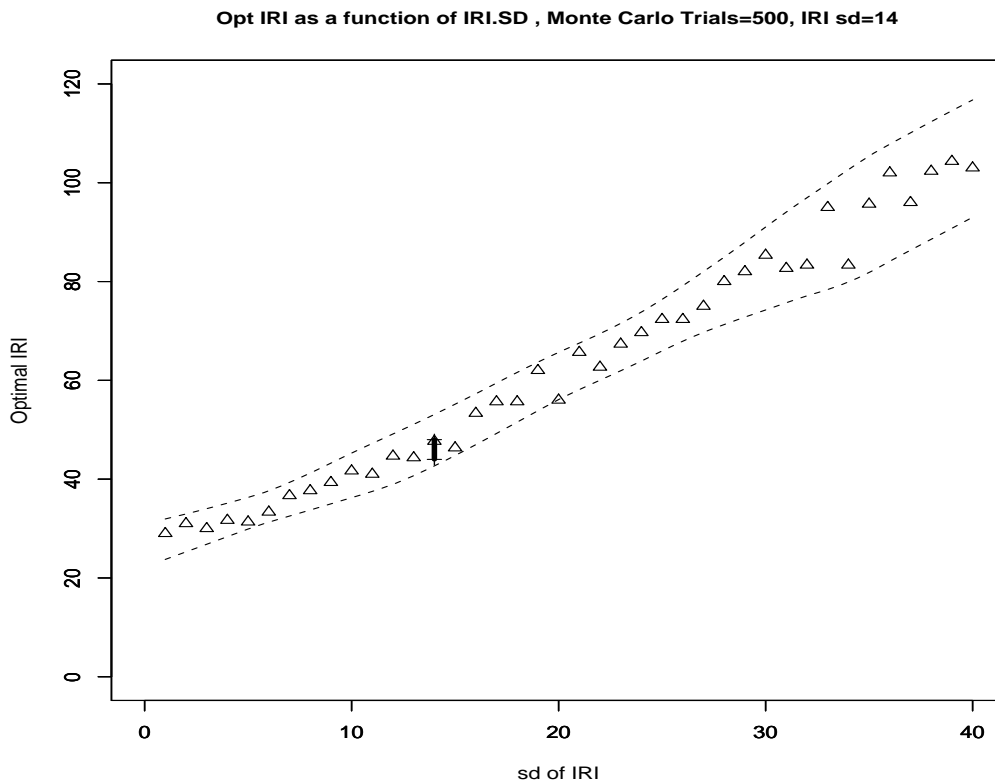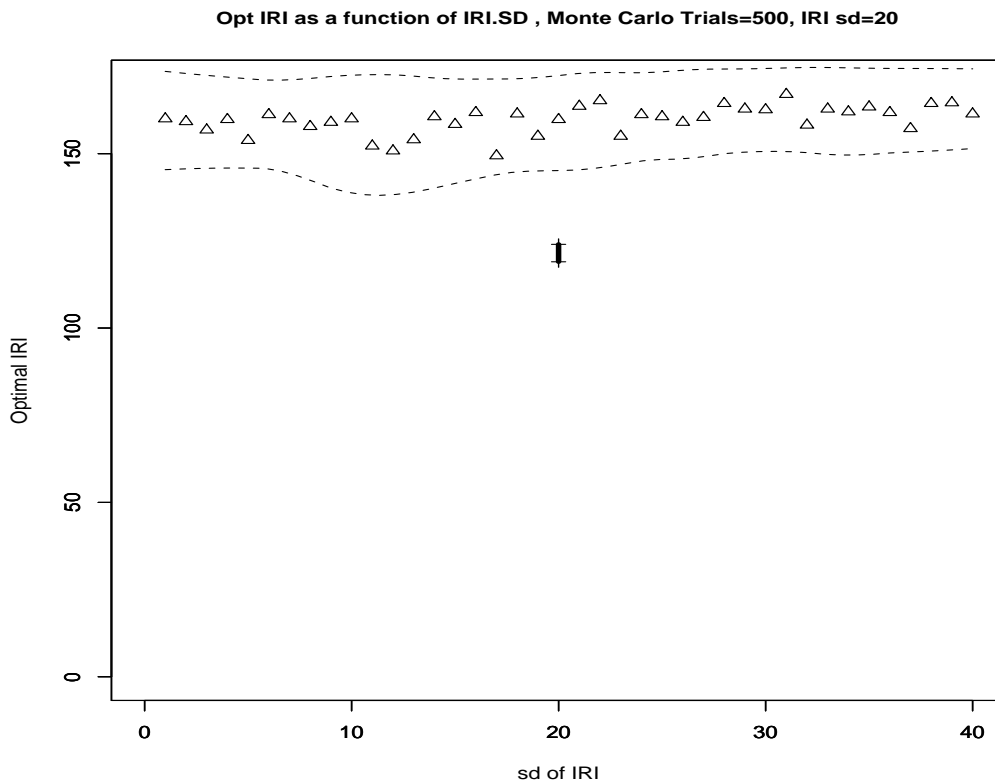**Opt IRI as a function of IRI.SD , Monte Carlo Trials=500, IRI sd=35**
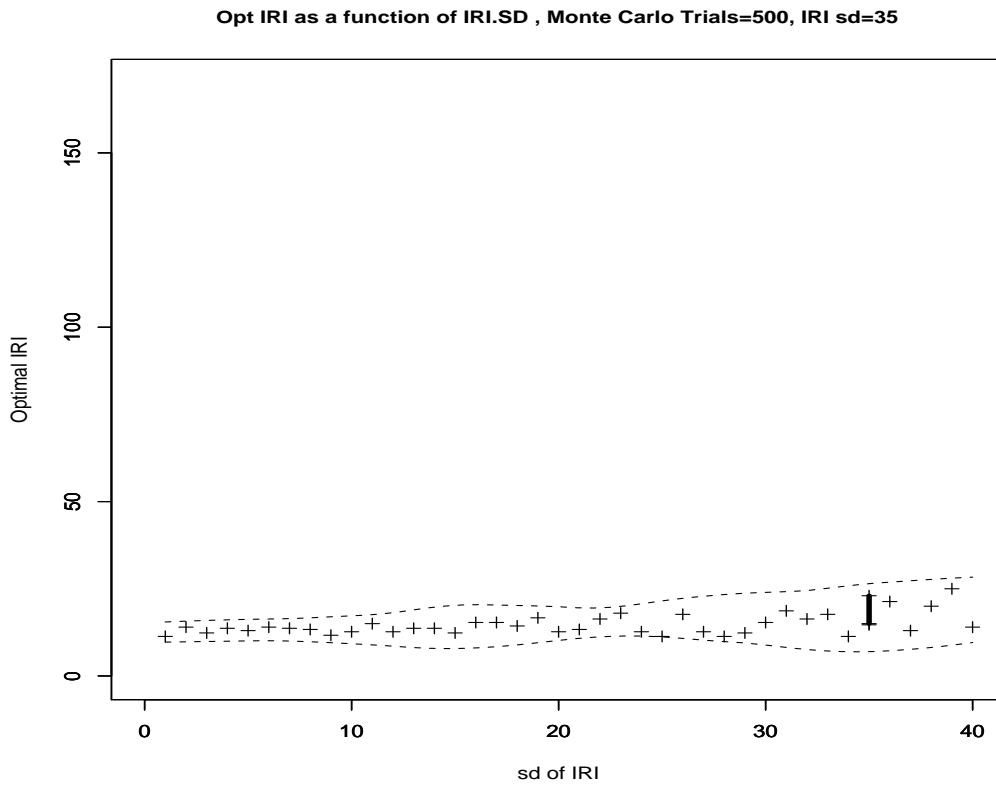
Figure 2.13. Individual Participant's IRI as a function of standard deviation of IRI in the low-external-noise, speed-payoff condition. The vertical black line is the participant's 95% CI for optimal IRI.

Difference from Optimal IRI as function of Total Noise
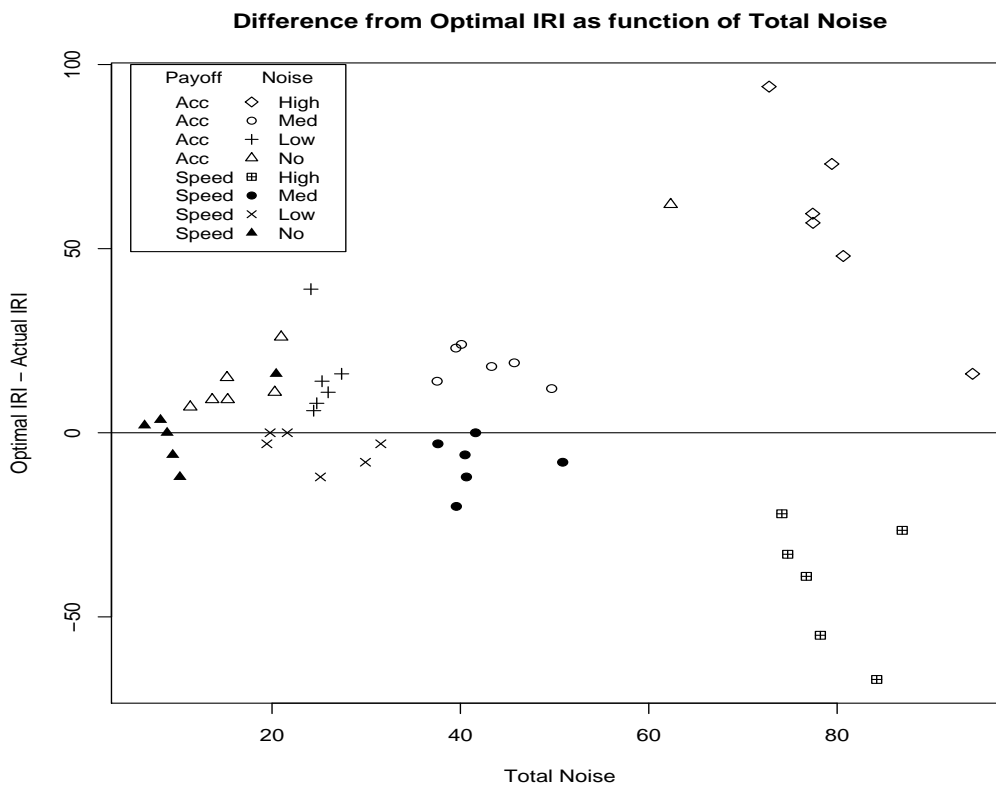
Figure 2.14. Individual participants difference from optimal IRI by condition.

result in optimal points.

2.6 General Discussion

Across our three experiments, we found clear evidence that people adapt so as to maximize the points they earn, by factoring in both an external payoff and their internal motor noise, even in tasks where IRIs are as little as 15 msec and no more than 150 msec. In both Experiments 2 and 3, we found a highly significant adaptation in the expected direction depending on the external payoff. We also formed 0-parameter mathematical models using just the external payoff and participants' level of internal motor noise and external experimentally-controlled noise, and made models that did a good job of explaining the human results. In cases with relatively low levels of total noise, we found that the average difference from mathematically-optimal performance was not significant. In a few other cases, when we we did find a statistically-significant difference, this would not have occurred if the average was just slightly different. In fact, due to possible measurement errors, it is probably best not to overemphasize the importance of any difference from optimal that is less than 10 msec.

We found that almost all participants had roughly similar amounts of internal motor noise in this simple-IRI task, making individual differences difficult to analyze. Although we were somewhat surprised by this initially, as evidenced by the fact that we did not think to add external noise in Experiment 1, the fact that all of our participants were college graduates who presumably have had numerous hours of practice with typing and text messaging may explain this result. Participants who have to type letters in a particular order have had extensive practice in doing an IRI task, so we may have found the limitation where internal motor noise cannot be reduced further.

By experimentally manipulating external noise, we managed to find the range where people are no longer able to achieve optimal strategy. Our participants adapted optimally

or near-so up until 25 msec of external noise was added to the internal motor noise, creating up to 45 msec of total noise participants had to adapt to. However, when we added 60 msec of noise, participants were no longer able to adapt optimally. Somewhere between 45 and 80 msec of total noise participants could no longer properly adapt in the allotted number of trials. We do not know if additional trials would have eventually allowed participants to adapt better, or if they had already reached an asymptotic performance level.

Our results provide us with a clear guideline for determining how people strategize: people's goal is to choose the strategy that allows them to earn the most money. Often finding this strategy is a non-trivial task, and it requires encouraging people to explore the strategy space, then giving them feedback so they can assess how well a given strategy works, and enough trials given the variability within a given strategy to ensure they reach the proper conclusion for each strategy. Nonetheless, it is crucial to establish what is the proximal goal in choosing a strategy if we want to understand human cognitive limitations. If a strategy is clearly optimal in terms of how many points participants will earn, but participants do not choose it, this may be a sign that there is some limitation in the mind preventing that strategy from being chosen.

However, we have to be careful in interpreting the lack of some optimal strategic decision as indicating a cognitive limitation. As we saw in the high-noise conditions, participants did not choose the optimal strategy. With enough noise, participants will not be able to discern whether one strategy is better or worse than another strategy, at least in a reasonable number of trials. We also have to be careful about participants finding local maximums in the strategy space. If a participant finds that Strategy A is better than Strategy B, a participant may just continue using that strategy throughout the experiment, never discovering that Strategy C would have been even better than both A and B. Hence it is very important to encourage participants to fully explore the strategy space, as we did in Experiments 2 and 3 via practice blocks and by having the experimenter encourage participants to try something different on these blocks.

An important result of this work is to show how important it is to use a point-system to try to restrict a participant's strategic choices. It may be that if you use participants who are in the experiment for course credit or for an hourly rate of pay, they can adopt any number of strategies that will maximize their goal. If for course credit, the participant may choose the strategy that lets them finish the experiment the fastest. If paid by the hour only, a participant may adopt the strategy that maximizes how much they earn, which may require them to go as slow as possible! Hence it may be impossible to interpret a participant's results unless we somehow encourage participants to perform in a specifically optimal way.

Although we used a point system where the points translated into money, there is some reason to believe the point system alone may be sufficient. For one thing, video games for years have used points to encourage performance by participants, even though often the points do not matter monetarily. Another reason points alone may be sufficient is that quite a few participants reported to the experimenter that they really enjoyed this experiment and felt that they were competing against themselves, by continually trying to improve on every block. It is thus clear that feedback is critically important in guiding performance, and in order to maximize this competition with oneself, in later experiments in this dissertation, we included feedback that showed how participants performed across different blocks, so they could see if they had improved or not.

The fact that we can now deduce that maximizing points guides what strategy a participant chooses, will be of crucial importance in Chapter 3, where we want to compare different computational models of human performance. In order to compare different models, we need some criterion, and that criterion is points earned. We can thus deduce what models maximize points, and compare those results to human performance.

There are some issues still to be resolved from these experiments. If we control for optimal performance, i.e. if two different strategies generate similar performance, is there any way we can determine which strategies participants would choose? Is the soft-constraints

hypothesis (Gray et al., 2006) correct, so that participants here would choose whichever strategy minimizes time to perform the task? Do participants instead try to minimize cognitive load? It also remains to introduce ideas from reinforcement learning to these types of basic tasks, to determine whether optimal performance will occur regardless of noise level if given enough trials, and to see whether human learning in these tasks matches reinforcement-learning models.

Chapter 3

Executive Control of Task Interruptions

3.1 Introduction

Imagine you are a pilot, engaged in the complex task of flying your plane. Suddenly, an alarm sounds, and you have to immediately focus your attention on averting a potential disaster by dealing with whatever caused the interruption to occur. How is this interruption going to affect you? Will you be able to resume flying the plane as soon as you are done dealing with the alarm? Will there be a delay before you can resume flying? What about dealing with this emergency - will you be slower because you are also processing information about the airplane you are flying? Do you retain your memories of the logistic details for the planes you are flying, its altitude and its heading, for example? This chapter helps provide answers to these questions.

Most people deal with interruptions all the time, albeit not in times where every millisecond is critical. A common type of interruption involves cases where you can finish some aspect of your initial task before responding to the interruption, such as with a telephone call. However, interruptions that involve immediate processing are also common in everyday life, including responding to a car horn while driving, a person tapping you on the shoulder, or a fire alarm sounding.

In this chapter, we will introduce a paradigm for studying task interruptions that require immediate attention and a theoretical interpretation of what occurs during these interruptions. We will begin by reviewing some of the major findings and procedures from other areas of multiple-task performance - such as task switching and perfect time-sharing, in part because these procedures can be modified to study task interruptions as well. In fact, the procedure we employed for task interruptions can be thought of as the diametric opposite procedure to that most common in task-switching, i.e. the opposite of the standard PRP procedure, or an Anti-PRP procedure.

Although there have been many studies on multiple-task performance, there have been relatively few studies of how we deal with interruptions. Many of these existing studies are naturalistic ones done by observing workers in their offices (e.g., Chisholm et al., 2001). There have also been studies about the effects of interruptions on higher-level cognitive processes such as decision-making (e.g., Speier et al., 1999). These studies, although useful for ergonomic purposes and helping workplace efficiency, do not help us understand the basic cognitive processes that occur when we have to respond to an immediate interruption.

Altmann and Trafton (2002) have begun to examine task interruptions in more depth. They considered task interruptions to involve activating the proper goal. For example, if you are doing Task 1 and then are interrupted by Task 2, you will have to activate the proper goal for Task 2 before you can start that task. They further suppose that activating a different goal takes time, a function of how often and how frequently that goal is used, as explained by the ACT-R cognitive architecture (Anderson & Lebiere, 1998). Given their theory, two hypotheses immediately follow. One hypothesis is that there will be an "interruption lag" when the goal for the interrupting task must be activated and the goal for the initial task must be suspended (Trafton et al., 2003; Altmann & Trafton, 2007). During this interruption lag, a person can engage in different strategies, such as trying to remember what they were doing in the initial task (retrospective rehearsal), or trying to reach a critical point in the initial task (see Seifert et al., 1994, for a discussion on

critical junctures in a task) so that when they suspend this task, they can look ahead for how to proceed (prospective goal encoding). The second hypothesis is that there will be a "resumption lag" such that after finishing the interrupting task, they will have to retrieve the goal for the initial task (Altmann & Trafton, 2004, 2007). They have found evidence supporting both of these hypotheses.

However, their tasks were relatively complicated, taking on the order of seconds to perform. The fact that they were so complicated makes it difficult to break them down into stages and understand the basic cognitive processes affected by interruptions. Further, the fact that the tasks were so complicated may lead different participants to develop different strategies for dealing with interruptions, making it hard to infer if these lags are unavoidable or instead caused by strategic considerations.

A related area of multiple-task performance where the basic cognitive processes have been studied is task switching. Task switching occurs when you have to complete two tasks in close temporal proximity, such that after completing a first task, an individual must shortly thereafter perform another, different task. One paradigm used to study task-switching is the Psychological Refractory Period (PRP) procedure (Welford, 1952; Meyer & Kieras, 1997a). In this paradigm, a person is first presented with a stimulus corresponding to a first task, which they must try to perform quickly and accurately. After some interval, known as the stimulus onset asynchrony (SOA), a second stimulus is presented corresponding to a second task that must be performed too. The plot of the reaction time for performing the second task as a function of the SOA generates the PRP curve. A common finding is that when the SOA is very short, such that one stimulus is presented quickly after the other, the reaction time for the second task is slow, much slower than when the SOA is larger. The finding that as the SOA increases the reaction time for the second task decreases is known as the PRP effect. Figure 3.2 shows an idealized version of the PRP Effect. Note that there is an easy and hard second task, but that at the small SOA they have the exact same RT, and slowly diverge as the SOA increases, until at longer SOAs the two

plots have asymptoted at their respective mean RTs.

Task 1
Stimulus 1

Task 1
Response 1

Warning
Signal

Task 1 Reaction Time

Time

SOA
Stimulus
Onset
Asynchrony

Task 2 Reaction Time

Task 2
Stimulus 2

Task 2
Response 2
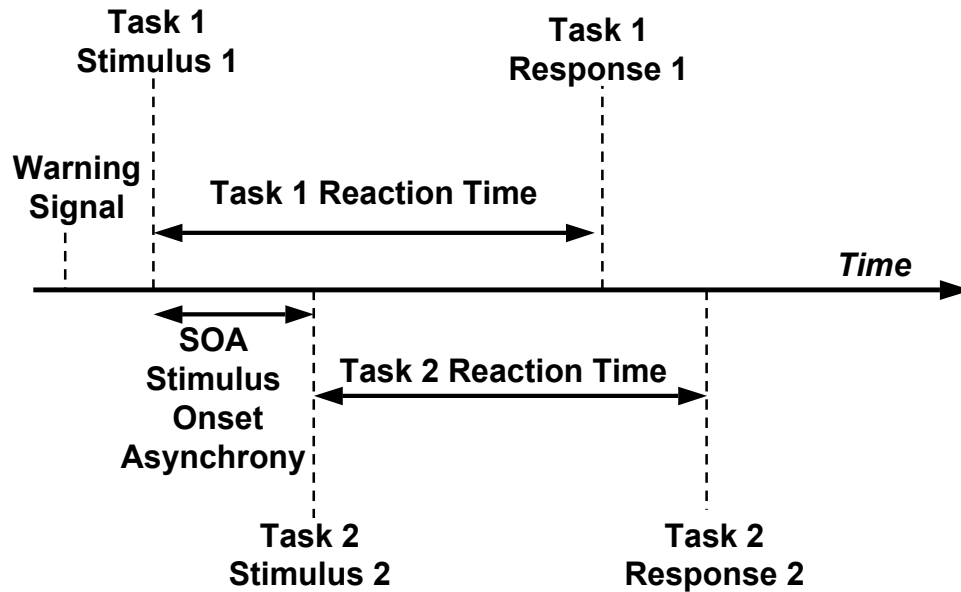
Figure 3.1. A typical trial in the PRP procedure (cf. Meyer & Kieras, 1997a).

What accounts for the PRP effect? To understand it, we must understand the discrete stage model of Saul Sternberg (1969a,b). The discrete stage model divides a task into temporally separate processing stages, including stimulus-encoding, response selection, and motor execution stages. According to Pashler (1994), the PRP effect is caused by a response-selection bottleneck (RSB), so that the response-selection stage cannot occur simultaneously for two different tasks. If so, then a person could not start selecting a response for the second task until the response was selected for the first task. When the SOA is large, the response for the first task will already have been selected before the second task begins, so there is very little or no overlap in choosing the response for the second task, resulting in no delay in completing the second task. So this bottleneck will result in increased reaction times when the SOA is small enough to allow response selection

51

to overlap, and the smaller the SOA the larger the overlap, accounting for the PRP effect.

Idealized PRP Effect With RSB
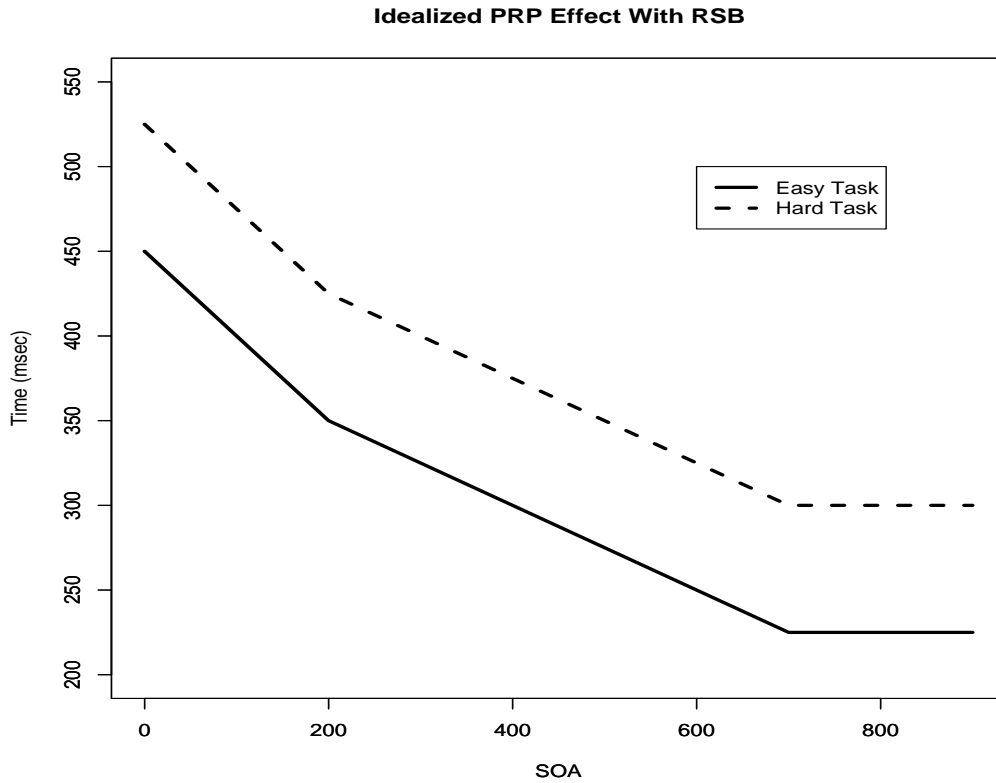
Figure 3.2. Idealized version of the PRP effect.

PRP studies usually manipulate both the SOA and the different stages of the various tasks, especially the response-selection stage. There are two major ways to change response-selection difficulty: Stimulus-Response (abbreviated S-R) numerosity and S-R compatibility. S-R numerosity refers to adjusting the number of stimuli you use, because the more stimuli the harder it is to choose a response for any one stimulus. S-R compatibility refers to making the responses less compatible with the stimulus, such as by having a spatial task where stimuli on the far left must be responded to with the right pinky finger (instead of the right index finger, which would be the most compatible response). By varying the difficultly of the first task, one can see what effect this variation has on the PRP effect. If there is a RSB, then making it more difficult to choose a response for Task 1 should have a corresponding delay in performing Task 2 at short SOAs. However, when the

SOA is longer, response-selection will be finished for Task 1 before Task 2 commences, so Task 2 RT should be the same in this case regardless of Task 1 response-selection difficulty.

Likewise in task-interruptions, it is important to manipulate the response-selection stage of the initial, interrupted task. By manipulating it, we can see whether the response-selection for the initial task affects how long it takes to do the interrupting task. If manipulating response selection does not affect how long it takes to do the interrupted task or to do the interrupting task, then this would suggest that response selection for the two tasks proceeds in parallel for the two tasks. If, instead, we find that manipulating response selection increases the interrupting-task reaction time similarly to how long the response-selection variation affects the uninterrupted early task, we will have support for a response-selection bottleneck. If there is a bottleneck, then when you are executing one response-selection stage, you have to wait for it to end before beginning the response-selection stage for the other task. So making the early task have a response-selection stage that takes longer would lead to a longer period when the participant has to wait while doing the interrupting task and thus yield longer reaction times.

The results from perfect time-sharing experiments suggests that as long as stimulus modality is not the same for the two tasks, there should be no response-selection bottleneck. However, there could still be a delay in the tasks caused by having two goals in memory during those blocks where there are interrupts compared to those blocks without. Previous researchers have found that until participants are highly learned in a task, they often adopt a conservative strategy that induces a response-selection bottleneck (Schumacher et al., 1999). However, once participants were highly learned in a task, they adopted a more daring strategy that no longer induced a seeming bottleneck. Hence we focused our analyses on late-sessions only (Session 3), because by then participants should be highly learned in the tasks and thus be able to select a wider range of strategies, including daring strategies.

Later theorists (Meyer et al., 1995) showed that there does not need to be a response-

53

selection bottleneck, but that people can set an arbitrary lock-out point before any one of the stages, based on what their strategy dictates. Thus, there could be a response-selection bottleneck, a motor-execution bottleneck, or a stimulus-identification bottleneck, or no bottleneck at all. Meyer et al. argue that people have adaptive executive control (AEC), so that people can adopt different strategies in order to maximize their success on complex tasks. If there need not be any bottleneck when performing multiple tasks, then, hypothetically, people should be able to do two tasks simultaneously at the same speed they could do each task individually.

Researchers have tested for performance bottlenecks with research on time-sharing, i.e., doing two tasks at the same time. If there is a bottleneck in the response-selection stage of a task, it seems likely that there should be a substantial delay in doing two tasks at the same time as compared to doing either task individually (but see Byrne & Anderson, 2001). However, Schumacher et al. (2001) have found that participants can perform two tasks at the same time without a delay in either task. Specifically, one task was a visual-manual task, where participants saw a visual stimulus and had to make a manual response; the other task was an audio-vocal task where the participant heard a tone and then had to make a vocal response. To achieve perfect time-sharing, it is imperative that the two stimuli be in different sensory modalities, because if they are in the same modality, there may be delays due to sensory or motor constraints, such as having to make an eye movement.

Schumacher et al. (2001) results also show that a person can simultaneously hold two goals in working memory and can use either goal without any delay. It is important to realize that the participants suffered no delay in doing both tasks at the same time compared to when they did a single task during mixed blocks where they sometimes had to do both tasks on a trial and sometimes only a single task . Participants did suffer a delay compared to blocks where, on every trial, they only performed one of the tasks, and thus likely stored only one goal in memory and never had to switch between tasks.

It is possible that having to store two goals in memory results in slower task perfor-

54

mance even when using just one of the goals, which supports the Altmann goal-activation theory (Altmann & Trafton, 2002, 2007) discussed below. This is also consistent with some of the findings of switch costs from task-switching experiments (e.g., Rogers & Monsell, 1995). If storing two goals in working memory results in delayed performance, and even if there is no response-selection bottleneck, then doing a task that is often interrupted would be slower than doing it on blocks where you only perform that task by itself.

Further, if there need not be a response-selection bottleneck, then while performing the interrupting task, a person may continue selecting a response for the initial task simultaneously with performing the interrupting task. Thus, changing the difficulty of the initial task may have no effect on how long it takes somebody to do the interrupting task, because the processing for the two tasks could occur in parallel. In fact, changing the difficulty of the initial task may not have any effect on how long it takes to perform the interrupted initial task, because a lot of the processing for this interrupted task could occur while the participant is performing the interrupting task. So if people adopt a strategy with no bottleneck whatsoever, they may be able to do the interrupted and interrupting tasks as quickly as they can perform these tasks individually.

An interruption not only provides information to start a new task, but it also informs you that you must stop doing the task that you were initially performing. Although multiple-task experiments where a participant must stop a task she has begun to perform are still fairly uncommon, there have been studies of stopping a task in single-task settings. These experiments use a countermanding procedure (Osman et al., 1986, 1990; de Jong et al., 1990). In this procedure, participants perform a simple or choice reaction-time task upon receiving some signal, which is usually called the "go" signal. However, on some percentage of trials, they may receive another signal some time after the go signal which tells them to "stop" performing the task. These studies have shown that there comes a "point of no return" beyond which the participant cannot stop performing the task, even if she receives the stop signal. This point is clearly important to consider when discussing interruptions,

because it is also the "point of no interruption." If participants are told to interrupt after this point, it will be too late for them, and they will fail to interrupt. It is thus very important to consider when exactly the interruption will occur, and at what stage of task processing the initial task is then. If the interruption occurs during motor execution, that would be past the point of no return.

So there are many studies of multiple-task performance that examine the basic cognitive processes underlying those skills, and there are some new studies that have begun to examine task interruptions. However, a study of task interruptions that examines the basic cognitive components is still needed. A paradigm that we could use to study these interruptions is an "Anti-PRP" procedure. In this procedure, you have an initial task, but if and when the second task comes along, you have to respond to this second task before responding to the initial task, and only after completing it would you resume the first task. This matches our intuitive idea of what an interruption is, while allowing us to use some of the same manipulations from the PRP procedure to see whether the task-interruption process is different than when the first task has higher priority.

This study is important for many reasons. The Anti-PRP Procedure allows us to test the response-selection bottleneck hypothesis in a different way than has been tested before. Another prospect is that it will allow us to compare task interruptions to task switching. There is no reason to assume that interruptions and switching are equivalent; interruptions give you a motivation to retain some memory of the interrupted task, because you will soon be resuming this task. Storing something in working memory takes time (Meyer & Kieras, 1997a, estimate a 25 msec gating time); since this is not necessary for task switching, it seems reasonable to expect that this will result in different performance than for task switching. Finally, the fact that we are dealing with very basic tasks allows us to computationally model human performance on task interruptions. This will help us gain more insight into different strategies people adopt to deal with interruptions and what stages of processing are affected by having interruptions.
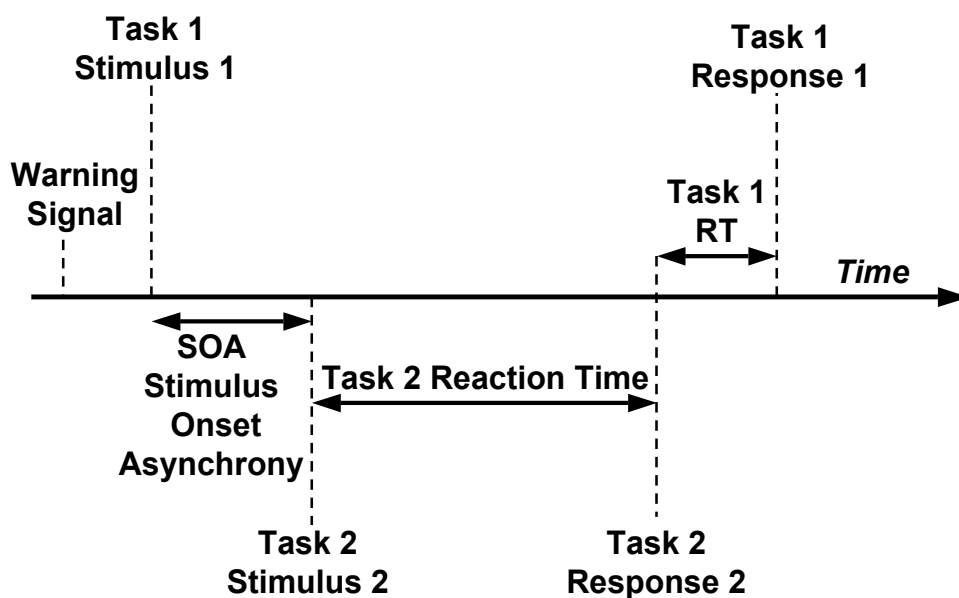
Figure 3.3. A typical trial in the "Anti-PRP" procedure.

We also manipulated the response mode that participants are in when performing a task (Meyer & Kieras, 1997a). In "immediate mode," a motor action is automatically produced as soon as a response is selected. In "deferred mode," the selected response is put into working memory, and a motor action is produced only after the response is retrieved from memory. The response mode you are in when doing a trial could have a large effect on how you interrupt the task. In immediate mode, there is no natural pause between stages. The participant is in a ballistic state where once they begin the task, they proceed to finish. The only way the participant may be able to abort the task is to pull the task goal from memory, so that the corresponding production rules associated with the task do not fire. However, pulling the goal may have hidden costs that have not been studied. For example, when resuming the task, where will you resume? Do you resume at the beginning of the stage in which you aborted? Do you resume where you left off during that stage? How much time

does it take to put the goal back in memory?

In deferred mode, though, there is a juncture in performing the task. When the selected response is put into working memory, a participant may consider the task effectively paused. It will not resume until she retrieves the response from working memory. Hence a person would not have to pull a goal from memory to successfully interrupt a task in deferred mode. Instead, she would just have to change control signals associated with retrieving a response from working memory. Hence if an interruption occurs during response selection while in deferred mode, a person could choose to finish selecting that response, store it in working memory, and then work on the interrupting task. Only after she has finished that task, or at least finished a sufficient amount of the interrupting task, will they retrieve their response from working memory for Task 1 and proceed to make a motor action based on it.

Given these considerations, it is quite possible that an interruption when a person is in immediate mode could be cataclysmic, but not so dire when the person is in deferred mode. We examine this in our Experiments 1 and 2. In Experiment 1, participants were trained to be in immediate mode, and in Experiment 2 participants were trained to be in deferred mode. To encourage immediate mode, in the pure blocks where participants only performed task 1, participants had to be faster than a deadline that encouraged them to respond as quickly as possible on every trial. To encourage deferred mode, on a given trial in the pure block participants were penalized if they responded before given a "go"signal, thus encouraging them to put their selected response in working memory and only proceeding when given the signal. In Experiment 3, we replicated Experiment 2, but with an important difference: there was no motor interference between the responses. Motor interference may contribute to some of the costs we associate with interruptions. Experiment 3 allows us to examine this.

We found in Experiment 1, as predicted, that interrupting a task underway in immediate mode had serious consequences, and seemed to make the participant lose the information

they had ascertained for Task 1. In Experiment 2, we surprisingly got results rather similar to those of Experiment 1, indicating that perhaps immediate versus deferred mode did not affect the strategy participants used to interrupt. However, a closer analysis showed that some of the results may have been caused by manual interference. Experiment 3 had less costs associated with the interrupting task or the uncertainty of whether there will be an interruption or not, and the modeling results provided support for the Strategic Response-Deferment model (Meyer et al., 1995; Meyer & Kieras, 1997a; Schumacher et al., 1999). A model with a single unlocking mechanism fit the data best, and the location of the unlocking seemed to vary across participants, consistent with them adopting different strategies.

We also made computational models for our results in Experiment 3, for each participant in each condition, using a system called CORE (Howes et al., 2007; Vera et al., 2004), which enables cognitively-bounded rational analysis. CORE is a Constraint-based Optimizing Reasoning Engine that allows us to generate models that formally specify architectural constraints and strategy spaces. This is extremely useful in modeling multitasking performance. By using CORE, we can construct models that assume there is a response-selection bottleneck, and other models that assume there is not. CORE also allows us to remain agnostic about the different assumptions made by different cognitive architectures, such as EPIC and ACT-R.

CORE, naturally, is in many ways similar to other cognitive architectures. However, instead of "production rules" of the form IF certain conditions are met THEN some action is performed, CORE uses a Information-Requirements Grammar (IRG; Howes et al., 2005). This approach, like production rules, supports temporal properties and information flows, i.e. the particular process will only occur when certain conditions are met. However, the grammar also requires a specification of which resource is needed to perform the process. For example, a selection process would require the cognition resource, so if that resource was already full via another process or processes, then that selection process could not proceed until cognition was sufficiently available.

That an IRG rule makes use of the resource is one way CORE models can independently vary the architecture. For example, ACT-R assumes that cognition is a unitary function akin to the RSB, in other words when cognition is engaged by one process it cannot be used by another process. EPIC assumes there is no RSB, so when cognition is engaged by one process, it can concurrently be engaged by another process. In CORE, we can change whether to model cognition as a unitary function or not. In this way, we can change the architectural assumptions, and then use the Information-Requirements Grammar to specify specific strategy knowledge. The IRG can also be used to specify a space of strategies, not just a single strategy. For example, in a PRP task the wait-time between the first key-press and the second key-press seems to be strategically selected. We could make an IRG rule for this wait time that has the wait process come from a Normal Distribution with a variable mean. When we then calibrate our model, we could generate a variety of different means for our wait-time, say from 1 msec to 500 msec by 25 msec increments. CORE models will automatically form and test models for all the different means we specified. If we had other processes that also vary strategically, the IRG would form hierarchical models across all the different strategic variables. So if another process varied from 1 to 100 msec by 25 msec increments, CORE would then generate models for the 20 different wait-times * the 4 different mean processes for the other variable, generating 80 different strategies. CORE then automatically compares these strategies and selects whichever strategy (or strategies) earns the most points per trial on average. CORE modeling was a natural fit for our desire to test a wide range of strategies.

3.2   Experiment 1

3.2.1   Method

Participants.   Fourteen undergraduate students participated as paid volunteers. All participants had normal or corrected-to-normal vision and were paid $8.00 per hour plus

bonuses based on the quality of their performance. Because of computer problems, three participants were removed from further analyses. All participants performed for three sessions over three days, with each session occurring no later than two days after the previous session.

Apparatus. Visual stimuli were presented on the display screen of a 17-inch Sony Trinitron monitor connected to a Pentium personal computer. Participants sat about 80 cm from the monitor in a quiet room. Responses were made with a piano-type response keyboard. It had two groups of three finger keys, with one group for each hand. The experiment was controlled by a program written in E-prime.

Experimental Design and Procedure

Tasks. Participants performed three different tasks on three different types of trials during the experiment. The three types of task were an easy-digit task, a hard-digit task, and a tone task. The three different types of trials were homogenous (pure) task, heterogeneous (mixed) single task, and dual task. There were also two types of trial-blocks, pure and mixed. In a pure block, the participant always performed a homogenous task, i.e. the same single task was performed on every trial, and the participant always knew what that task would be. In a mixed block, half of the trials were dual task trials, and half of the trials were heterogeneous single task trials. A heterogeneous single task trial was a trial where the participant performed a single task on that trial, but was not aware before the trial started that it would be a single-task trial. A dual-task trial required the participant to perform two tasks. The first (early) task was always a digit task, but a second (late) task occurred on these trials. The late task was always the tone task.

Pure Single-Task Trial Blocks (Easy-Digit Task). On each trial of the pure single-task blocks with the easy digit task, the participant first saw a 42 by 42 pixel (0.438 x

0.438 inches) square in the center of the display screen. After 500 msec, a digit written in

36-pixel Arial font appeared in the center of the square. The digit was either 1, 2 or 3. If

the digit was 1, the participant made a right-index finger key press. If the digit was 2, the

participant made a right middle-finger key press. If the digit was 3, the participant made a

right ring-finger key press. Figure 3.4 is a graphical representation of one of these trials.

Note that basically the participant is taught to respond as soon as the stimulus is present,

consistent with immediate mode.
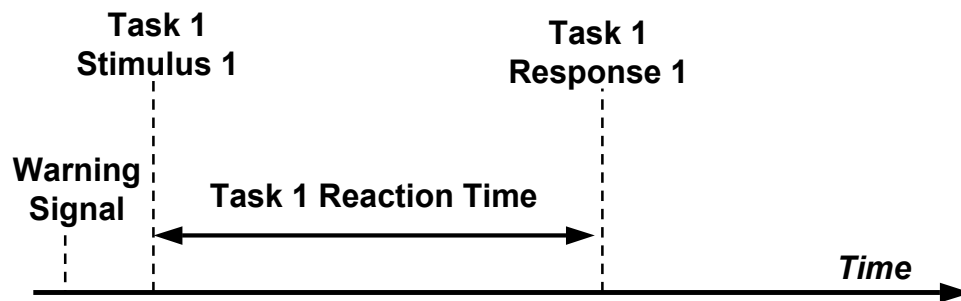
Figure 3.4. A typical trial on a single-task trial. Note it encourages "immediate mode"
because participants are taught to respond as soon as they can.

A deadline was imposed based on the participant's previous reaction times for this trial

type. The deadline was computed for the next block by averaging the 75th percentile of

the RT distribution from the last block of this type with the previous deadline for this

block type. Participants received 200 points minus their reaction time divided by 10 if they

responded correctly and beat this deadline. They lost 200 points for an incorrect answer.

62

They lost 100 points for being slower than the deadline. During Sessions 2 and 3, if a participant responded within 50 msec after the deadline, she received one-fourth of the points that would have been given if she had beaten the deadline. Visual and auditory feedback were provided after each trial, telling the participant if her response was correct, almost fast enough, too slow, or incorrect. She was also told how many points she earned or lost. The inter-trial interval was 500 msec.

Each of these blocks had 36 trials. During Session 1, participants performed this task for 6 blocks; during Sessions 2 and 3, participants performed this task for 5 blocks.

Pure Single-Task Trial Blocks (Hard-Digit Task). The pure single-task blocks with the hard digit task were similar to those with the easy digit task. On each trial, the digit could be 1 - 9. The participant responded with her right index finger if the digit was 1, 4 or 6. If the digit was 2, 7, or 8, the participant responded with her right middle finger. If the digit was 3, 5 or 9, the participant responded with her right ring finger. The deadline was based on reaction times from previous trial blocks of this type and was computed in the same way as was the deadline for the easy-digit single-task blocks.

Pure Single-Task Trial Blocks (Tone Task). On each trial of the single-task blocks with the tone task, participants saw the same fixation as for the digit tasks. After 500 msec, a tone occurred. If the tone was high-pitched (3195 Hz), the participant made a left index-finger key press. If the tone was medium-pitched (880 Hz), the participant made a left middle-finger key press. If the tone was low-pitched (196 Hz), the participant made a left ring-finger key press. The tones were chosen to be easily discriminated. A deadline was imposed on performance of the tone task in the same manner as were the deadlines for the digit tasks. The points and feedback were the same as for the digit task. The inter-trial interval was 1000 msec. Participants performed the tone task for 10 pure single-task blocks of 36 trials during each of the three sessions.

Easy Mixed Dual-Task Trial Blocks. On each trial of the mixed dual-task blocks with the easy digit task, participants saw the same fixation square as on single-task trial blocks. For half of these trials, the events were exactly the same as on easy-digit single-task trials, although the point system differed. Participants still received 200 points - RT/10 for responding correctly before the deadline, and still lost 200 points for incorrect responses. To emphasize responding quickly and prevent strategies that focused exclusively on accurate performance, participants now lost 200 points for responding slower than the deadline. To compensate for this harsher penalty, during Sessions 2 and 3 if these participants responded within 100 msec (not 50 msec as in the pure single-task trial blocks) of the deadline, they were given one-fourth of the points that they would have gotten for responding correctly before the deadline.

For the other half of these trials, participants first saw a digit, but after a certain time (the stimulus onset asynchrony, SOA), they heard one of the three tones of the tone task. Synchronously with the tone, a large (97 by 97 pixels, or 1.01 by 1.01 inches) box surrounded the original square and digit was displayed. Participants had to respond to this tone (in the same way as they responded to the tones on single-task tone blocks) before responding to the digit. After the response to the tone, the large square disappeared from the screen, leaving the original square and digit on the display. Participants then responded to the digit after responding to the tone. See Figure 3.3 for a graphical representation of these interrupt trials.

Following both responses, participants received visual and auditory feedback about how they had performed on the digit and tone tasks. Deadlines were imposed for both tasks; they had the same magnitudes as for the immediately prior corresponding single-task blocks. To discourage participants from initially ignoring the digit task and just waiting for the tone task to occur, and to encourage a quick response to the second interrupting task, we invoked a harsher penalty for responding after the deadline. A response after the deadline resulted in a loss of 200 points on these trials, for both the early digit and the late tone task. However,

because the dual-task blocks were more difficult than the corresponding single-task blocks, there was a 100 msec window after each deadline in which the participant was told "almost" and received one quarter of the points she would have gotten for beating the deadline.

The SOA was set by a staircase-tracking algorithm so that on 70% of the dual-task trials, participants were successfully able to postpone the digit task. On the other 30% of these trials, the tone occurred but participants did not postpone responding to the digit task. For these "failures to postpone," participants lost 100 points, and received no feedback about how they did on the digit task. In all other respects, the feedback was the same as on single-task blocks; performance on each task was awarded points independently of performance on the other task. Treating the two tasks independently allowed participants the greatest range of strategies. So even if a participant responded incorrectly to the tone task, they could still receive points if they successfully responded to the digit task. The inter-trial interval was 1000 ms. Participants had 9 dual-task blocks (each with 36 trials) during Session 1, and 10 blocks during Sessions 2 and 3.

Hard Mixed Dual-Task Trial Blocks. The hard mixed dual-task blocks were like the easy mixed dual-task blocks except for two changes. First, participants performed the pure hard early task throughout these blocks. Second, the SOA was set to have the exact same distribution as for the previous easy dual-task block. This meant that the tone occurred earlier than it would have if it had been set by a staircase-tracking algorithm to successfully postpone 70% of the responses for the hard-digit task.

3.2.2 Results

We are most interested in the participants performance after they learned the task; we are not as interested in the learning process. Therefore we will only analyze data from the final session, after participants had already learned how to perform the task. Two participants never mastered the task, having poor accuracy even on their final session. These

participants were removed from further analyses.

We divided our participants into two subsets. One subset we call "groupers" because they grouped their finger-presses on dual-task trials, responding to the digit task within 100 msec after completing the tone task, consistent with the idea that these participants were "grouping" their two manual responses. The other subset we call "switchers" because participants took a relatively long time to respond for the resumed digit task, consistent with the idea that they were switching from doing the tone task to doing the digit task. Figure 3.5 shows a density distribution of these groups. Note that groupers have very little variance and are less than 100 msec for both the easy and hard-digit case. Switchers have much larger variance, are 72 msec slower on average in the hard-digit case than in the easy-digit case (compared to 21 msec differences for switchers), and are slower than the groupers in both the easy and hard-digit case. We analyzed these groups of participants separately.

We will begin by examining performance on the pure digit blocks, to ensure that our manipulation of response-selection difficulty worked as planned. We will then examine performance on the pure tone blocks, which is useful in understanding the results from the dual-task blocks. Through all of these blocks, we will examine overall statistics across participants and also examine switchers and groupers separately, to see how fundamental the differences between them are. After examining the pure blocks, we will examine performance on the easy-digit mixed-trial block. We will examine the uninterrupted trials, to see what the effect of often being interrupted is. Then we will look at results from interrupted trials, examining groupers and switchers separately, because they perform these tasks differently. Finally, we will perform the same analyses for the hard-digit mixed-trial block.

Pure Task Blocks
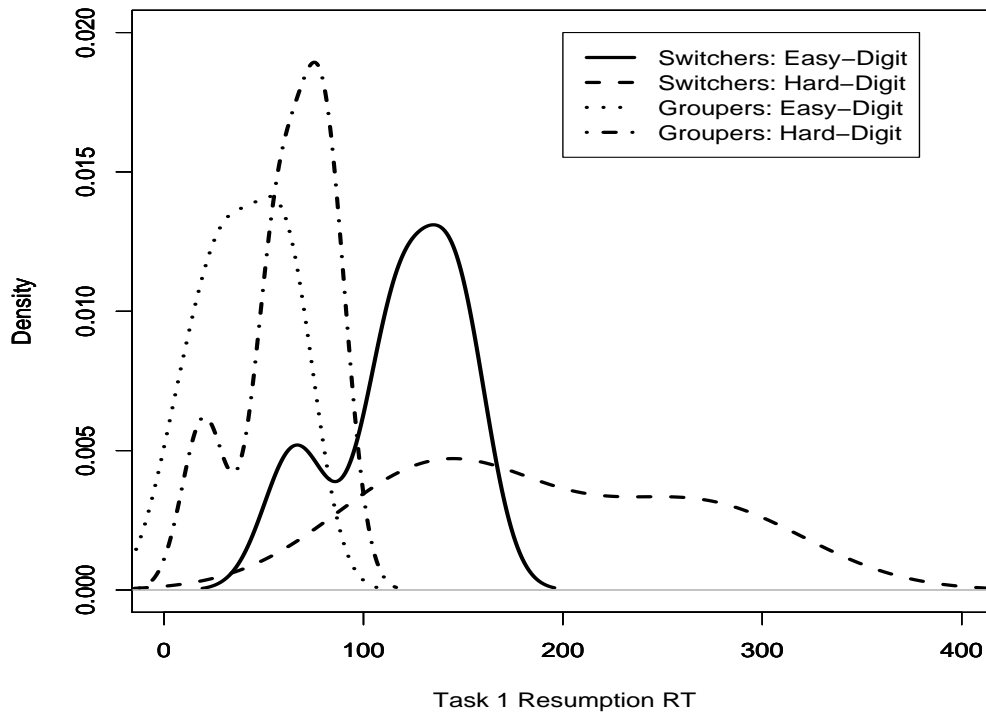
Figure 3.5. Distribution of "groupers" and "switchers" in easy and hard-digit task.

Digit Task. Before interpreting the effects of manipulating the difficulty of the digit task, we first have to ensure that our manipulation actually affected response selection. Analyses across all participants showed that the average median reaction time (the mean of participants median reaction times) for the easy-digit task was 310 msec (standard deviation = 20 msec, inter-quartile range = 26 msec), with an average accuracy of 92.5%; for the hard-digit task, the average median reaction time was 396 msec (sd = 35 msec, IQR = 54 msec), with an accuracy of 86.6%. Because the accuracy was lower for the hard-digit task, we know that this cannot be a case of the speed-accuracy tradeoff. A t-test verified that this was a significant difference (t(8) = 6.37, p < .001). All participants were slower on the hard-digit task than on the easy-digit task. We also examined separately the reaction time of trials during the hard-digit task where the digits 1, 2, 3 were used. These trials were the exact same as the easy-digit task, so differences here can give a good idea of

how much response-selection was affected. On these trials, the average of the median reaction times was 363 msec (sd = 27 msec, IQR = 32 msec), with an accuracy of 90.93%. This was a significant difference (t(8) = 4.73, p < .001), supporting our goal of affecting response-selection (see Table 3.1). For the pure easy-digit task blocks, the average of the median deadlines was 347 msec and participants were slower than the deadline 19.04%, close to the 20% we desired. For the pure hard-digit task blocks, the average of the median deadlines were 478 msec and participants were slower than the deadline 17.65%.

	Pure Easy-Digit Trials	Pure Hard-Digit Trials	Pure Tone Trials
Median RT (msec)	310	396	362
Mean Accuracy (%)	92.5	86.6	88.3
Median RT for 1,2,3 Trials	310	363	—
Groupers - Median RT	307	391	344
Switchers - Median RT	312	412	386
Ultra-Switchers - Median RT	315	433	436

Table 3.1. Results for pure-task blocks.

Groupers. The five participants classified as groupers had an average median RT of 307 msec (sd = 25 msec, IQR = 31 msec), and an average accuracy of 91.89% for the easy-digit task. For the hard-digit task, groupers had an average median RT of 391 msec (sd = 26 msec, IQR = 30 msec) and an accuracy of 87.5%.

Switchers. The four participants classified as switchers had an average median RT of 314 msec (sd = 14 msec, IQR = 24 msec) and an average accuracy of 93.2% for the easy-digit task. For the hard-digit task, switchers had an average median RT of 402 msec (sd = 48 msec, IQR = 76 msec) and an accuracy of 85.8%. There was not a significant difference between groupers and switchers for the easy-digit task (t(7) = .47, p = ns) nor for the hard-digit task (t(7)=.45, p = ns). Note that the ultra-switchers and switchers had similar RTs for the digit task, but that the ultra-switchers were slower in the pure-tone task, although with only two participants in each condition, we cannot test if this difference is significant. Hence for these analyses, we averaged together switchers and ultra-switchers.

Tone Task.   The pure tone-task block showed that participants were able to do the tone task fairly quickly, with an average median reaction time of 362 msec (sd = 53 msec, IQR = 31 msec) and accuracy of 88.3%. However, individual median reaction times varied more for this task than for either of the two digit tasks, with a standard deviation of 53 msec compared to 20 msec for the easy-digit task and 35 msec for the hard-digit task. The mean of each participant's individual deadline was 451 msec and participants were slower than the deadline 18.8%. Although groupers were slightly faster than switchers, 344 msec compared to 384 msec, this difference was not significant (t(7) = 1.02, p = ns).

Easy-Digit Mixed-Trial Blocks

Uninterrupted Trials.   In the mixed-task blocks it seemed likely that often being interrupted should slow performance even on those trials that had no interruption. We will refer to this delay as the "uncertainty effect" because it is caused by the participant being uncertain about whether the next trial would involve an interruption. The average median RT for the uninterrupted easy-digit task, across all participants, was 410 msec. The average uncertainty effect across participants for the easy-digit task was 100 msec. This is a considerable cost, when we consider that the easy-digit task only took 310 msec on the pure blocks where there was never an interruption, a 32% increase. However, individuals varied considerably in how large an uncertainty effect they had for the easy-digit task. The standard deviation was 37.95 msec (IQR = 40 msec), and participants had uncertainty effects that ranged from 145 to 22 msec. Accuracy on these trials was also quite high at 98.7%. Although this suggests a speed-accuracy tradeoff, a look at the percentage of time they were slower than the deadline (for which they lost points) shows that the gain in accuracy was not worth the time they lost. Participants were too slow on 81.6% of the uninterrupted trials. Thus, participants were slower than the deadline 62.5% more in this condition than in the pure easy-digit task.

Groupers: The five groupers had an average median RT of 402 msec, with a standard

	Groupers	Switchers	Ultra-Switchers
Uninterrupted Digit-Task RT (msec)	402	417	419
Uncertainty Effect	95	105	104
Mean SOA	219	241	241
Interrupted Digit-Task RT	261	368	364
Tone-Task RT	489	498	624
Interrupting Lag	146	112	188

Table 3.2. Results for easy-digit mixed blocks.

deviation of 54 msec, an inter-quartile range of 78 msec, an average accuracy of 98.4%, and were too slow on 78.3% of these trials. The average uncertainty effect was 95 msec.

Switchers: The four switchers had an average median RT of 420 msec, with a standard deviation of 46 msec, an IQR of 70.5 msec, an average accuracy of 98.7%, and were too slow on 85.7% of these trials. No difference was significant between groupers and switchers. The average uncertainty effect was 106 msec. Note that switchers and ultra-switchers had very similar patterns of data on non-interrupted trials, so these two subsets were collapsed together here.

Interrupted Trials. Groupers: As Table 2 shows, the average SOA for the groupers was 219 msec. At this point, they heard the tone, and then had to respond to it. They were unable to postpone doing the easy-digit task on 30.3% of the trials, as determined by our staircase tracking algorithm. They then started the tone task. The average of the median RTs for the tone task was 489 msec (sd = 53 msec, IQR = 62 msec), with an average accuracy of 90.7%. This is considerably slower than on the pure tone-trial blocks. We will refer to this difference between the pure tone-trial reaction time and the interrupting tone reaction time as the "interrupting lag." The average interrupting lag for groupers was 146 msec.

The reaction time for an interrupted digit task did not include the time during which the participant was performing the tone task. Hence the total reaction time for an interrupted digit trial included the SOA and the time from when the participant responded to the tone

Easy–Digit Interrupted Trials Results
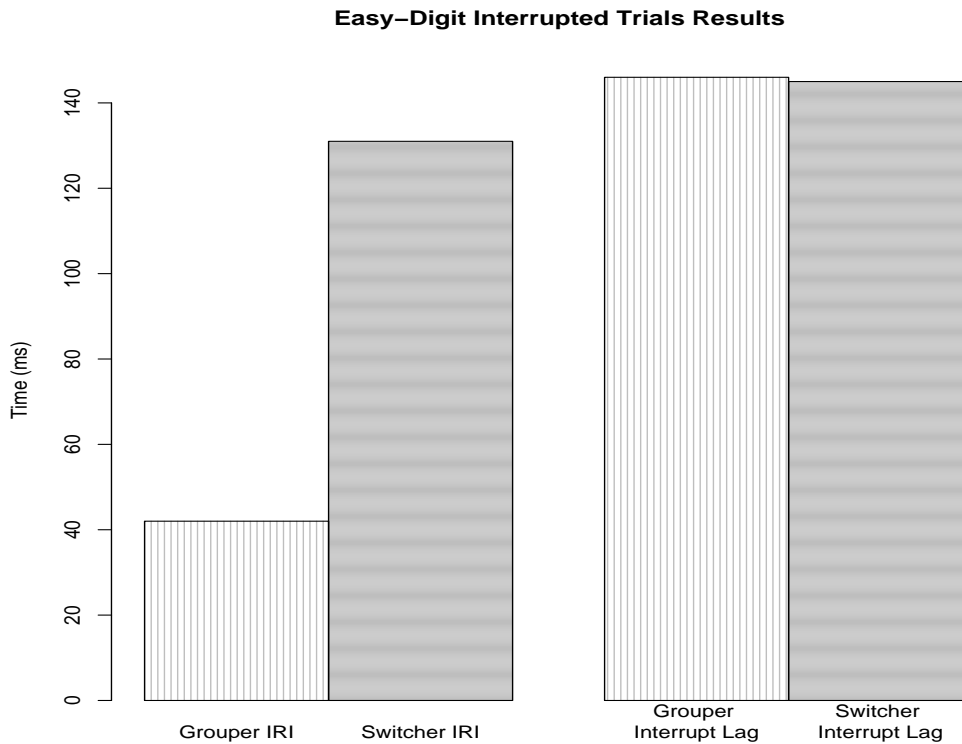
Figure 3.6. Inter-response interval differences between "groupers" and "switchers", both groups have similar lags in doing interrupted tone task compared to pure tone task.

task until they responded to the digit task. For the interrupted easy-digit task, the average of their median RTs was 261 msec (sd = 69 msec, IQR = 36 msec), but subtracting the average SOA from this time, this corresponds to taking 42 msec after responding to the tone to respond to the digit. The average accuracy on the interrupted easy-digit task was 96.4%.

We also examined the correlation between interrupting tone RT and the time from tone onset until the response to the interrupted digit RT. Because these two variables have the same onset, they are not independent and thus we expected a strong correlation, but nonetheless we expected that groupers should have a particularly strong correlation. We found that the average correlation for the groupers was .93.

Switchers: The average SOA for the four people who were classified as switchers was 241 msec. These participants failed to postpone 28.9% of the time, around the 30% we

desired. For the tone task, the average of their median RTs was 529 msec (sd = 97 msec, IQR= 63 msec) with an accuracy of 91.6%. The average of their interrupting lags was 145 msec. This is equivalent to that of the groupers. However, the average RT for the interrupted digit task was 373 msec (sd = 41 msec, IQR = 48 msec) with an accuracy of 97.4%. This corresponds to taking 132 msec to complete the digit task after it resumes, considerably longer than for the groupers.

The average correlation for the switchers between tone RT and the time from tone onset until the response to interrupted digit RT was .81.

Although we collapsed switchers and ultra-switchers here, an argument could be made that they seem to differ on interrupted tone-task RT. However, this difference was largely caused by one participant who had a median RT for the interrupting tone task of 756 msec, whereas the ultra-switcher had a median interrupting tone-task RT of 492 msec. This participant's tone-task RT is very similar to that of the two participants classified as switchers, so we decided just to collapse these four participants together into one subset.

Hard-Digit Mixed-Trial Blocks

Uninterrupted Trials. The average RT for the uninterrupted hard-digit task was 450 msec, with an average uncertainty effect of 54 msec across subjects and a standard deviation of 28 msec. Individual participants ranged from having a -5 msec (being faster on the uninterrupted trials of the mixed blocks than on the pure blocks) to having a 93 msec uncertainty effect. Accuracy here was also higher (90.9%) than on the corresponding pure blocks (85.7%), t(8) = 2.61, p<.05. Participants in this condition were too slow on 41.6% of the trials, not much more often than for the pure hard-digit task, as only 23.9% of the time more often are they too slow.

Groupers: The five participants had an average median RT of 442 msec (sd = 35 msec, IQR = 60 msec) on the uninterrupted hard-digit task, with an average accuracy of 89.7%. They were slower than the deadline 42.1% of the time. Their average uncertainty effect

72

was 52 msec.

Switchers: The four participants had an average median RT of 460 msec (sd = 36 msec, IQR = 48 msec), with an average accuracy of 91.9%. They were slower than the deadline 40.9% of the time, and had an average uncertainty effect of 59 msec.

	Groupers	Switchers	Ultra-Switchers
Uninterrupted Digit-Task RT (msec)	442	449	471
Uncertainty Effect	52	47	69
Mean SOA	222	243	228
Interrupted Digit-Task RT	284	381	497
Tone-Task RT	552	530	636
Interrupting Lag	208	193	235

Table 3.3. Results for hard-digit mixed blocks.

Interrupted Trials. Groupers: The average SOA for these participants was 222 msec, similar to the SOA for the easy-digit task. These participants failed to postpone doing the hard-digit task 23.4% of the time. This is what we expected, because an interruption in an earlier stage of task processing makes one more likely to be able to postpone that task. These participants had an average RT on the tone task of 552 msec (sd = 61 msec, IQR = 70 msec), with an accuracy of 86.4%. The average interrupting lag was 208 msec, and overall the interrupting lag was larger for the hard digit task than for the easy dual-task block (t(16) = 2.96, p < .01). Note, however, this difference was less than the difference between the pure easy-digit task and pure hard-digit task RT. The average median RT for the interrupted hard-digit task was 284 msec (sd = 69 msec, IQR = 7 msec) with an accuracy of 92.1%. This indicates that these participants took 62 msec to respond to the digit after responding to the tone.

Switchers: Switchers in this condition were further sub-divided into two subsets: one called "switchers" and one called "ultra-switchers." The switchers are ones who have the same pattern of results for the hard-dual trials as the easy-dual trials. Ultra-switchers had a different pattern for the hard-dual trials, being considerably slower on both the interrupting
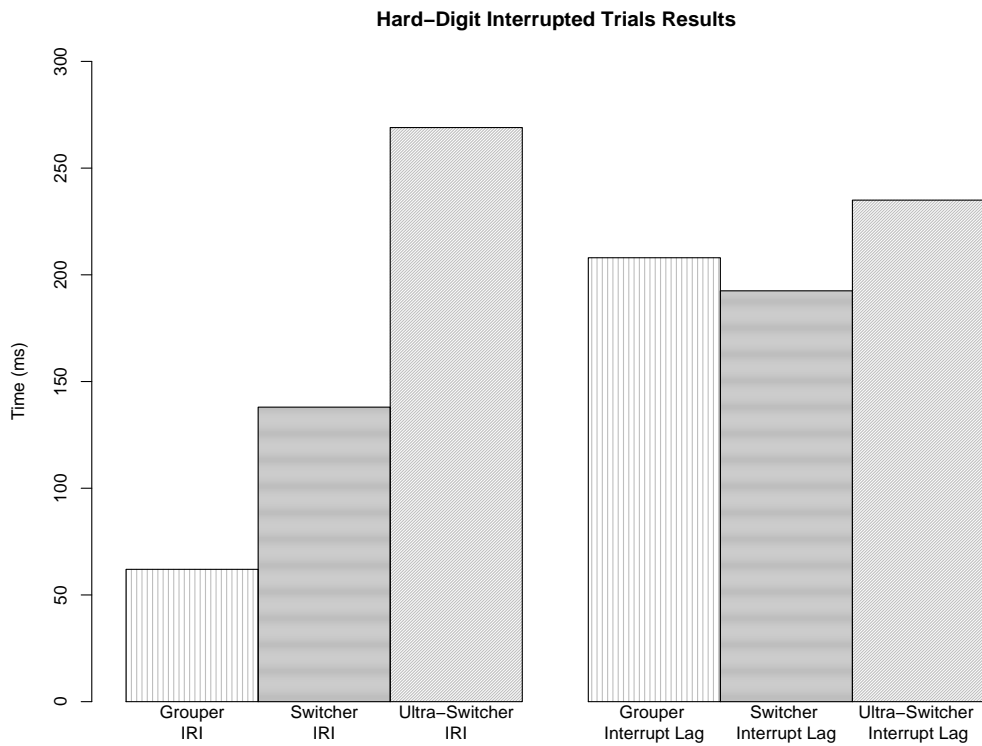
Figure 3.7. Inter-response interval differences between "groupers," "switchers" and "ultra-switchers;" all groups have the same slowdown in doing interrupted tone task compared to pure tone task, which we refer to as the "interrupting lag."

tone task and the interrupted hard-digit task than switchers or groupers. Two participants were classified as switchers, and two as ultra-switchers. Switchers had an average SOA of 243 msec and failed to postpone 24.3% of the time. For the tone task, the average of their median RTs was 530 msec with an accuracy of 82.1% and an average interrupting lag of 193 msec. The difference in interrupting lag between switchers and groupers was not significant here (t(7) = .16, p=ns). For the interrupted digit task, these participants had an average median RT of 381 msec with an accuracy of 92.2%. These participants spent, on average, 138 msec on the resumed hard-digit task.

Ultra-Switchers: Ultra-switchers had an average SOA of 227.5 msec. They failed to postpone 18.9% of the time. For the tone task, the average of the median RTs was 636 msec, with an accuracy of 93.1%. The average interrupting lag was 235 msec. For the interrupted hard-digit task, the average median RT was 497 msec with an accuracy of 93.9%. So these participants spent 269 on this task.

3.2.3   Discussion

In Experiment 1 we discovered how difficult it is to interrupt when the interrupted task, Task 1, is performed in immediate mode (Meyer & Kieras, 1997a). This can be deduced by looking at the "interrupting lag" for Task 2, the slowdown in responding to the tone task on dual-task trials compared to performing the tone task on pure-task trials. For the easy-digit mixed-trial blocks, both sets of participants (i.e. "groupers" and "switchers") had about a 145 msec slowdown, an approximately 40% slowdown compared to how quickly participants performed when just doing the tone-task. When doing the hard-digit mixed-trial blocks, the interrupting lag was even greater, being about 200 msec, about a 50% slowdown. This was done in spite of the payoff being structured to emphasize performance on Task 2. Why was there such a large lag?

One clue is the fact the interrupting lag differed in the easy and hard-digit task. Note that the tone task is the exact same in the conditions, so this difference indicates that the

response-selection difficulty of the digit task affected performance on the tone task. It appears that some of the processing of the digit-task response selection occurred while the participants should have processed only the tone task, consistent with the response-selection bottleneck hypothesis. However, note that the difference in interrupting lag was only about 50 msec, only about half as much as the 90 msec response-difficulty effect. However, without careful modeling of this task, we cannot be sure if this fact is consistent with the RSB or not.

That we found evidence of processing of Task 1 when people should be processing Task 2 is not surprising for certain sub-sets of our participants. We found that some of our participants seemed to group their responses to Task 2 and Task 1, such that from the time they make their first key press, they took less than 50 msec to make their second key press. This speed is not enough for them to engage in response-selection for the interrupting task after making their key press for Task 2, hence response selection for Task 2 must be completed before these participants began their key press for Task 1. Yet the participants who grouped their responses did not have larger interrupting lags than did those participants who seemed to switch between doing the two tasks.

Switchers was the label we applied to those participants who took longer to respond to the interrupted digit-task. These participants took about 140 msec to resume their response to the digit-task, or as much as 250 msec in the case of those participants we labeled "ultra-switchers." These participants could have engaged in response-selection for Task 1 simultaneously with their motor-production for Task 2, in which case they should not have any lag in doing Task 2, because motor-production rules for one task should not interfere with cognitive rules for the other task. And yet, they still exhibit just as much lag as groupers, indicating that they had still engaged in response selection for Task 1, before selecting a response for Task 2, and then they chose a response again for Task 1.

We have some clues about individual differences that contribute to making a person a "grouper," "switcher," or "ultra-switcher." The biggest factor seems to be how quickly

76

participants performed the pure-tone task. Participants who performed the pure-tone task quickly tended to become groupers. Those participants who were slowest on the pure-tone task tended to become ultra-switchers. It is possible that participants who took the longest to do the tone task realized that they had to prioritize finishing the tone-task response selection as soon as possible, so were more willing to abort task 1 response selection before gating their response into working memory, resulting in the need to do response-selection for Task 1 again, which they did either in parallel with motor preparation for Task 2 (switchers) or after motor preparation was finished for Task 2 (ultra-switchers).

Evidence for the differences between switchers and groupers comes in part from the slowdown in uninterrupted task 1 trials on mixed-blocks compared to on pure-blocks, the so-called "uncertainty effect." From Chapter 2, we know that participants tend to be rational and try to find the optimal solution to maximize their points. And yet we found that our participants slowed down on trials during the mixed-blocks that were the exact same as on the pure-blocks. Why? Because participants knew that there was some chance that the trial would be interrupted during a mixed-block, so they were trying to optimize the points they earned on average across both single-task trials and dual-task trials. To do this, they probably realized it was much easier if they interrupted while they were still perceptually processing Task 1. Once they had begun selecting a response for Task 1, participants found that they could not really interrupt without severe consequences, which will be discussed below. So participants would wait longer on a given trial to see if it was an interruption, and only when they were confident it was a single-task trial would they begin selecting a response for Task 1. Of course, we used a variable SOA, so even if participants waited to begin selecting a response, the SOA would change such that they were still interrupted after they had begun selecting a response for Task 1.

Participants were thus stuck between a rock and a hard place: if participants did Task 1 as fast as they could, then interruptions would cause severe problems and cost them points on dual-task trials. If instead participants waited to see if the trial was a dual-task trial, then

the SOA would change such that as soon as the participant began to process Task 1, only then would the interruption occur.

The interruption caused such severe interference perhaps because participants had learned to do Task 1 in immediate mode. This meant that as soon as participants had finished selecting a response, they immediately proceeded to make a response. This clearly was a problem if they were interrupted at some point by Task 2, before the point-of-no-return where they should still have been able to inhibit their response to Task 1 (Logan & Cowan, 1984; de Jong et al., 1990). In the case of this interruption, then, participants had to pull their goal associated with Task 1 in order to stop responding to Task 1 before their responded to Task 2. Pulling the goal meant participants still had to process Task 1, as the work they did in selecting the response was not stored in working memory. Hence participants, especially switchers, indeed seemed to select a response for Task 1 twice, as they did not remember what response they stored the first time. Groupers, in contrast, may have somehow remembered their response to Task 1, perhaps by pulling their goal later in processing. Their interrupting lag, then, is caused only partially by a delay before they began selecting a response for Task 2. The primary cause of their lag is electing to do a "patter response" where they pressed one finger on one hand and then quickly pressed a finger on the other hand. Patter responses are complicated responses, so take longer to prepare, estimated in EPIC to take 200 msec, since four features need to be specified (hand/finger for each response) compared to 100 msec for a single finger response when hand must also be specified.

Thus, participants in immediate mode do not seem capable of successfully interrupting without a large cost. We also made it too hard for participants to come up with an optimal strategy of dealing with the variable SOA and at the same time processing single-task trials.

3.3 Experiment 2

Experiment 1 failed to achieve some of our original objectives; we did not get any evidence of parallel response-selection in our participants, but a possible reason was problems in the design of Experiment 1. The variable SOA made it so that participants did not even have a chance to let response-selection overlap for the two tasks. We also encouraged participants to be in immediate mode, so that as soon as they chose a response for Task 1 they proceeded to produce it. We needed to encourage them more fully to be in deferred mode to see if they could interrupt more easily when responses initially went to working memory instead of directly to motor processors. Hence Experiment 2 was designed to correct these faults, while at the same time making the procedure a bit easier for participants.

3.3.1 Method

Participants.   Eight undergraduate students participated as paid volunteers. No participant was in Experiment 1.

Apparatus.   The apparatus was the same as in Experiment 1.

Experimental Design and Procedure.   The tasks were the same as in Experiment 1. However, we made numerous changes in the design to further emphasize the interrupting task on dual-task trials, and to allow for the possibility the response-selection overlap.

   Pure Digit Tasks.   We made two significant changes to the pure-digit tasks from Experiment 1. First, we manipulated encoding difficulty by having intact legible and degraded digits (digits that had pixels removed, making it harder to identify them). This has been shown to affect the stimulus-encoding stage of a task, but not response-selection (Sternberg, 1969a). This manipulation made it more likely that the response-selection stage of the two tasks would overlap and thus could reveal how people strategize needing to choose

two responses simultaneously.

Another change concerned the fact that on single-task trials during mixed-blocks, participants had to worry about there being an interruption, so this affected their strategy and made it hard to compare their results to the pure-task trials. For the mixed-blocks, based on concepts from EPIC, we believed participants were in a "deferred" mode whereby the chosen response went directly to working memory. For pure-blocks of Experiment 1, however, we believed that participants were in "immediate" mode whereby the chosen response went directly to motor processors to produce movements. We thus felt participants had to be in deferred mode on pure blocks, to allow for direct comparisons between pure and mixed blocks. To do this, participants saw a large red box around the digit on pure blocks, and they were not allowed to respond until the box turned green. The box had three different SOAs for the easy digit task: 0, 200, or 400 msec. For the hard digit-task, the SOAs were 0, 250, or 500 msec.

To aid the participants in deciding how fast they should respond, we made the deadline visible to them via a large animated box that collapsed toward the smaller box that contained the digit. When the large box reached the digit-box, the participant reached her deadline and would be "Too Slow." Participants thus had a visual guide to know how quickly they should respond.

The tone task was the same as before, except that the animation was now present so that participants could judge their deadline for the tone task, too.

Mixed Dual-Task Trial Blocks. We fixed the SOA on mixed blocks to be 100 msec, so the interrupting tone always occurred then. On non-interruption trials, the digit task was as above, with a SOA of 0, 200, or 400 msec for the easy-digit task (0, 250 or 500 for the hard-digit task). The payoff was also changed to further emphasize the interrupting tone task, by making it so that if participants were Too Slow or erroneous on the tone task, they did not receive any points for how they did on the digit task (although they could still lose
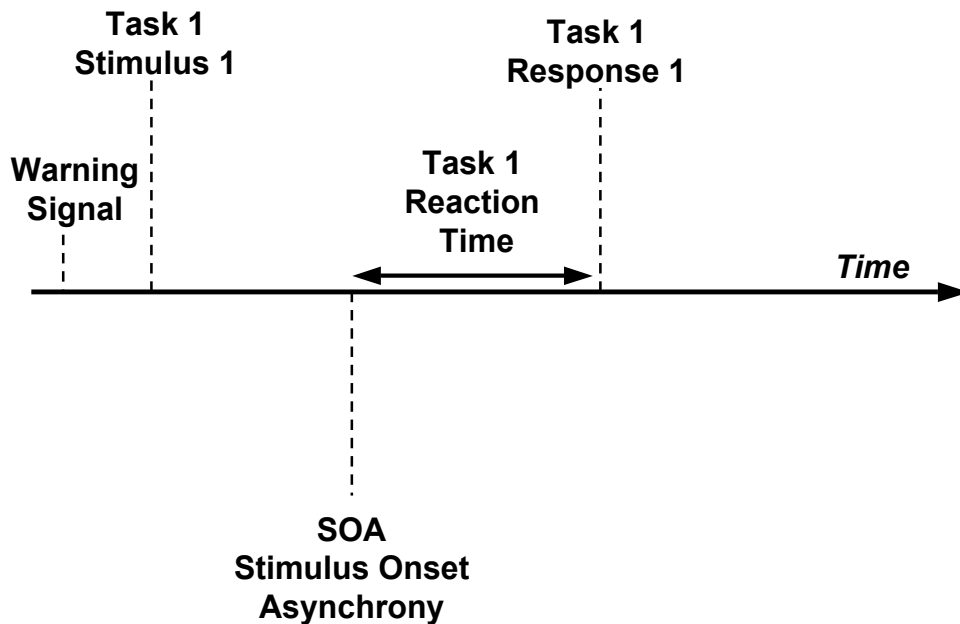
Figure 3.8. A typical trial on a single-task trial. Note it encourages "deferred mode" because participants are taught to respond only after the SOA.

points if they responded incorrectly).

3.3.2 Results

Because we are generating so many statistics and comparisons, due to the added manipulation of having three different SOA's, all analyses were done on the mean of the median reaction times for each block. One participant was removed from further analyses due to poor performance on single-task trials of mixed blocks.

Pure-Task Blocks

Digit and Tone Task. Because each digit task had three different SOAs, we will refer to these as low (for SOA 0), medium (for SOA 200 in the easy-task, SOA 250 for the
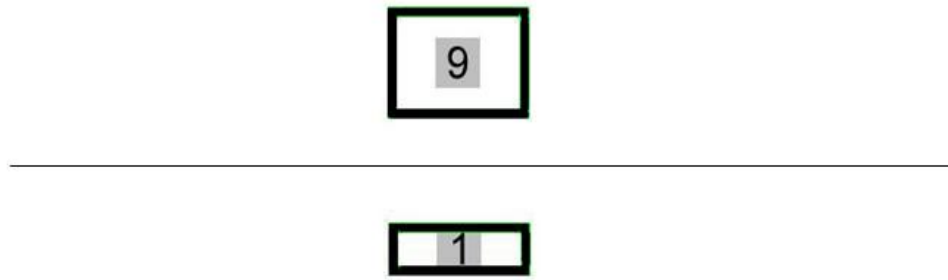
Figure 3.9. Screenshot of typical trial from digit task. Top view shows the start of trial. Bottom shows what participant sees when deadline is reached. Note that the original color of surrounding square is green.

hard-task), and high (SOA 400 easy, SOA 500 hard). Unlike in Experiment 1, we found no evidence that participants were grouping responses, so we did not separate the participants into different groups. All t-tests were done within subject, so paired t-tests were used.

We first examined our response-selection effect. Table 3.4 summarizes the mean and standard deviation for each condition. Responses to whole stimuli (stimuli that were not degraded), at Low-SOA were significantly slower in the hard-digit task (M = 423 msec, sd = 25.9 msec) than in the easy-digit task (M = 330 msec, sd = 19.2, paired t-test t(6) = 9.18, p < .001). Note that the Low-SOA RT is about 20 msec longer for both the easy and hard-digit task than in Experiment 1, which may be caused by the participants having to perceive the go signal. At Mid-SOA, we again found that the RT for the hard-digit task with a mean of 241 msec was larger than the RT for the easy-digit task with a mean of

215 msec (t(6) = 5.73, p < .005). Unexpectedly, at High-SOA, there was a difference in the hard (M = 208) and easy (M = 196) case, t(6) = 3.82, p < .01. We found the same pattern for degraded stimuli, such that at all three SOAs, response-selection difficulty had a significant effect (Low-SOA: hard-digit RT of 471 > easy-digit RT of 368, t(6) = 8.48, p < .001; mid-SOA: hard RT of 266 > easy RT of 235, t(6) = 7.45 p< .001; high-SOA: hard RT of 209 > easy RT of 194, t(6) = 3.31, p < .02). However, the difficulty effects were very small in the High-SOA condition.

We next examined whether our degradation had an effect. At Low-SOA, we found a significant effect for both the easy and hard task. In the easy task, we found that degraded RT of 368 was significantly greater than the whole RT of 330 msec, t(6) = 5.29, p<.005, and that in the hard-digit task, the degraded RT of 471 was significantly greater than the whole RT of 423, t(6) = 5.75, p< .005. At Mid-SOA, we found for the easy-digit task that degradation still had an effect, degraded RT of 235 was significantly greater than whole RT of 215, t(6) = 5.93, p< .005, as well as for the hard-digit task, where degraded RT of 266 > whole RT of 241, t(6) = 4.90, p < .005. At High-SOA, we found no degradation effect for either difficulty level. For the easy-digit task, whole RT of 196 = degraded RT of 194, t(6) = 0.70, p=ns. For the hard-digit task, whole RT of 208 = degraded RT of 209, t(6) = 0.15, p = ns.

The tone task yielded a mean time of 384 msec, indicating similar performance to that in Experiment 1.

We also examined "catch" trials, where the participant was never given a "go" signal and so was not supposed to respond at all. We found that amongst these trials, participants incorrectly responded on 11.6% of easy, whole trials, and 5.3% of easy, degraded trials. However, these numbers belie quite a range, from 0.0% of these errors to 29.6% of these errors in the easy, whole condition for one participant. This seems to indicate a clear strategic difference, as one participant anticipated more than another. In the hard-digit task with whole stimuli, participants made 7.9% such errors, and 4.8% in the hard, degraded

stimuli condition.

Condition (msec)	Low-SOA	Mid-SOA	High-SOA
Whole Easy	330 (19.21)	215 (12.79)	196 (14.51)
Degraded Easy	368 (27.26)	235 (14.44)	194 (12.03)
Whole Hard	423 (25.92)	241 (10.75)	208 (12.06)
Degraded Hard	471 (36.20)	266 (21.61)	209 (20.78)
Tone	384 (30.76)	—	—

Table 3.4. Results for pure-task blocks.

Mixed-Trial Blocks

Uninterrupted Trials. Table 3.5 shows results from the single-task trials of the Easy-Digit Mixed-Trial Blocks. There was still a significant response-selection difficulty effect at SOA 0, such that the hard task took longer for both whole and degraded stimuli (whole hard RT of 464 msec > whole easy RT of 420 msec, t(6) = 7.07, p<.001; degraded hard RT of 511 > degraded easy RT of 433, t(6) = 7.00, p< .001). At Mid-SOA, the RT difference for hard degraded stimuli compared to easy degraded stimuli was marginally significant (t(6) = 2.00, p=.09), and it was significant at High-SOA (t(6) = 2.45, p< .05). For the whole stimuli, at Mid-SOA the hard task took significantly longer than the easy task (t(6) = 2.89, p< .05), while there was no significant difference at High-SOA (t(6) = 1.87, ns).

The degradation effect on RT was no longer significant except in the hard-digit task at SOA 0 (t(6) = 4.12, p<.001). Even at the Low-SOA for the easy task, the degradation effect was not significant, t(6)=1.29, p=ns.

Although the degradation effect was greatly reduced, reaction times in these single-task trials were significantly greater than in the pure blocks. This phenomenon we referred to as the "uncertainty effect" since it is caused by a participant not being certain a priori whether the trial will be a single or dual-task trial. This uncertainty effect was significant in all cases except the High-SOA. For the mixed-blocks uninterrupted whole easy-digit Low-SOA task, RT of 420 > pure whole Low-SOA RT of 330, t(6) = 8.88, p<.001, mixed-blocks whole

Figure 3.10. Comparison of pure and uninterrupted RT for easy-digit task. Top-graph is at Low-SOA, middle-graph is at Mid-SOA, and bottom-graph is at High-SOA. Note the different scales.

Mid-SOA uninterrupted RT of 275 > pure whole Med-SOA RT of 215, t(6) = 5.89, p< .001, although High-SOA mixed-blocks uninterrupted RT of 196 = pure whole High-SOA RT of 196 t(6) = .01, ns. For the degraded easy-digit task, mixed-blocks degraded Low-SOA uninterrupted RT of 433 > pure degraded Low-SOA RT of 368, t(6) = 6.44, p<.001, mixed-blocks degraded Mid-SOA uninterrupted RT of 282 > pure degraded Mid-SOA RT of 235, t(6) = 6.65, p<.001, and mixed-blocks degraded High-SOA uninterrupted RT of 197 = pure degraded High-SOA RT of 194, t(6) = .84, ns.

85

Condition (msec)	Low-SOA	Med-SOA	High-SOA
Whole Easy	420 (40.42)	275 (31.97)	196 (20.18)
Degraded Easy	433 (41.58)	282 (33.09)	197 (15.51)
Whole Easy Uncertainty Effect	90	61	0
Degraded Easy Uncertainty Effect	65	47	3
Whole Hard	464 (42.40)	295 (23.31)	219 (34.20)
Degraded Hard	511 (42.17)	298 (46.19)	213 (22.35)
Whole Hard Uncertainty Effect	40	54	11
Degraded Hard Uncertainty Effect	41	33	4

Table 3.5. Results for uninterrupted trials on mixed-trial blocks.

In the whole hard-digit task case, we get a similar pattern. Mixed-blocks whole Low-SOA uninterrupted RT of 464 > pure whole Low-SOA RT of 423, $t(6) = 3.19$, $p< .02$; mixed-blocks whole Mid-SOA uninterrupted RT of 295 > pure whole Mid-SOA RT of 241, $t(6) = 5.89$, $p<.001$, but not at High-SOA, mixed-blocks whole uninterrupted RT of 219 = pure whole High-SOA RT of 208, $t(6) = .87$, ns. The same pattern holds for the degraded stimuli: mixed-blocks degraded Low-SOA uninterrupted RT of 511 > pure degraded Low-SOA RT of 471, $t(6) = 3.74$, $p< .01$, degraded Mid-SOA mixed-blocks uninterrupted RT of 298 > pure degraded Mid-SOA RT of 266, $t(6) = 2.70$, $p<.05$, but again not at High-SOA, where mixed-blocks degraded High-SOA uninterrupted RT of 213 = pure degraded High-SOA RT of 209, $t(6) = 1.37$, ns.

Interrupt Trials. Table 3.6 shows the results from interrupt trials. As is obvious, the median RT for the tone task was longer on dual-task trials than it was on pure-task trials. This cost we call the "interrupting lag." It is clear that degradation had no effect on the tone-task RT, but it turns out response-selection difficulty did have an effect, i.e. if the initial digit task was easy, participants responded faster to the tone than if the initial digit was hard. If we combine easy-whole and easy-degraded tone RTs into one measure (since they did not differ), and do the same for hard, we find that for the easy-digit task, tone
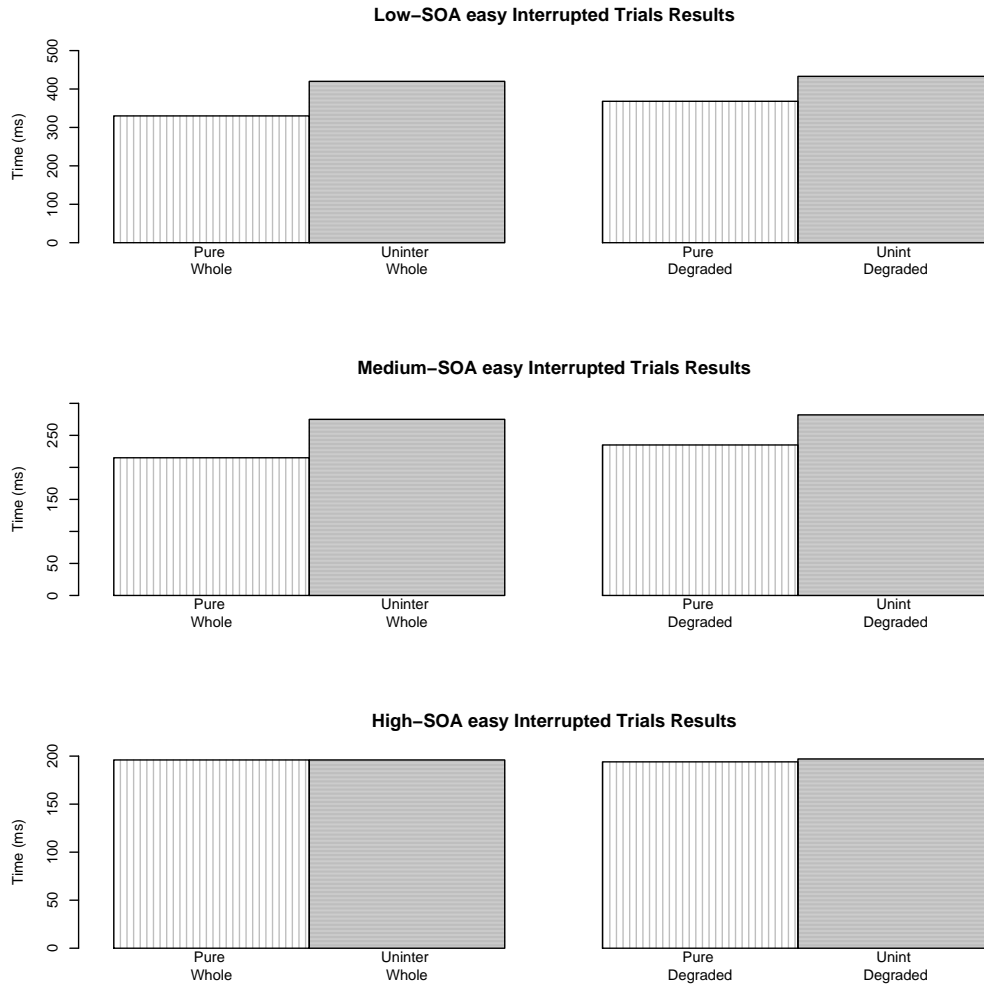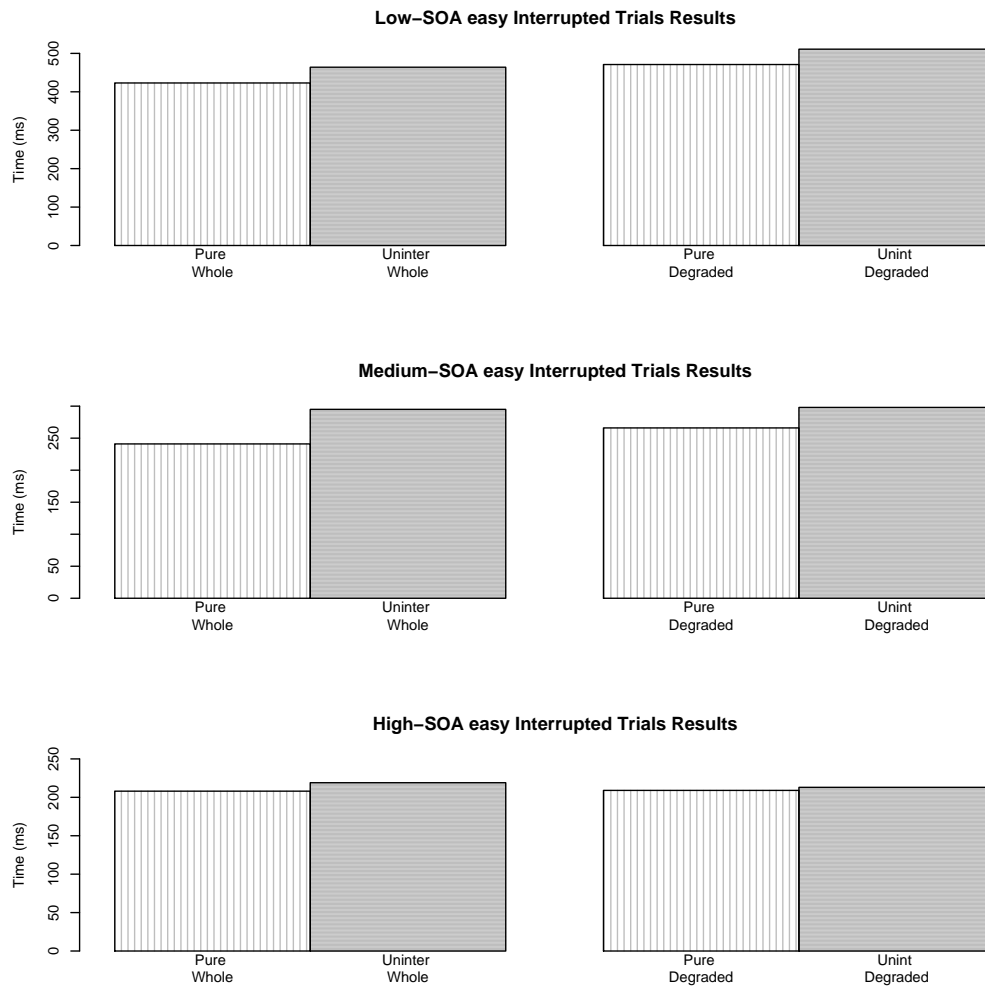
Figure 3.11. Comparison of pure and uninterrupted RT for easy-digit task. Top-graph is at Low-SOA, middle-graph is at Mid-SOA, and bottom-graph is at High-SOA. Note the different scales.
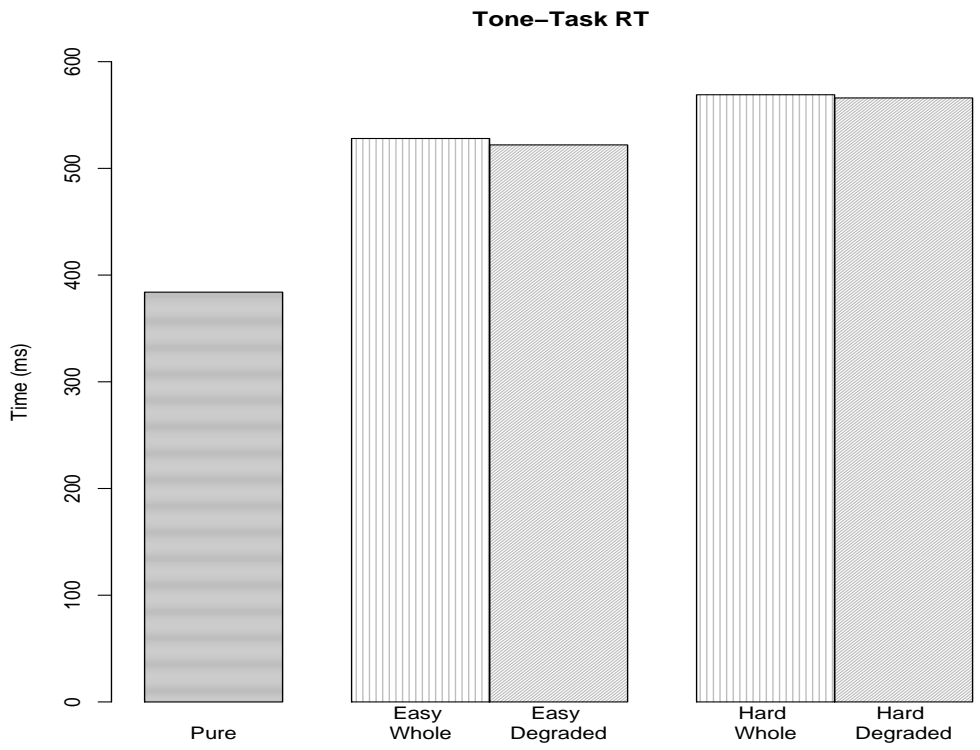
Figure 3.12. Comparison of tone task from pure blocks and from interrupt trials, depending on the type of digit task.

RT (m = 525) is less than for the hard-digit task (M = 568 msec), t(13) = 3.89, p< .005. Note, though, that this 43 msec difference is less than the 100 msec difference between pure easy-digit RT and pure hard-digit RT.

	Easy-Whole	Easy-Degraded	Hard-Whole	Hard-Degraded
Interrupted Digit-Task RT	194 (30.77)	195 (31.68)	261 (36.63)	258 (29.68)
Tone-Task RT	528 (91.16)	522 (82.55)	569 (126)	566 (108)
Interrupting Lag	144	139	186	183
Fail to Interrupt (%)	6.82	5.66	1.73	1.73

Table 3.6. Results for interrupted trials on mixed-trial blocks.

We also examined how often participants failed to interrupt by responding to Task 1 before they responded to Task 2. There were relatively few of these errors. Amongst all dual-task trials, participants failed to interrupt on 6.8% of the trials if Task 1 was the easy-digit whole-stimuli condition, 5.7% in the easy-digit, degraded-stimuli condition. However, there was interesting variation between individuals, with a range of 0.00% such errors to 11.96% such errors.

3.3.3 Discussion

Interestingly, the results of Experiment 2 were similar to those of Experiment 1. Partic- ipants still had significant "uncertainty effects" of around 90 msec for the easy-digit task and around 40 msec for the hard-digit task. In this case, there was a fixed SOA at 100 msec, early enough that participants should not have had to wait so long to determine whether it was a dual-task trial. We also found a similar interrupting lag as in Experiment 1, around 140 msec for the easy-digit task and 180 msec for the hard-digit task.

These results rather surprised us, as we expected that being in deferred mode should make interruptions much easier, and yet this did not seem to be the case. However, Experi- ment 2 yielded little evidence of ""groupers" and ""switchers" as in Experiment 1. Almost all participants took around 100 msec to resume the easy-digit task. One participant seemed

like a clear grouper, taking 30 msec to resume the easy-digit task (but 110 msec in the hard-digit task). However, the next fastest participant took 84 msec in the easy-digit task, and all the other participants had around 100 msec. There were also no clear ""groupers" in the hard-digit task.

A question that we had from Experiment 1 remains: what is causing the interrupting lag? Perhaps Experiment 2 still involved a confound. Participants on dual-task trials had to make two manual responses, so it was possible that there was interference between them (Peters, 1977; Caroselli et al., 1997). If so, that would create a delay in responding to Task 2, and possibly explain our interruption lag.

The uncertainty effect may also just be caused by manual interference. Participants in pure blocks can specify their hand before a trial begins, so only one feature needed to be specified. However, in dual-blocks, participants are not sure with which hand they have to initially respond, so they now have two features they need to specify. This would take 100 msec to specify, resulting in a 50 msec "uncertainty effect." This is similar to the uncertainty effect in the degraded-easy, whole-hard, and degraded-hard conditions (and in fact none of these conditions significantly differed from 50, although the whole-easy uncertainty effect did significantly differ from 50).

In other ways, though, Experiment 2 worked much better than Experiment 1. In Experiment 1, there was a huge range of different strategies in how participants performed. In Experiment 2, participants seemed to have relatively homogeneous strategies. This makes the results much easier to interpret.

3.4   Experiment 3

The problem with Experiment 2 is that the interrupting tone task and the original digit task both required manual key presses. Naturally these must interfere with one another, and this interference may dwarf any effect of parallel response-selection. To correct this problem,

90

we changed the interrupting tone task to be an audio-vocal task with vocal responses. In all other ways, Experiment 3 was the same as Experiment 2.

3.4.1 Method

Participants. Four undergraduate students participated as paid volunteers, none of whom were in a previous experiment.

Apparatus. The apparatus was the same as in Experiments 1 and 2.

Experimental Design and Procedure. The design and procedure was the same as in Experiment 2, except that the interrupting tone task involved vocal responses. Participants heard a tone and had to say "1" "2" or "3" depending on whether the tone was low, medium, or high-pitched.

3.4.2 Results

One participant was removed from further analyses, due to having an excessively large tone-task RT on dual-task trials (160 msec more than next nearest participant). Since we only had three participants remaining, we will not report many t-tests, as we do not expect them to be significant due to the lack of degrees of freedom.

Pure-Task Blocks

Digit and Tone Task. We again examined our response-selection effect. Table 3.7 summarizes the mean and standard deviation for each different condition. RTs for whole stimuli (stimuli that were not degraded), at Low-SOA, were significantly slower in the hard-digit task (M = 398 msec, sd = 60.5 msec) than in the easy-digit task (M = 319 msec, sd = 47.8, paired t-test t(2) = 21.10, p < .005). We found the same result for degraded stimuli,

91

t(2) = 14.72, p< .005. At Mid and High-SOA, we did not find a significant RT difference, but results were consistent with our findings in Experiment 2.

We next examined whether stimulus degradation had an effect, but did not find any significant differences. However, the trend was similar to that found in Experiment 2, i.e. degraded stimuli took longer to process than whole stimuli.

The RT for the tone task had a mean of 389 msec, similar to performance in Experiment 2.

We again examined "catch" trials, where the participant was never given a "go" signal and so was not supposed to respond at all. We found that participants incorrectly responded on 17.3% of easy whole trials, and 8.6% of easy degraded trials. For the hard-digit task with whole stimuli, participants made 11.1% catch-trial errors and 9.3% for the hard degraded stimuli.

Condition (msec)	Low-SOA	Med-SOA	High-SOA
Whole Easy	319 (47.8)	200 (24.7)	180 (11.9)
Degraded Easy	373 (61.9)	225 (37.4)	184 (12.7)
Whole Hard	398 (41.6)	213 (31.5)	190 (16.9)
Degraded Hard	469 (60.5)	240 (42.0)	191 (20.0)
Tone	389 (19.2)	—	—

Table 3.7. Results for pure-task blocks.

Mixed-Trial Blocks

Uninterrupted Trials. Table 3.8 shows the results from the single-task trials in the Easy-Digit Mixed-Trial Blocks. The pattern is very similar to what we saw in the pure trial blocks, as can be seen by the fact that we now have rather small uncertainty effect. Response selection was significantly slower for the hard-digit task than for the easy-digit task in the Low-SOA condition (for whole stimuli, t(2) =10.95, p< .01; for degraded stimuli, t(2) = 18.94, p< .005).

Condition (msec)	Low-SOA	Med-SOA	High-SOA
Whole Easy	344 (23.6)	235 (26.2)	183 (14.0)
Degraded Easy	394 (47.4)	241 (26.1)	185 (12.4)
Whole Easy Uncertainty Effect	24	35	4
Degraded Easy Uncertainty Effect	20	16	1
Whole Hard	422 (28.5)	235 (31.9)	193 (20.00)
Degraded Hard	447 (51.2)	252 (34.2)	188 (15.9)
Whole Hard Uncertainty Effect	24	22	3
Degraded Hard Uncertainty Effect	-22	12	-3

Table 3.8. Results for uninterrupted trials on mixed-trial blocks.

The only uncertainty effect that significantly differed from 0 occurred with in the Mid-SOA for easy whole stimuli, i.e. 235 msec > 200 msec, $t(2) = 6.70$, $p<.03$. Of course, the small sample size means we cannot interpret the null effects of these other effects, but the trend was for a much greater reduction in the uncertainty effects compared to those in previous experiments.

Interrupt Trials. Table 3.9 shows the results from interrupt trials. The greater emphasis that we placed on the interrupting tone task seemed to have worked, as the interrupting lag was much lower than in previous experiments. We also found no real difference in RTs for the tone task across conditions, although there was a marginal difference between easy-whole and easy-degraded conditions, $t(2) = 3.20$, $p=0.085$. The easy-digit tone RT (M = 388) did not differ from the hard-digit tone RT (M = 403), $t(5) = 1.72$, p=ns.

We do find a significant effect of task difficulty on interrupted digit-task RT. Combining easy-whole and easy-degraded RTs compared to hard-whole and hard-degraded RTs, we found that the easy-digit task interrupted RT of 136 msec < hard-digit task interrupted RT of 171 msec, $t(5) = 4.01$, $p<.02$.

We also examined how often participants failed to interrupt, responding to Task 1 when they should have responded to Task 2. There were more of these errors than in Experiment

	Easy-Whole	Easy-Degraded	Hard-Whole	Hard-Degraded
Interrupted Digit-Task RT	139 (61.7)	132 (48.8)	160 (80.1)	181 (56.6)
Tone-Task RT	376 (33.9)	401 (37.7)	396 (25.2)	410 (22.6)
Interrupting Lag	-13	13	8	21
Fail to Interrupt (%)	13.91	6.90	5.73	2.44

Table 3.9. Results for interrupted trials on mixed-trial blocks.

2. With whole stimuli in the easy-digit case, participants made these errors on 13.9% of the trials, compared to 6.9% for degraded stimuli, versus 5.7% for the hard-digit task with whole stimuli, and 2.4% for the hard-digit task with degraded stimuli.

Computational Modeling. All models were made using CORE, the constraint-based optimizing reasoning engine (Howes et al., unpublished manuscript). CORE allows the modeler to disentangle strategy space from architectural constraints. Details of the modeling are discussed below.

3.4.3 Modeling Results

All models were made using CORE (constraint-based optimizing reasoning engine). All models were calibrated based on parameters estimated from pure-block results in the Low-SOA condition only. Models were then made for each subject in each condition. So each model was fit to six different data sets, each participant's easy and hard dual-task blocks, for each of three different participants. All the models used estimates from EPIC (Meyer & Kieras, 1997a,b) as the bases for the duration of different processes. For example, all of our models included a time to detect the tone, which we asssumed came from a Gamma Distribution with a mean of 50 msec. Because only one hand was ever used, we assumed motor preparation also had a mean time of 50 msec, as well as motor initialization. We used these assumptions for vocal responses, as well. We assumed a 25 msec mean for the vocal apparatus to record the response, and a 10 msec mean for the manual apparatus. We assumed that the easy-digit response selection only took 1 cognitive cycle, whose duration

came from a Gamma distribution with mean 50. We assumed that the tone task, a compatible task, also only took one cycle for selecting a response. For the hard-digit task, we assumed that it took 2.67 cycles to select the response.[1] All the models discussed below also include an "unlock" process that must be called before Task 1 can be completed. We used the EPIC assumption that the mean unlocking time is 100 msec (Meyer & Kieras, 1997a, p. 26).

Parameter	Estimated Mean (msec)
Motor Preparation (1 feature)	50
Motor Initiation	50
Whole Digit Identification	132
Degraded Digit Identification	193
Easy-Digit Response-Selection (1 cycle)	50
Hard-Digit Response-Selection (2.67 cycles)	133
Tone Response-Selection (1 cycle)	50
Tone Detection	50
Tone Identification	189
Unlock	100
Lock	50
Working Memory Gating Time	25
Vocal Apparatus	25
Manual Apparatus	10

Table 3.10. Estimated mean times for parameters in models (derived from EPIC, Meyer & Kieras, 1997a,b).

Models were compared to human data along the following criteria: % Fail to Interrupt, Tone-Task RT, Interrupted-Digit RT, and Uninterrupted-Digit RT in Low/Medium/High-SOA conditions. Architectural assumptions will be discussed concerning each individual model. Models are presented using a program called CogTool (John et al., 2004). To make comparisons as easy as possible, all models shown below were derived from the same participant in the same condition, i.e. participant 1 in the easy-digit condition. See 3.11 for a summary of models.

---

[1]This follows from the fact it takes one cycle to respond to the digits 1 - 3, and we assume the digits 4 - 9 are done serially and take a cognitive cycle each. Hence 1/3rd of the time 1 cycle, 2/3rd of the time 3.5 cycles on average = 1/3 * 1 + 2/3 * 3.5 = 2.67

Model	Architecture	Summary	Results
Deferment	RSB	Model with no response-reversal errors because Task 1 resumes only after Task 2 completed	Too slow at fitting interrupted RT
Early-Deferment	RSB	Task 1 proceeds as soon as wait-time ends if tone not detected; if tone detected Task 1 proceeds after Task 2 completed	Poor fit for interrupted RT
No Deferment/No Unlock	RSB	Task 1 RS occurs as soon as Task 2 RS finished, then proceeds immediately to motor execution	Too many response reversals
No Deferment/Unlock	RSB	Task 1 RS occurs after unlock when Task 2 RS finished	Generally good fit
Lock	No-RSB	Task 1 RS first locked if interruption detected, then unlocked after Task 2 RS finished	Interrupted Digit Task too slow
Motor Unlock	No-RSB	Motor Prep. begins locked for Task 1, unlocked after RS for Task 2 finished	Generally good fit

Table 3.11. Summary of computational models, RS=response selection, RSB = response-selection bottleneck.

Deferment Models. In these models we assume that there is a response-selection bottleneck, so response selection can only occur for one task at a time. A visualization of an interrupt trial from our first model is shown in Figure 3.13. The grey boxes are associated with processes for Task 2, the tone task, and the black boxes are processes associated with Task 1, the digit task. The scale above shows the time in seconds. If we follow the path along the grey boxes, we see an initial SOA of 100 msec, which is when the tone would occur on a dual-task trial. Then we have to detect and identify the tone. Tone identification occurs in the bottom row, tone detection occurs in the grey box right above it. After perception, the participant immediately proceeds to response selection, and then proceeds to motor preparation, motor initiation and then apparatus-recording time. So in this model, there is no delay in performing Task 2, as all of these stages also have to occur on pure-tone trials. Hence the interrupting lag should be 0.

The black boxes are associated with processes for Task 1. At time 0, two black boxes

begin: Wait-for-task and perceive. Perceive is the process to identify the digit. Note that perception can occur in parallel, so although we see the grey box associated with tone detection in-between the two black boxes, that is occurring simultaneously with digit identification. Wait-for-task is a participant's mental waiting time before beginning response selection for Task 1. A participant has to be worried that if she begins response selection for Task 1, and then has to do response selection for Task 2, this may result in a significant delay. Hence if participants have not yet detected any "go" signal, they will wait some time before they begin to select a response for Task 1. If participants do not wait long enough, then they will begin response selection for Task 1 as soon as digit perception is completed, resulting in a delay in Task 2 if they then detect the tone. If, as in Figure 3.13, tone detection occurs before wait-for-task is complete, then participants do response selection for Task 2 first. The mean duration of wait-for-task is varied across strategies, ranging from 150 msec to 350 msec. Note that if participants get the "go" signal before wait-for-task is completed, then they know it is a single-task trial, and they will proceed to Task 1 response selection as soon as they finish digit identification.

After Task 2 response selection is completed, this model ensures it will not make an response-reversal error[2], by waiting until the motor initialization has been completed for Task 2, and only after a deferment will Task 1 resume being processed. The durations of these "defer" times also are strategically varied, from 1 to 25 to 50 msec. After the defer time has ended, Task 1 motor processes are unlocked, and then Task 1 response is made.

In our "early-deferment" model, when the wait-for-task process is not long enough, then no deferment or unlocking for Task 1 is needed, and a response reversal is likely. Figure 3.14 shows a trial where the modified model makes a response reversal. Here, the black boxes associated with Task 1 for motor preparation and initiation occur before the corresponding grey boxes associated with Task 2. We made this change because almost all human participants in all conditions had some response-reversal errors (except for one par-

---

[2]Response-reversal errors could equally well be considered as fail-to-interrupt or, as in Chapter 2, order errors
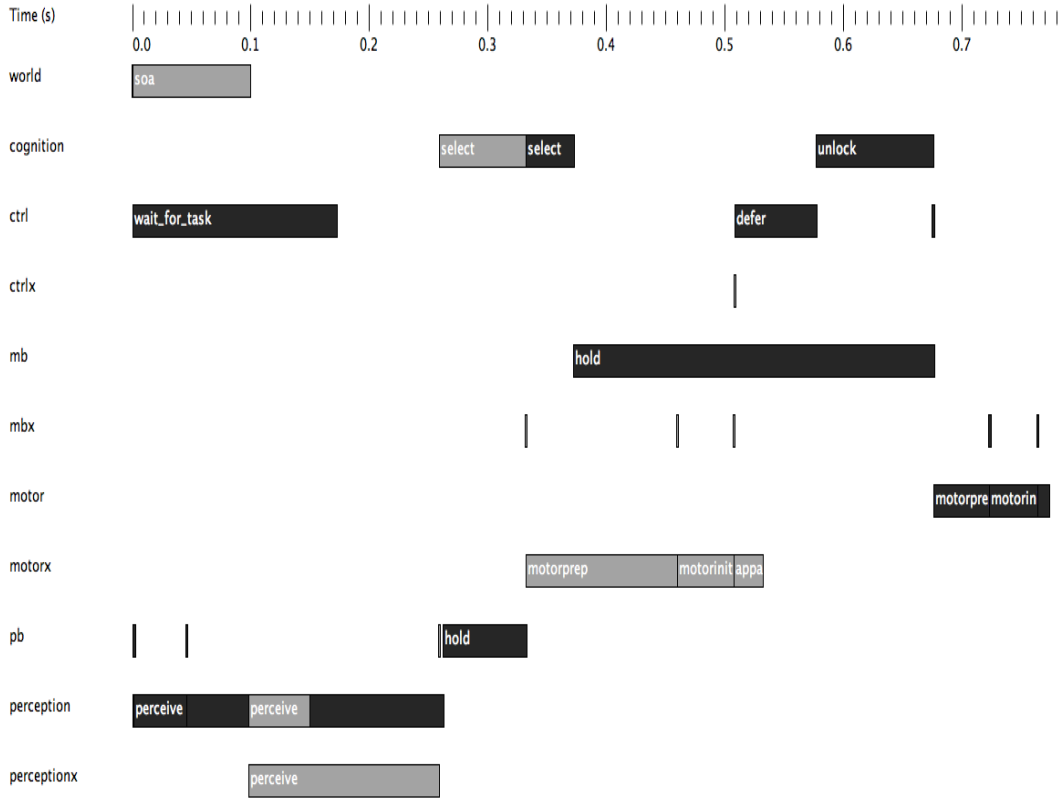
Figure 3.13. Visualization of "deferment" model, easy-digit condition.

ticipant in the "hard-digit" condition), so we felt it necessary that our model could produce these as well.

Our CORE models automatically compared these different strategies, and found out which ones were optimal in terms of maximizing points earned per trial. With this version of "defer" the same strategies were chosen as optimal for all participants in all conditions: a long wait time (300 or 350 msec, there was no difference in these strategies), and "defer" duration did not matter, except in one case where the long duration was suboptimal. Since all the other strategies were equally optimal, we will examine the strategy with a 300 msec wait-for-task time and a 1 msec mean duration for defer.

Our model at this time predicts no response-reversal errors. Table 3.12 shows a comparison of the model and human results. Note that the human data was calculated a bit

Figure 3.14. Visualization of "early-deferment" model response-reversal error, easy-digit condition.

differently than above, to make it more directly comparable with the model data.[3]

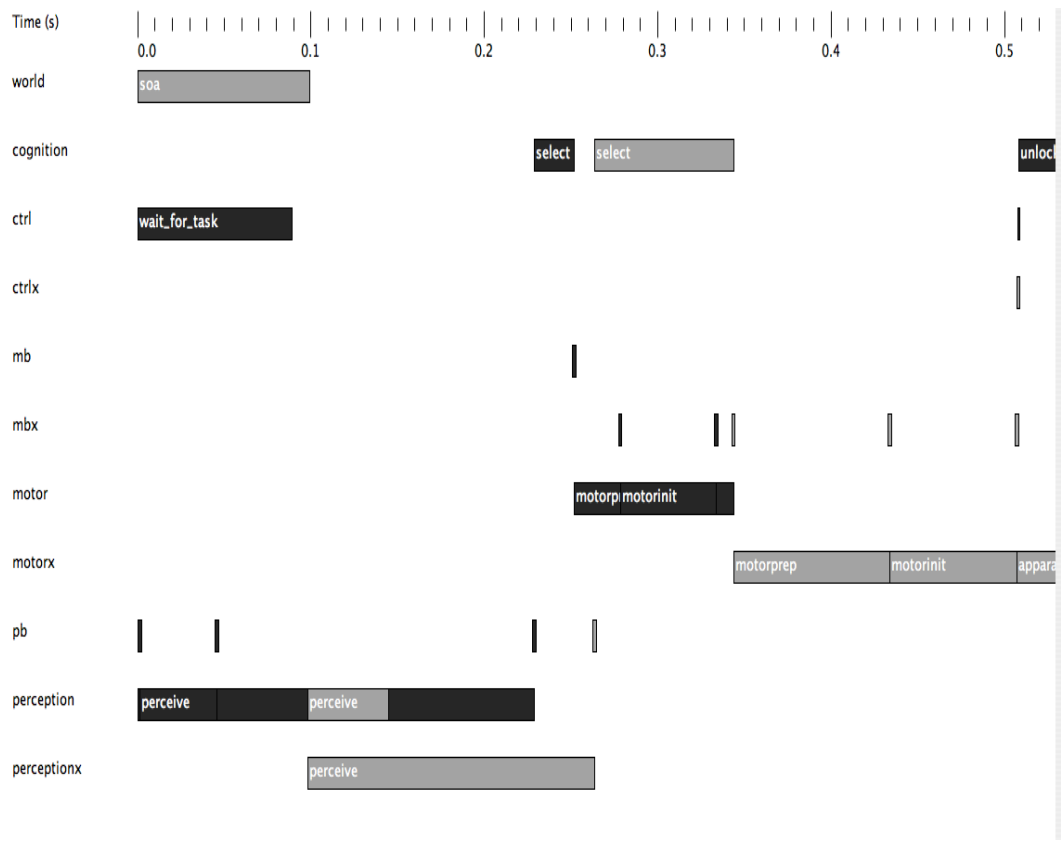|	Human Easy	Model Easy	Human Hard	Model Hard
Interrupted Digit-Task RT	159	191	195	231
Tone-Task RT	403	385	437	388
Response-Reversal Errors (%)	10.42	0.77	4.11	0.03
Whole Single-Task Trials, SOA low	349	316	418	387
Degraded Single-Task Trials, SOA low	398	376	461	468
Whole Single-Task Trials, SOA med	240	170	244	215
Degraded Single-Task Trials, SOA med	253	196	263	242
Whole Single-Task Trials, SOA high	181	159	210	161
Degraded Single-Task Trials, SOA high	186	155	194	165

Table 3.12. Comparison of human data and "Deferment" model results for mixed-trial blocks.

Many differences leap out in this comparison. First of all, the optimal strategy makes very few response-reversal errors. Note that some of the suboptimal strategies do make more response-reversal errors, however otherwise they tend to have the same problems as this optimal strategy. The fit for the tone task is better in the easy-digit condition than in the hard-digit condition, because our model's tone-task RT is not affected by the difficulty of the digit condition, whereas our participants were slower in doing the tone task in the hard-digit condition than in the easy-digit condition. For all three participants there was a significant difference between the model and actual tone-task RT in the hard-digit case. The model is also too slow on the interrupted trials. Hence the dual-task fit is significantly off. On uninterrupted trials, the model is too fast at High-SOA and at Mid-SOA. The model does a much better job at the Low-SOA (note that although the model differs by 20 or 30

---

[3]We did not examine results by block, as we did above, because the model did not generate data by block; we also compared distributions and there would not have been enough trials per block to do this

msec often on average, there is more noise in these distributions, so these differences are often not statistically significant).

We first modeled "deferment" strategies because some models of the PRP task used deferment strategies to prevent response-reversal errors (Howes et al., unpublished manuscript). However, the defer model makes too few of these errors and takes too long to resume the interrupted digit task. The fact that "deferment" is involved in both of the ill-fitting models discussed above indicates human do not use a "deferment" strategy. Note that if we assumed a no-bottleneck architecture, we would still have these problems. Hence the next strategy we will consider is a "no-deferment" model. Given our instructions did not particularly emphasize preventing response-reversal errors, we felt it was reasonable that participants would adopt a simpler strategy that resulted in faster overall performance, but at the risk of making more of these types of errors. Figure 3.15 shows a visualization of an interrupt trial from the model.

Under this model, Task 1 begins every trial with response selection locked. It is unlocked as soon as it gets a control signal, either from response selection finishing in Task 2 or if wait-for-task finishes before the tone is detected. Note that due to variance, at times unlock and the motor processes associated with Task 1 could finish before the motor processes associated with Task 2, although this is rare. As in our previous model, we find that optimal strategy is to wait 300 or 350 msec (although for most participants and conditions, 250 was also optimal). In Table 3.13 we compare the results from an optimal strategy of waiting 300 msec with human data. This model is a much better fit on the single-task trials than the previous model, and virtually every statistic is closer to the human data than the previous model, and it is more parsimonious, using only a parameter for mean wait-time compared to two parameters (for mean defer-time and mean wait-time) in the previous model.

Although this is arguably our best model (it is arguable because some models fit best for an individual participant in a particular condition, but not overall), there is still some
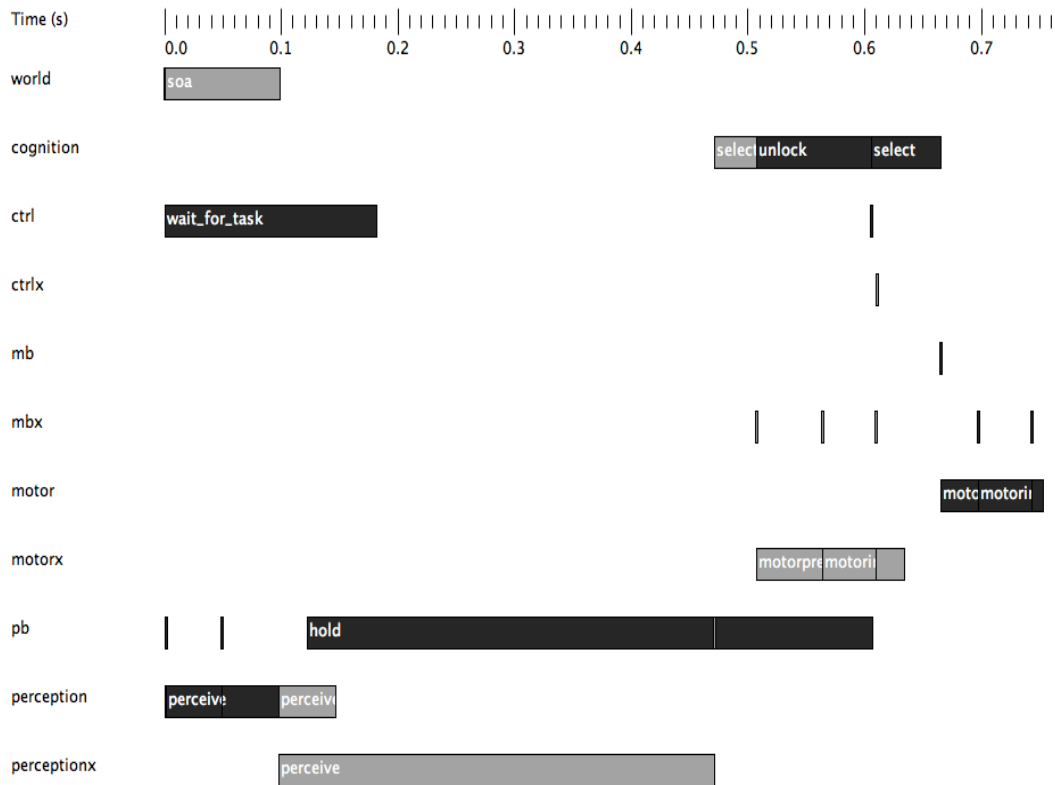
Figure 3.15. Visualization of "no-deferment" model, easy-digit condition.

things that need to be improved. Namely, the tone-task RT in the model does not vary if we use the easy or hard-digit task as Task 1, but it does vary in the human data. In most other ways, though, this model is a good fit. We have also tried a few other variations of this no-deferment model. In one, we did not "unlock" Task 1 response selection, we just had it immediately follow Task 2 response selection. This model did a very poor job of fitting the data, giving strong support to the AEC notion of "unlock."

We also did a variation of our "no-deferment" model where there was no structural response-selection bottleneck, i.e. two responses could be selected simultaneously. In this case, Task 1 motorprep was initially locked at the start of every trial. If wait-for-task finished before the tone was detected, then "unlock" immediately proceeded, simultaneously with response-selection for Task 1. As soon as unlock and Task 1 response selection were finished, motor preparation began. Note that for Task 2, no process began the trial "locked"

subject 2 strategy 1 easy trials task 2 density function

subject 2 strategy 2 easy trials task 2 density function

subject 2 strategy 3 easy trials task 2 density function

subject 2 strategy 4 easy trials task 2 density function

Model Strategy 1
Subj. 2,
Tone Task

Model Strategy 2
Subj. 2,
Tone Task

Model Strategy 3
Subj. 2,
Tone Task

Model Strategy 4
Subj. 2,
Tone Task

Task 2 RT
average strategy 1 payoff 73.5

Task 2 RT
average strategy 2 payoff 125.5

Task 2 RT
average strategy 3 payoff 161.6

Task 2 RT
average strategy 4 payoff 165.2

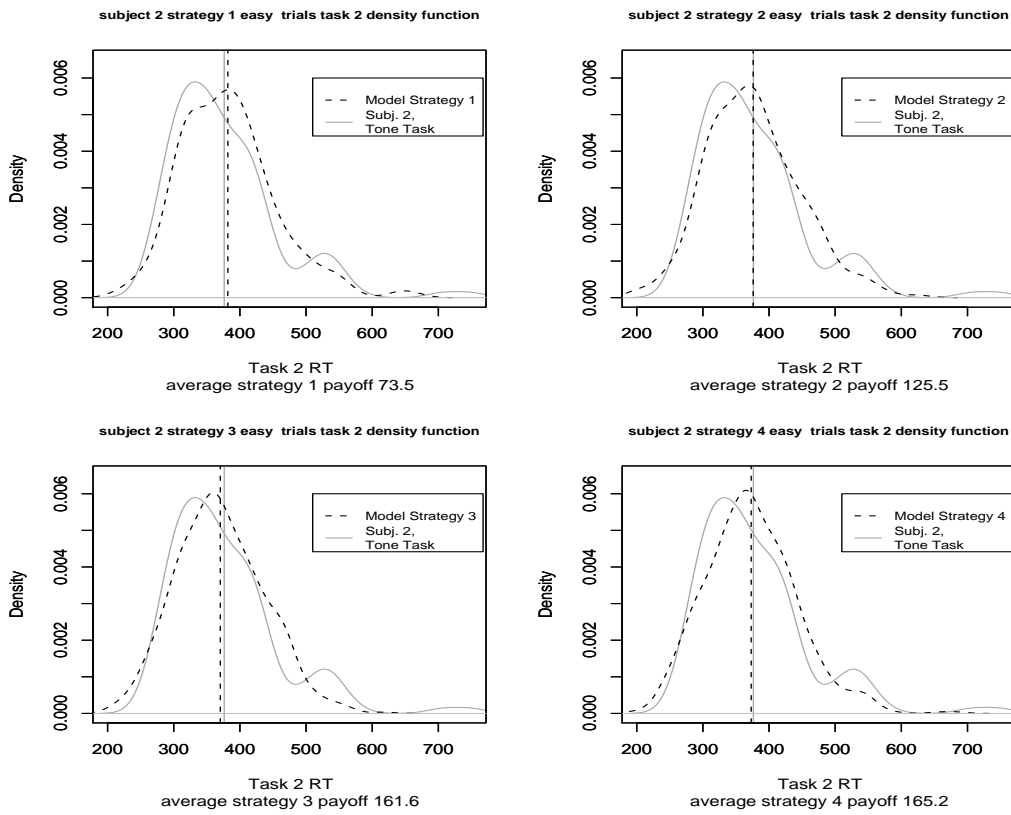Figure 3.16. Density plot of tone-task RT on interrupt trials on mixed-trial blocks for a particular subject across strategies compared to No-Deferment/Unlock model.

	Human Easy-RT	Model Easy-RT	Human Hard-RT	Model Hard-RT
Interrupted Digit-Task RT	159	139	195	220
Tone-Task RT	403	386	437	390
Order-Errors (%)	10.42	1.54	4.11	0.52
Whole Single-Task Trials, SOA low	349	322	418	405
Degraded Single-Task Trials, SOA low	398	369	461	458
Whole Single-Task Trials, SOA med	240	191	244	270
Degraded Single-Task Trials, SOA med	253	218	263	277
Whole Single-Task Trials, SOA high	181	184	210	177
Degraded Single-Task Trials, SOA high	186	186	194	184

Table 3.13. Comparison of human data and "No-Deferment" model results for mixed-trial blocks.

so no unlock was needed. Hence Task 2 response selection immediately followed Task 2 perception, and motor preparation immediately followed response selection. If the tone was detected before wait-for-task was finished, then "unlock" began immediately following Task 2 response selection. Hence, on these trials, Task 1 motor preparation processes began about 100 msec after Task 2 motor preparation began. However, these models were too fast on average, predicting interrupted RT should all be around 100 msec, regardless of the difficulty of the digit task. These models were also too fast for the uninterrupted trials and for the tone task. However, this model was a better fit than the standard "no-deferment" for one participant.

We also made one model that did not begin in deferred mode. Only if the tone was detected before the wait for it was over would Task 1 response selection then be "locked," a process that took 50 msec on average. After locking, Task 1 response selection would be "unlocked" after Task 2 response selection was complete. This model was a poor fit for every participant.
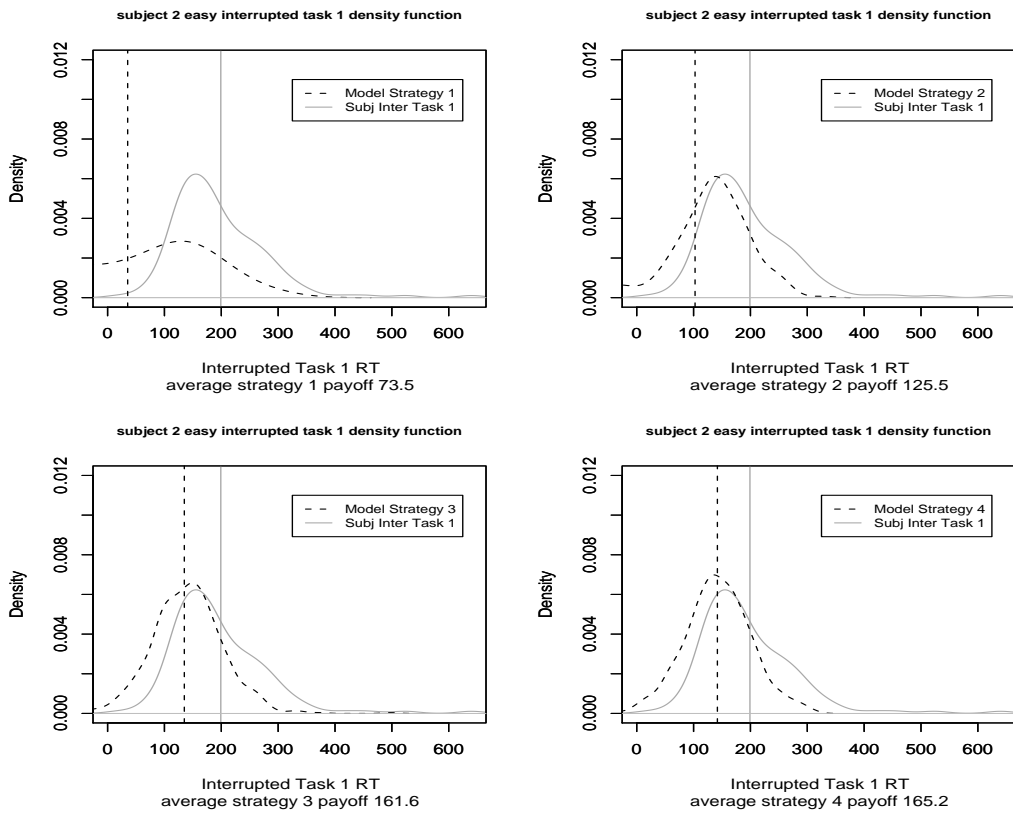
Figure 3.17. Density plot of interrupted easy-digit RT on interrupt trials on mixed-trial blocks for a particular subject across strategies compared to No-Deferment/Unlock model.

3.4.4 Discussion

In Experiment 3, we developed a way to get people to successfully interrupt. By ensuring there was no motor interference, the uncertainty effects that concerned us in Experiment 2 largely went away. More importantly, the interrupting lag also largely disappeared, indicating participants could do the interrupting task as quickly as they could do that task by itself, which is important given the payoff we gave participants emphasizing this.

Modeling the experimental results in CORE was quite illuminating. First of all, we found that a very parsimonious model was our best-fitting model. This model required us to assume that participants began each trial with response selection locked at the start of each trial, similar to the conservative strategy from the adaptative executive control (AEC) model (Meyer et al., 1995; Schumacher et al., 1999). By unlocking response selection for Task 1 depending on a wait time and Task 2 processes, we created a reasonable approximation of all participants' results.

However, in some ways our model was still lacking, and that helps us realize what we still need to work on in understanding how people interrupt. For example, in our models the tone-task RT was the same if performed in an easy-interrupt or a hard-interrupt block, i.e. the difficulty of the initial digit task did not affect performance of the interrupting tone task. Participants, though, were 40 msec slower when performing the tone task during hard-interrupt blocks than during easy-interrupt blocks. However, with only three participants in this experiment, we need more data to see just how robust this finding is.

We also examined optimal strategies in each of our models, and found that most optimal strategies had fewer response-reversal errors than humans made. In fact, optimal strategies that maximized points per trial for a given model found there were no response-reversal errors. Because participants were making these errors, the model strategies that best fit human data were suboptimal strategies. Figure 3.18 shows a plot of uninterrupted RT by SOA, and how the optimal strategy is not a particularly good fit, whereas one of the suboptimal strategies fits the data much better. It is unclear if participants were choosing a

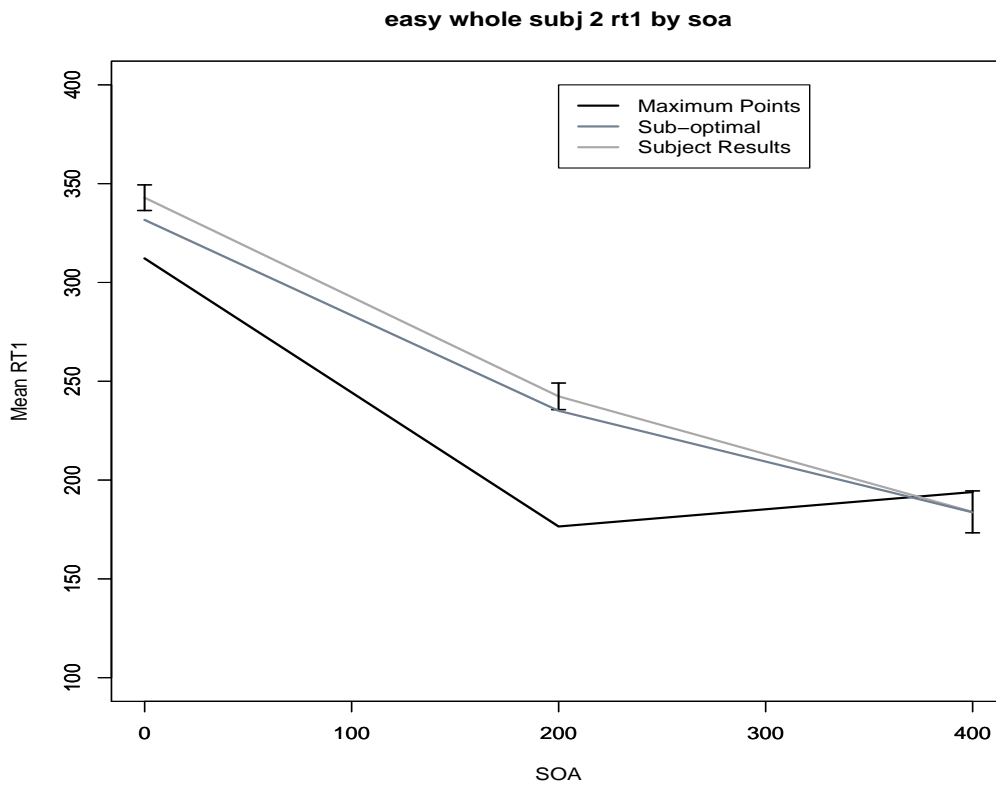**easy whole subj 2 rt1 by soa**

Figure 3.18. Plot of easy-digit RT by SOA for uninterrupted trials on mixed-trial blocks for a particular subject; note that maximum points represents the optimal strategy for No-Deferment/Unlock model, and is a poor fit to the human data, whereas a suboptimal strategy is a good fit to the human data.

107

suboptimal strategy, if there is some cognitive constraint preventing optimal performance, or if there is some other factor contributing to these response-reversal errors not addressed in our models.

3.5   General Discussion

To return to our example from the introduction, if a pilot is jarred from his flying duties by an alarm, we have learned is that it is possible to train the pilot to deal with the interruption quickly. However, it does not seem possible to deal with the alarm if the pilot was in immediate mode on the previous task. In Experiment 1, we learned how debilitating an interruption is when trying to do the initial task as quickly as possible. Interruptions force you to lose the previous progress you made on that initial task, and participants find it difficult to stop progress in the initial task. We found a much larger interrupting lag than that predicted by EPIC assumptions, which assume merely a 50 msec mean cost in order to remove the goal for Task 1 from memory. So it appears trickier to interrupt than straightforward analyses of cognitive architectures would predict.

It is worth nothing that interruptions were very stressful for our participants. Once we trained participants to do a task as quickly as possible, it was clear they did not feel natural interrupting that task. Quite a few participants referred to it as the most difficult experiment they had ever done. As such, it was very interesting to learn what strategies they were exploring, as some of the strategies seemed to be based on the goal of relieving their stress level rather than on maximizing their points. As such, we had to use many different methods just to help participants get more comfortable with the task and accept the fact that they were not able to interrupt 100% of the time. In Experiment 1, we think we made participants try as hard as possible to maximize their points, but the stressfulness of the task, even after three sessions of training, means that we cannot entirely discount the possibility that they may have adopted suboptimal strategies that were less cognitively-

demanding.

These other sub-goals may explain in part why we found so many different strate-gies amongst our participants. Although we do not have records of the ethnicity of the participants in Experiment 1, it would be interesting to see if there were cultural differ-ences amongst the participants. As we will see in Chapter 4, it is possible that Western participants favor strategies that require switching, but that East Asian participants prefer strategies that require grouping because then the two tasks may be done relatively simulta-neously.

Experiments 2 and 3 were far less stressful for our participants, as we had trained them to perform Task 1 in deferred mode. In this mode, response selection goes to working memory, instead of directly to motor processors, allowing a natural interruption point. In-terestingly, Experiment 2 had rather similar results as in Experiment 1, with the notable exception that the range of strategies was much more constrained. This allowed us to begin to look at participants as a group, not just as a collection of subgroups.

We realized to best interpret the results, we had to first remove other sources of inter-ference, which we did by making the responses in different modalities in Experiment 3. The results of that experiment seemed quite meaningful and interpretable, so we modeled performance using CORE. We found the best overall model to be one that assumed that each trial began with response selection for Task 1 locked, and only unlocked after Task 2 response selection was completed or after the participant's internal clock told them they had waited long enough to see if it was a dual-task trial or not.

Hence for our pilot having to deal efficiently with an alarm, we had better hope a vocal-response is required, and that he was in deferred-mode originally. This may also be part of the reason that we have copilots on flights, as one pilot could always be doing a task that allows them to remain in deferred mode while the other is in immediate mode, allowing one pilot to always be able to deal with emergency interruptions that occur.

Our paradigm and results are a form of multitasking not heavily studied, with poten-

tially broad implications. The Psychological Refractory Period (PRP) studies have been heavily used to study the existence of a response-selection bottleneck (RSB). Unfortunately, as our modeling shows, our results are not conclusive one way or the other concerning whether a RSB exists or not. Although our results were certainly consistent with there being a RSB, because our best model only had one response selected at a time, in fact we think our model better supports the Meyer et al. (1995) model of Adaptive Executive Control. One participant had his Interrupted-Digit RT fit better via the no-bottleneck model than the bottleneck model (this participant had 100 msec RTs for the interrupted-digit task in both the easy and hard-digit case). Because we only had a few participants in this experiment, concern about over-fitting means we wanted to use the model that fit all of the participants best overall. But this result at least tentatively supports the AEC model of different strategies used in multitasking.

One concern in our study is that the "go" signal on single-task trials was visual, so that may result in some interference with processing the stimuli. Plus, the color of our visual go-signal indicated if the participant had to do Task 1 or Task 2. But because the go-signal involved motion, participants may have detected the motion before they detected the color of the signal. This may have resulted in more response-reversal errors than our model would predict, as a participant may interpret the motion as the go-signal for Task 1. Because our model does not consider perceiving color and motion separately, this may be why our optimal strategies underestimate the number of response-reversal errors.

Because there are so few studies in task-interruptions, the results here are rather tentative. More research is needed: how does training affect performance, what if Task 1 has a vocal response and Task 2 a manual response, do we need to encourage participants to explore the strategy space more? Will there be an "Anti-PRP" effect where the lower the SOA, the less interference in Task 1 and Task 2?

Chapter 4

Cultural Differences in Multi-Tasking

4.1   Introduction

The movie *Pushing Tin* featured two rival air-traffic controllers competing to see who was the better controller. One of these controllers, Nick Falzone, was an American who was the best at processing all the information needed to guide planes, and was shown in the movie switching between processing and relaying information for one plane and then rapidly proceeding to the next. This sequential strategy made him the best controller, until the newly-hired rival controller, Russell Bell, proceeded to outshine Nick as a controller. Russell was half-Choctaw Native American, so he was likely raised in a different culture than Nick. Choctaw culture, like most Native-American cultures, is considered more collectivistic than European-American culture, which is considered individualistic. Could this cultural influence have aided Russell in his career pursuit, influencing him to choose different, more simultaneous multitasking strategies, garnering him an advantage as an air-traffic controller? Although the movie did not delve into these issues, this chapter documents how people's socio-cultural milieu influences people's multitasking strategies.

Is there reason to believe that culture could influence how people choose to multitask? Research has demonstrated that cognitive processes can be modified, sometimes dramati-

cally, by socio-cultural contexts in which people are engaged. For example, Polk & Farah (1998) found that Canadian postal workers who often read zip codes with both letters and numbers had different performance on visual-search tasks than normal controls that primarily read numbers and letters in separate contexts. More generally, research has demonstrated that cognitive and attentional characteristics of people vary widely as a function of largely demarcated cultural contexts such as West and East (Kitayama et al., 2008; Nisbett, 2003). The current work seeks to contribute to this emerging literature by examining whether individuals with different cultural backgrounds vary in their strategies to perform two complex sensory discrimination tasks.

Over the last two decades, it has been proposed that whereas practices and meanings of Western cultural contexts emphasize the independence and separatedness of each distinct individual, those of East Asian cultural contexts highlight the interdependence and connectedness of people constituting a social group (Kitayama et al., 2008; Markus & Kitayama, 1991; Triandis, 1995; Shweder & Bourne, 1982). They have also argued that these different cultural contexts entail certain cognitive consequences. In particular, by engaging in the Western independent practices and meanings, individuals may become more attentive to each distinct individual and draw inferences about him or her while giving only scant attention to the social surrounding. In contrast, by engaging in the Eastern, interdependent practices and meanings, individuals may become more attentive holistically to the entire context in which the target person is embedded. In support of this analysis, numerous studies have shown that when asked to explain another person's behavior, people from Western contexts are more likely than those from Eastern contexts to refer to dispositional (e.g., the protagonist's personality or attitude) rather than situational factors (e.g., social atmosphere and norms; Kitayama et al., 2006; Morris & Peng, 1994; Nisbett et al., 2001).

More recently, it has become evident that the cognitive consequences of culture are not limited to the domains of social perception and social inference. In fact, analogous cross-cultural differences in attention have been shown even when the stimulus at issue

112

is entirely non-social (Nisbett & Masuda, 2003). For example, Masuda & Nisbett (2001) had Japanese and American participants watch 10 videos of a clear focal fish swimming in a virtual aquarium, replete with background fish and scenery. Participants were then given a recognition memory task either of the focal fish or a novel fish with the previously-seen background, no background, or a novel background. Americans were only influenced by the focal fish; the background made no difference in performance. The Japanese, in contrast, made more errors when the background was different than in the original video. The evidence is consistent with the hypothesis that compared to Americans, Japanese are more holistic, with attention distributed over the whole scene, to the extent that people from the two cultures appear to even look at images differently (Chua et al., 2005).

Fish can still be interpreted in anthropomorphic terms. Yet, Kitayama, Duffy, Kawamura, and Larsen (2003) have shown a similar cross-cultural difference in attentional allocation with purely geometric stimuli. Participants were presented with a frame with a line printed in it. They were then shown another frame of different size. With this Framed Line Test (FLT), the researchers found that Americans performed better when asked to draw a line of the exact length in the second frame than to draw a proportional line, suggesting that their attention is narrowly focused on the goal object - namely, the line. In contrast, Japanese performed better when asked to draw a proportional line than to draw the exact line, indicating that their attention is more holistically allocated to both the goal object (the line) and its surrounding frame. More recently, Hedden, Ketay, and Aron (2008) used a modified version of the FLT and measured brain activations via fMRI. They found a significant frontal activation when Caucasian Americans engaged in the relative judgment that was made very difficult (relative to when the judgment was easy). Conversely, Asian Americans showed a similar frontal activation when they engaged in the absolute judgment that is made difficult (relative to when the judgment was kept easy). This suggests that individuals needed to actively regulate their attention when performing a task that was culturally incongruous.

Extending these previous studies, our research examines cross-cultural differences in multitasking with perceptual-motor and cognitive tasks that involve several information-processing stages: stimulus encoding, perceptual discrimination, response selection, and movement production (Meyer et al., 1988; Sternberg, 1969a). In the three experiments that follow, we have tested participants from Western individualistic and Eastern collectivist societies. In experiments 1 and 2, these participants must perform two concurrent tasks while trying to schedule processing stages for each task so as to complete them as quickly and accurately as possible. In these experiments, performing two tasks optimally would require participants to minimize total costs, to process the two tasks as simultaneously as possible. However, if it is not possible for an individual to process both tasks simultaneously with no costs[1], then a new optimal strategy may be to process one task as quickly as possible, and then proceed to the other task. This would result in the minimum cost of the two tasks being as small as possible, but having a relatively large maximum cost on a trial. In experiment 3, participants have to switch between two tasks, but only respond to one task on a given trial, so optimal performance would require participants to be able to switch between two tasks while minimizing any associated cost (Allport et al., 1994; Allport & Wylie, 1999; Rubinstein et al., 2001).

These studies help answer fundamental questions about the diversity of human information processing, the heterogeneity of people's cognitive architectures, and the adaptability of their task-scheduling strategies (cf. Meyer & Kieras, 1997a,b, 1999; Welford, 1952). Do people from individualistic and collectivist societies all have a common structural cognitive bottleneck? To what extent may they adopt different strategic modes of multitasking? Are members of some cultures more prone than others to select responses and produce movements simultaneously for multiple tasks?

To obtain answers, we have used a procedure first introduced by Schumacher et al.

---

[1]These individual differences may be caused by different architectures across individuals, such as some individuals having a response-selection bottleneck and others not. However, another explanation is that perceptual times vary across individuals, resulting in different probabilities for response selection for the two tasks overlapping

(2001). It requires participants to perform visual-manual and auditory-vocal discrimination tasks quickly and accurately under both single-task and dual-task conditions. On dual-task trials, which are designed to encourage "perfect time-sharing," both tasks must be performed concurrently with equally high priority, whereas on single-task trials, only one task must be performed. By comparing reaction times (RTs) and error rates (ERs) from the two trial types, measures of dual-task performance costs are derived, and inferences are made about how participants cope with the challenging demands of this procedure. Specifically, we may discover systematic cross-cultural differences in participantsŕ preferred strategies of task scheduling, and assess to what extent parallel or serial processing is manifested by members of one culture (e.g., East Asians) versus another (e.g., North Americans).

We predicted that in Experiments 1 and 2 East Asians would exhibit a tendency toward "parallel" strategies, where they process the two tasks as simultaneously as possible. Westerners, in contrast, favor a sequential strategy where they first processed one task and then the other. Hence East Asians should have less total dual-task costs, and a lower maximum cost. In Experiment 3, we predicted that because Americans tend to favor sequential strategies, they should have more experience and thus be better at switching at tasks and thus have less switch-costs. Results were largely consistent with our hypotheses.

4.2   Experiment 1

4.2.1   Method

Participants.   Thirty-two Japanese students at Kyoto University and 26 European-American (American students of European descent) students at the University of Michigan participated as paid volunteers. Participants were closely matched on several ancillary factors (e.g., age, educational level, scholastic aptitude, and SES). They received payments for participation and earned monetary bonuses for good performance.

Apparatus.   Visual stimuli were presented on the display screen of a 17-inch Sony Trini-tron monitor connected to a Pentium personal computer. The monitor and computer were of the same builds in Japan and the US. Participants sat about 80 cm from the monitor in a quiet room. Responses were made with a piano-type response keyboard. It had two groups of three finger keys, with one group for each hand. The experiment was controlled by a program written in E-prime.

Experimental Design.   Each participant was tested during two sessions on successive days. A session included 24 trial blocks, divided into 8 consecutive "epochs" with 4 blocks per epoch. In each epoch, the first block contained auditory-vocal (AV) single-task trials, the second block contained visual-manual (VM) single-task trials, and the third and fourth blocks contained dual-task trials involving both tasks. There were 24 single-task or dual-task trials per block. During each block, all relevant stimulus-response (S-R) pairs occurred equally often in random order.

For the VM task, 4 S-R pairs were included. Their stimuli each had three dashes and a capital "O", forming either the stimulus O - - -, - O - -, - - O -, or - - - O. In response to these stimuli, respectively, the participant pressed one or another of 4 keys with either the right-hand ring, index, little, or middle finger (cf. Fitts & Seeger, 1953). For the AV task, 3 S-R pairs were included. Their stimuli were tones having frequencies of 196, 880, and 3520 Hz. The responses paired respectively with these stimuli were the spoken digits "1", "2", and "3".

During each dual-task trial block, all possible combinations of one S-R pair from the AV task and one S-R pair from the VM task occurred at least once in random order.

Procedure.   For each AV single-task trial, an initial 500 msec warning signal with 4 horizontal dashes appeared on the video display. Next, a 40 msec stimulus tone occurred. The participant responded by saying the digit that had to be paired with the stimulus, and RT was measured. An experimenter scored the response as being correct or incorrect. Then

the participant received 500 msec of visual feedback about response accuracy and bonus points earned on the trial.
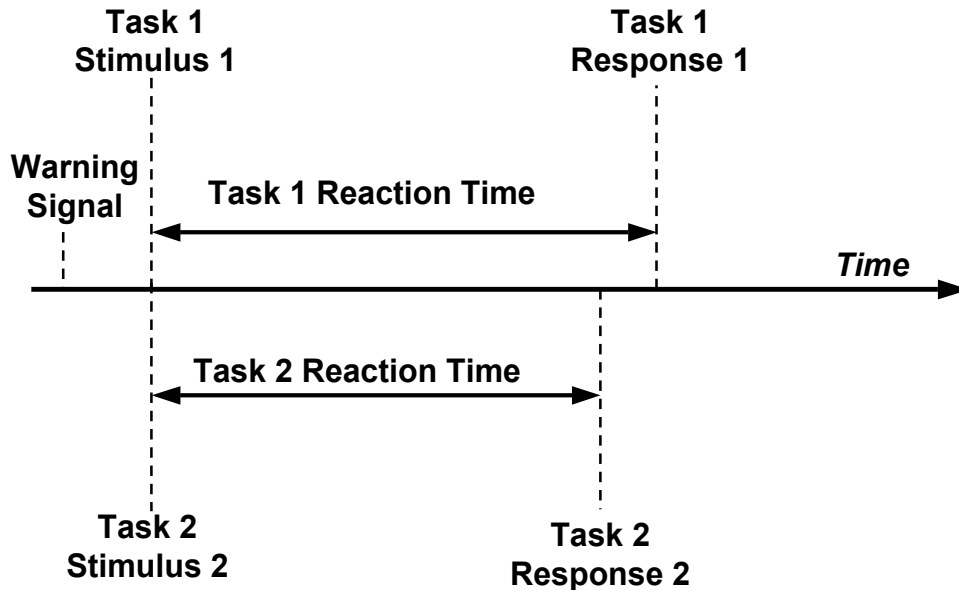

Figure 4.1. Timeline for a typical dual-task trial in a time-sharing study. Note that Task 1 and Task 2 are arbitrary designations, as both begin at the same time.

Each VM single-task trial started like those with the AV task. After the visual warning signal, one of its 4 dashes changed into a capital "O", presenting a 250 msec VM-task stimulus. The participant responded by pressing the finger key that had to be paired with the stimulus. Then there was 1000 msec of post-response visual feedback.

Each dual-task trial combined events from the AV and VM single-task trials. Stimuli for the two tasks occurred simultaneously after the initial warning signal. Participants responded to both stimuli as best possible. Following the responses, feedback was presented about the accuracy and points earned for both tasks.

In addition, participants received visual feedback after every trial block. It summarized their performance on the block, including total correct responses, mean RTs, number of

"too slow" responses, and points earned for each task. Participants were instructed to maximize their bonus points and monetary rewards.

4.2.2 Results

Outlier Removal. Outliers were removed based on bits transmitted per second. This is a measure that combines speed and accuracy into one statistic. We calculated bits transmitted per second for each participant in each task. Participants were removed based on the difference in bits transmitted per second in dual-task compared to how the participant performed in the single-task case. Participants who were more than 2.33 standard deviations (i.e. 1% of the time) away from the mean of their culture were eliminated this way. Participants were eliminated in this manner because this was the only way to interpret dual-task performance. Three Japanese and two American participants were eliminated this way and removed from future analyses.

Mean RTs and Error Rates appear in Table 4.1 for each trial type, task, and participant group.

Culture	Trial-Block Type	Task	Mean RT (msec)	Error Rate (%)
Japanese	Single-Task	AV	438	2.5
		VM	452	5.9
	Dual-Task	AV	583	5.9
		VM	562	5.7
American	Single-Task	AV	473	3.4
		VM	522	6.2
	Dual-Task	AV	696	7.2
		VM	688	7.2

Table 4.1. Comparison of mean RTs across cultures

Reaction Times. On average, RTs for the AV and VM tasks were approximately equal (M = 548 and 556 msec, respectively); F(1, 51) = 0.58, p > 0.4. Dual-task trials yielded reliably longer RTs (M = 632 msec) than did single-task trials (M = 471 msec); F(1, 51) =

103.2, p < 0.001. The dual-task time cost (DTTC; i.e., dual-task minus single-task RTs) was higher for the AV (M = 184 msec) than VM (M = 138) task; $F(1, 51) = 8.65$, p < 0.01. However, supporting our initial hypothesis, Japanese participants had reliably lower DTTCs (M = 125 msec) than did Americans (M = 195 msec); $F(1, 51) = 4.51$, p < 0.05. Lower Japanese DTTCs occurred for both tasks: the triple interaction involving task, trial type, and culture was unreliable; $F(1, 51) = 0.47$, p > 0.4.

This pattern prevailed throughout Session 2 (Figure 4.2). Mean RTs decreased reliably across trial-block epochs; $F(6, 306) = 26.6$, p < 0.001. So did DTTCs; $F(6, 306) = 5.06$, p < 0.001. Yet none of these trends interacted with the cultural factor (p > 0.5 in all cases).

We also examined the maximum and minimum cost on each trial, regardless of on which task. As predicted, Americans had larger maximum costs, M=270 msec, compared to Japanese, M=179 msec, $t(51) = 2.02$, p<.05. The minimum cost on each trial did not differ across cultures, Americans: 85 msec, Japanese: 61 msec, $t(51) = 1.17$, p=ns.

Error Rates.   ERs were generally low (M = 5.4%) and decreased with practice; $F(6, 306) = 2.13$, p = 0.05. Fewer errors occurred on single-task (M = 4.4%) than dual-task (M = 6.4%) trials; $F(1, 51) = 52.2$, p < 0.001. However, the practice effect was greater for dual-task trials; $F(6, 306) = 2.11$, p = 0.05. Americans committed slightly more errors (M = 5.8%) than did Japanese (M = 4.9%), but this difference was not reliable; $F(1, 51) = 0.37$, p = 0.5. Nor did the cultural factor interact with other effects on ERs. Thus, major speed-accuracy tradeoffs did not contribute to the observed pattern of mean RTs (Pachella, 1974).

Cross-Cultural Clustering of DTTCs.   To help further interpret the observed cross-cultural differences between DTCCs, we performed a partitioning-around-medoids (PAM; Kaufman & Rousseuw, 1990) clustering of individual participants' data. Figure 4.3 shows the results of the cluster. Initially we found that two clusters fit the data best, as can be seen. A chi-squared test was performed on whether the number of people from each cul-
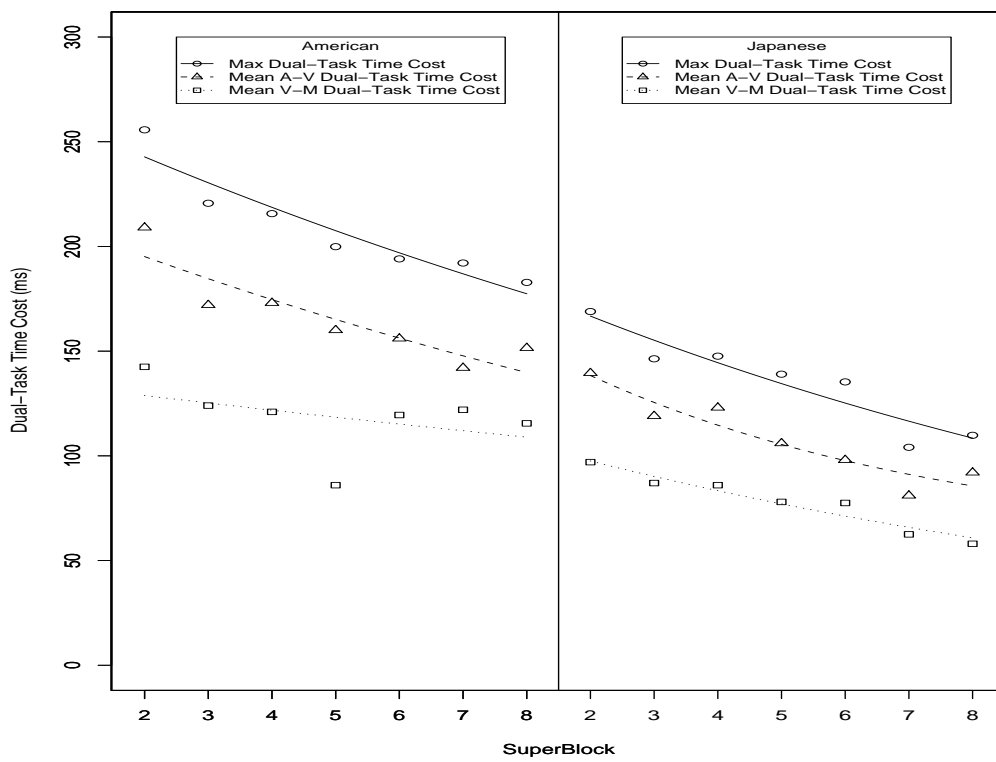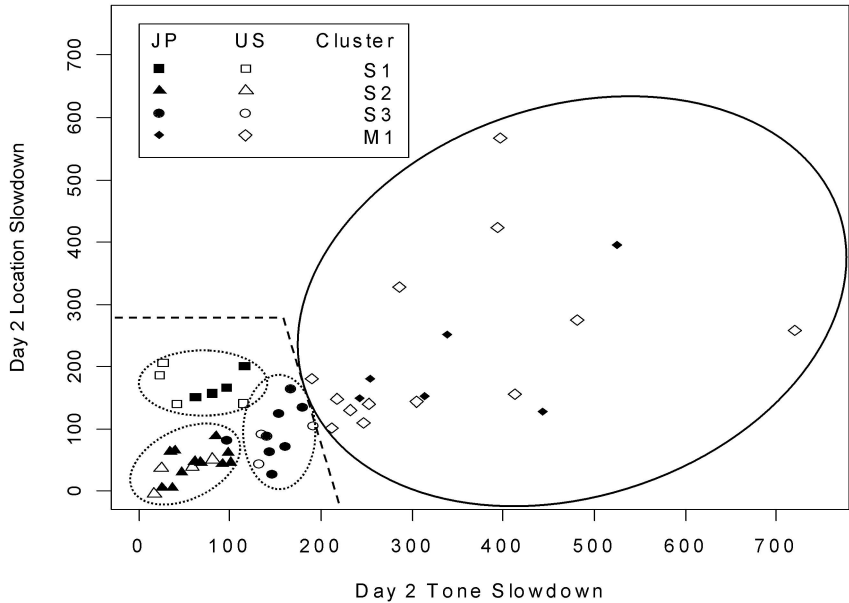
119

Figure 4.2. RTs by block

Figure 4.3. Cluster and sub-clusters of individual participants

ture significantly differed, $\chi^2(1) = 5.03, p < .05$. Participants in the large circle were basically those who had the most trouble with the task, so were not particularly theoretically interesting, but this significant difference indicates there was more variability in the American population than in the Japanese population. The second cluster (those participants within the dotted lines) were those participants who were capable at the task and so had reason to strategize. We re-clustered based just on those participants. Each of these sub-clusters is circled in Figure 4.3. There was no longer a cultural difference in these sub-clusters $\chi^2(2) = 1.49, p = ns$. However, note that the three sub-clusters do seem to exhibit strategic differences. One cluster seems to have low costs in both tasks, consistent with performing both tasks in parallel. The other two clusters have low costs in one task, but larger costs in the other task, consistent with doing the tasks sequentially.

Points. We also examined how many points were earned by participants from each culture. Japanese earned on averaged 61,993 points, marginally more than the American average of 53,846 points, t(51) = 1.86, p=.069. However, when we only examine results from the sub-clusters, we find the American mean of 63,800 points did not differ from the Japanese mean of 66,265 points (t(32) = 0.46, p=ns).

4.2.3 Discussion

Our results provide new insights about cross-cultural differences in cognition and performance. Previously, it has been established (e.g. Kitayama et al., 2003) that while making unspeeded perceptual judgments, members of collectivist societies distribute their attention more broadly than do members of individualistic societies. We have found that this propensity for distributed "attention to perception" is complemented by distributed "attention to action" during speeded multitasking. The lower DTTCs of Japanese participants suggest that they more strongly prefer and/or can better implement task-scheduling strategies whereby response selection and movement production occur simultaneously for multiple tasks. Such parallelism further manifests the diversity of human information processing and the flexible use of people's cognitive architectures (cf. Meyer & Kieras, 1997a,b, 1999). That this diversity and flexibility arise even in performing basic choice-reaction tasks illustrates the power of cultural factors to extend well beyond routine social situations.

4.3 Experiment 2

In Experiment 2 we generalized the procedure of Experiment 1, using another type of choice-RT task: color identification. We chose a color-identification task because there was no reason to think color identification should vary across cultures, whereas a spatial task could use different strategies even at a perceptual stage. For example, a person could choose to perceive the 0 and three dashes as a spatial task depending on the location of the

0, or they could group the 0 and three dashes into a single stimulus object, and thus press a key depending on which of four stimuli they perceive.

4.3.1 Method

Participants.   Twenty-four Japanese and 23 European-American university students participated as paid volunteers. None had been in Experiment 1. The same payment scheme was used again. One Japanese participant had a computer malfunction during Session 2, so her data was never analyzed.

Apparatus, Design and Procedure.   The apparatus, design, and procedure were as before except that, for the VM task, each test stimulus was a disc colored green, yellow, blue, or red. In response, participants pressed either a right-hand index, middle, ring, or little finger key, respectively.

4.3.2 Results

Outlier Removal.   Outliers were removed based on bits transmitted per second, as in Experiment 1. One American and one Japanese participant were removed from further analyses.

Mean RTs and Error Rates appear in Table 4.2 for each trial type, task, and participant group.

Reaction Times.   On average, RTs for the AV and VM tasks were approximately equal (M = 496 and 492 msec, respectively); $F(1, 41) = 0.07$, $p > 0.5$. Dual-task trials yielded reliably longer RTs (M = 566 msec) than did single-task trials (M = 422 msec); $F(1, 41) = 165.0$, $p < 0.001$. The DTTC was slightly but not reliably higher for the VM task (M = 150 msec) than AV task (M = 138); $F(1, 41) = 0.36$, $p > 0.5$. Again, supporting our initial hypothesis, Japanese participants had reliably lower DTTCs (M = 122 msec) than

Culture	Trial-Block Type	Task	Mean RT (msec)	Error Rate (%)
Japanese	Single-Task	AV	399	2.3
		VM	399	7.4
	Dual-Task	AV	524	4.4
		VM	517	7.0
American	Single-Task	AV	455	3.0
		VM	435	6.6
	Dual-Task	AV	605	7.5
		VM	617	5.1

Table 4.2. Comparison of mean RTs across cultures

did Americans (M = 166 msec); F(1, 41) = 3.91, t(41) = 1.98, p < 0.05, one-tailed. Lower Japanese DTTCs occurred for both tasks: the triple interaction involving task, trial type, and culture was unreliable; F(1, 41) = 0.83, p > 0.3. As in Experiment 1 (Figure 4.4), the pattern of Experiment 2's results prevailed during all of Session 2. Mean RTs decreased reliably across trial-block epochs; F(6, 246) = 16.0, p < 0.001. So did DTTCs; F(6, 246) = 5.85, p < 0.001. Yet none of these trends interacted with the cultural group factor (p > 0.2 in all cases).

We also examined the maximum cost on a given trial, regardless of which task. Americans, as expected, had significantly larger average maximum cost, M=240 msec, compared to the Japanese cost of 170 msec, t(41) = 2.11, p<.05. As a sanity check, we also examined average minimum cost, regardless of task, which we did not expect to differ. Americans averaged 71 msec, Japanese 68 msec (t(41) = 0.14, p>.5).

Error Rates. ERs were generally low (M = 5.4%). Fewer errors occurred on single-task trials (M = 4.8%) than dual-task trials (M = 6.0%); F(1, 41) = 72.1, p < 0.001. Americans committed slightly more errors (M = 5.5%) than did Japanese (M = 5.3%), but this difference was not reliable; F(1,41) = 0.58, p > 0.4. Nor did the cultural factor interact with any other effects on ERs (p > 0.4 in all cases).
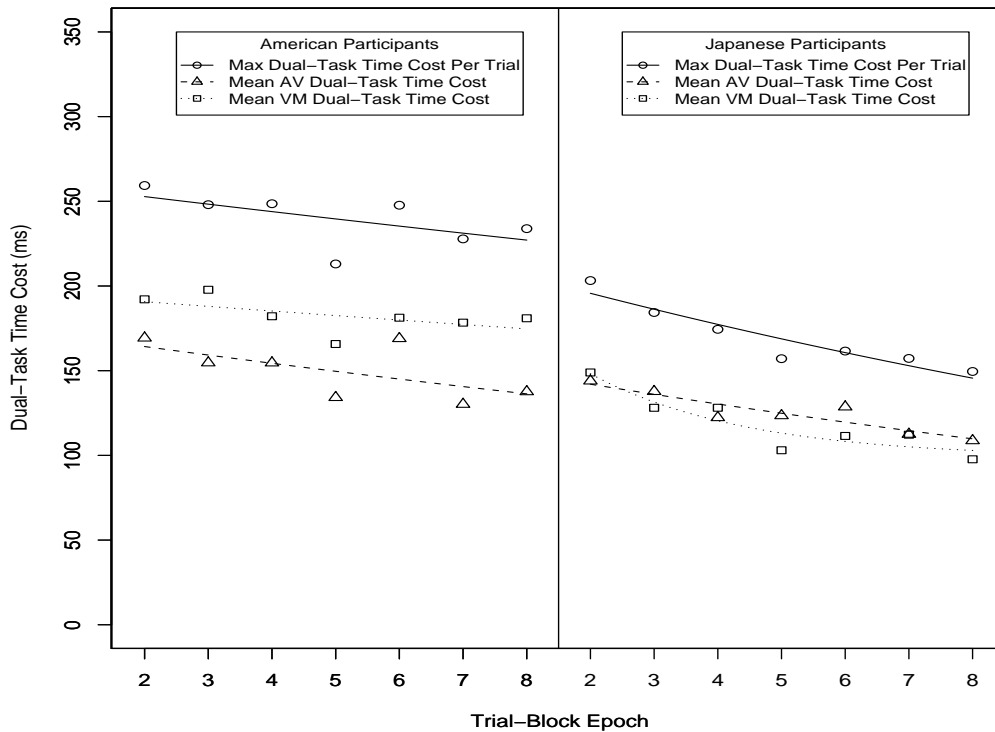
Figure 4.4. RTs by block

Cross-Cultural Clustering of DTTCs. As in Experiment 1, we clustered based on dual-task trial costs. The clustering revealed that except for one aberrant case, each participant belonged to one or another of three principal clusters in a space defined by individual pairs of mean DTTCs. Cluster 1 contained a preponderance of participants whose mean DTTCs were relatively low (M ≤ 80 msec) for both the AV and VM tasks; they apparently tended to perform the two tasks simultaneously without extreme dual-task interference. In contrast, Cluster 2 contained participants whose mean DTTCs were considerably higher for the VM task (M = 328 msec) than AV task (M = 133 msec), tending more toward sequential performance with the AV task having first priority. Complementarily, Cluster 3 contained participants whose mean DTTCs were considerably higher for the AV task (M = 225 msec) than VM task (M = 131 msec), reversing the priorities for sequential performance manifested by Cluster 2.
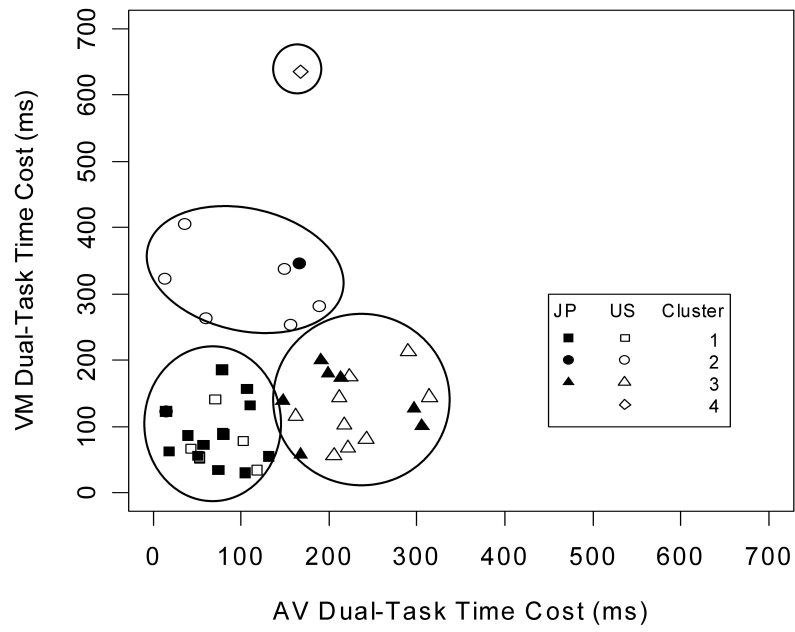
Figure 4.5. Cluster of individual participants

The distribution of participants across these three clusters depended systematically on culture. A large majority (68%) of the participants in Cluster 1 were Japanese, whereas a large majority (86%) of the participants in Cluster 2 were Americans, and there was also a majority (56%) of Americans in Cluster 3. This overall cultural contingency was reliable; $\chi^2(2) = 6.42$, $p < 0.05$, supporting our a priori hypothesis that members of collectivist societies differ from those of individualistic societies in their proneness to simultaneous rather than sequential task scheduling.

However, it must be stressed that preferences for particular scheduling strategies and the ability to implement them are not uniform within either Japanese or North American cultures. Reinforcing this proviso, Figure 4.6 shows scatterplots of paired DTTCs from individual dual-task trials for various participants (Panels A-D) who belonged to Clusters 1, 2, and 3, respectively. The participant in Panel A, a Japanese member of Cluster 1, had DTTC pairs clumped relatively close to the zero origin, manifesting essentially simultaneous multitasking. A few Americans had DTTCs from individual dual-task trials that looked like this Japanese participant's DTTC. Unlike him, though, the participant in Panel B, an American member of Cluster 2, had DTTC pairs mostly scattered near the vertical axis as if the AV task was usually performed before the VM task. Also, the participant in Panel C, an American member of Cluster 3, had some DTTC pairs scattered near the vertical axis, but even more of his DTTC pairs were scattered near the horizontal axis, indicating that he tended more often to perform the VM task before the AV task rather than vice versa. Complementing this pattern, the participant in Panel D, a Japanese member of Cluster 3, had DTTC pairs scattered mostly parallel to but slightly below the positive diagonal, manifesting another "mixed" sequential strategy that favored the VM task.

Points. US participants earned on average 50,323 points, which did not differ from the Japanese average of 50,505 points, t(41) = 0.06, p=ns.
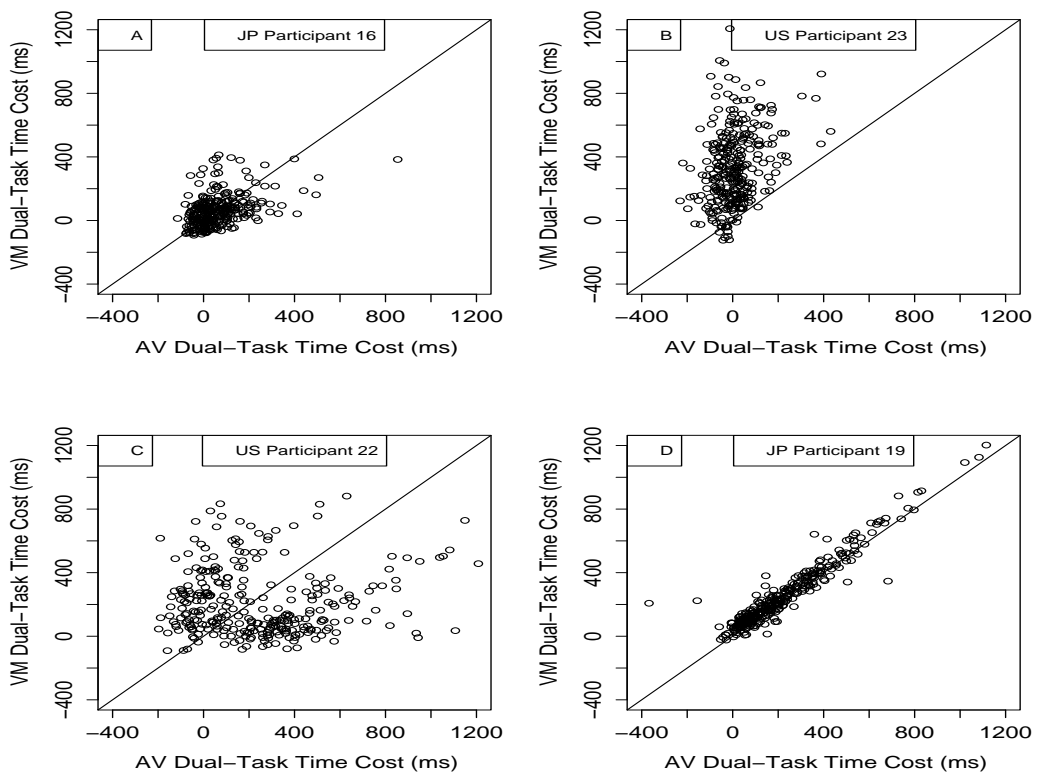
Figure 4.6. Scatterplots of dual-task trial costs for individual dual-task trials

4.3.3 Discussion

Results in Experiment 2 were consistent with those in Experiment 1: Japanese participants had less dual-task cost than Americans, but without any corresponding gain in points earned. The individual-subject analyses were even more clear in this experiment, as there were fewer participants who struggled so mightily with the simplified color task, as again we found the three clusters we found in our initial sub-cluster in Experiment 1, and we found that participants from each culture differed in what clusters they tended to be in.

The fact that the clusters were consistent with our notion that Japanese would tend to choose a parallel strategy and Americans tended to choose sequential strategy certainly helps our hypotheses. The trial-by-trial analyses of all trials was even more compelling, showing clear structure consistent with sequential or parallel strategy.

Experiments 1 and 2 were thus consistent with our hypotheses that culture can affect strategy.

4.4 Experiment 3

Experiments 1 and 2 looked at performing two tasks simultaneously, i.e. time-sharing. However, another form of multi-tasking we can look at is task-switching. Initially, it seems that because you only do one task at a time, a sequential strategy would be optimal, favoring Americans. So we predicted that if you compare Americans who switch between tasks with Japanese who switch between tasks, Americans should have less switch costs.

4.4.1 Method

Participants. Twenty-four Japanese students at Kyoto University and 24 European-American students at the University of Michigan participated as paid volunteers. They received payments for participation and earned monetary bonuses for good performance.

Apparatus and Design. The apparatus and experimental design were the same as before. The stimuli were the same as for the Visual-Manual task in Experiment 1. Although the stimuli were the same regardless of task, participants were primed as to which task to perform by the color of the fixation. The fixation dashes were either green or red, indicating which set of stimulus-response pairs participants needed to use. The blocks had a Pure Green block, where every trial had a green fixation and required the green S-R responses and a corresponding Pure Red block. On a Dual-Block, there were twenty-four trials with 12 red trials and 12 green trials, randomly interspersed.

Procedure. For each trial, 4 green or red horizontal dashes appeared on the video display for 500 msec. Then one of the four stimuli from Experiment 1 appeared on screen for up to 2000 msec or until the participant made a response. The participant then received 500 msec of visual feedback about response accuracy on the trial.

In addition, participants received visual feedback after every trial block. It summarized their performance on the block, including total correct responses, mean RTs for each task, number of "too slow" responses for each task, and points earned for each task. Participants were instructed to maximize their bonus points and monetary rewards.

4.4.2 Results

Outliers were removed based on bits transmitted per second, as in Experiments 1 and 2. One US and one Japanese participant were removed from further analyses.

All results taken from Session 2 only, excluding the first block to get participants re-oriented to the experiment. Because we found no difference in performance for the green and red tasks, all results were collapsed across these tasks.

Mean RTs and Error Rates appear in Table 4.3 for each trial type, task, and participant group.

Culture	Trial-Block Type	Trial Type	Mean RT (msec)	Error Rate (%)
Japanese	Single-Task	Pure-Trials	527	6.5
	Dual-Task	No-Switch Trials	551	8.5
		Switch Trials	591	11.8
American	Single-Task	Pure-Trials	571	7.0
	Dual-Task	No-Switch Trials	617	10.0
		Switch Trials	691	13.3

Table 4.3. Comparison of mean RTs across cultures

Reaction Times. On average, RTs for the no-switch trials on Mixed Blocks were not significantly greater than RTs for the Pure Blocks (M=584 and 549, respectively, t(94) 1.63, ns). RTs for the switch trials on Mixed Blocks (M = 641) were significantly larger than for trials on the Pure Blocks, t(94) = 3.68, p < .001. Reaction times for switch trials were significantly larger than for no-switch trials, t(94) = 2.11, p < .05.

Japanese and American participants did not differ in Pure RT (M = 527 and M = 571, respectively, t(46) = 1.61, ns). The US Mixing Cost of 46 msec was significantly larger than the Japanese Mixing Cost of 24 msec (t(46) = 2.04, p < .05, and the US Switch Cost of 74 msec was significantly larger than the Japanese Switch Cost of 39 msec (t(46) = 2.56, p < .05.

ANCOVA results were also used to control for Pure RT differences. When we control for Pure RT, then Switch Cost is still significant, t(45) = 2.12, p<.05, and Pure RT does not significantly differ across cultures in this model. If we control Pure RT when looking at the Mixing Cost, then culture is no longer significant t(45) = 1.46, p=ns, but Pure RT becomes highly significant, t(45) = 2.94, p<.01.

Error Rates. ERs were larger than in Experiments 1 and 2 (M = 9.53%). Fewer errors occurred on pure-task trials (M = 6.76%) than no-switch trials (M = 9.24%, t(94) = 2.18, p<.05) and switch-trials (M = 12.60%, t(94) = 3.88, p<.001). There were fewer errors on

no-switch trials than switch trials, as well (t(94) = 2.18, p<.05). Americans did commit slightly more errors (M = 10.1%) than did Japanese (M = 8.9%), but the difference was not reliable, t(46) < 1, p > .5. Culture did not interact with any other effect on ERs, (p > 0.4 in all cases).

Cross-Cultural Clustering.   As in Experiment 2, we clustered using PAM based on dual-task costs, here Mixing Costs and Switch Costs. We initially found that two clusters were optimal, with an average silhouette width of .482. We then examined one of the sub-clusters. We found, as in Experiment 2, that 3 sub-clusters were optimal, with an average silhouette width of 0.47. Figure 4.7 shows a plot of individual subjects in the sub-cluster. Although it is harder to interpret these clusters compared to those in Experiment, of most concern to us is Cluster 1, the cluster of individuals who have no Mixing Cost and no Switch Cost. In this cluster, 75% are Japanese. The distribution of individuals in these three clusters is marginally significant, $\chi^2(2)$ = 5.86, p = 0.053.

This sub-cluster of parallel participants gives the initial impression that Japanese have less switch-cost than Americans. However, if we combine the other two sub-clusters (which we assume have some form of sequential strategy, or at least a non-parellel strategy), these 11 American participants have significantly less switch costs (M = 37.3 msec) than these 9 Japanese participants (M = 53.8 msec), t(18) = 2.29, p< .05. Mixing costs for the two non-parallel sub-clusters did not differ (t(18) = 1.19, p = ns).

If instead we just remove the parallel cluster, but look at all remaining participants (not just those in the sub-clusters above, but including those in the original cluster), we do not find a significant difference in switch or mixing costs. Americans do have slightly larger switch costs, M = 75.5, compared to the Japanese participants, M = 66.4, t(30) = 0.59, p>0.5. Mixing costs for Americans (M=41.5) also did not differ compared to those of the Japanese (M=35.6), t(30)=0.46, p>0.5.
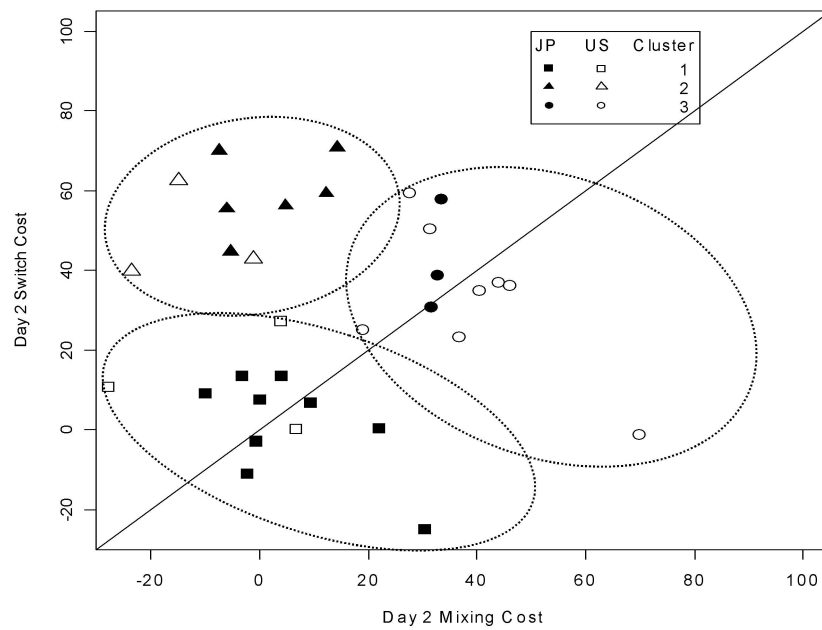
Figure 4.7. Subcluster of individual subjects

Points. Japanese participants averaged 49,948 points, slightly more than the American average of 47,912 points, but this difference was not significant, t(48) = 0.70, p=ns. However, of more interest is the average points in the sub-clusters. In the parallel sub-cluster, Japanese averaged 47,733, basically the same as the American average of 47,400 points (t(10)=0.04, p=ns). In the two sequential sub-clusters, Americans averaged slightly more points (M=51,800) than did the Japanese (M=48,689), although this difference was not significant (t(18) = 0.96, p=ns).

4.4.3 Discussion

Our initial results that Japanese had less switch costs than Americans surprised us. But closer analysis of our data supported our hypothesis that Americans favor sequential strategies and have less switch costs than Japanese. When we compared Japanese who clearly switched between tasks with Americans who clearly switched, we found Americans had less switch costs. We also found more Americans than Japanese who engaged in a clear switching strategy, consistent with our hypotheses.

The reason our overall results were not as we initially expected is that numerous participants adopted a parallel strategy that we did not anticipate, which required no task-switching. These participants adopted a parallel strategy whereby they held two rule sets in parallel, so they had absolutely no interference, neither switching nor mixing-cost. It was not surprising that these participants with no switching or mixing-cost tended to be Japanese participants. The fact is this parallel strategy is clearly more optimal than a sequential strategy, as it results in no costs. So the fact that not all participants chose this strategy seems to contradict our findings from Chapter 2. However, the problem may be our training and instructions encouraged participants to switch between tasks. Our English instructions actually used the word "switch" in it, and thus word alone may have prevented Western participants from considering parallel strategies. Our Japanese instructions used the Japanese translation of "switch," but the word may not have the same connotation in a

culture where people favor parallel strategies and so are more likely to search the strategy space to find a parallel strategy.

Our evidence would have been more compelling if we found that Americans had less switch costs than Japanese across all other participants except those who engaged in the parallel strategy. We did not, largely because there seemed to be more variance in the American population than in the Japanese population. The American population had three participants who were potential outliers, but because there were three of them, there was no good statistical test that would remove all three of them. The Japanese population, in contrast, had one potential outlier (although outlier tests came up negative). Americans, for example, had six participants with switch costs greater than 100 msec, Japanese had one. The fact is, in both cultures, it is hard to interpret these participants. Are they trying, or are they content with the base pay? If we are arguing that one culture is better on average than sequential tasks, then should we care about the tail about the distribution that skews the mean? Examining the sub-cluster constrains the variance so that it is roughly equal across the two cultures, and these are participants who are trying hard in the task. The fact we had roughly equal number of participants from each culture in these sub-clusters is yet another reason why we should take these results seriously.

4.5 General Discussion

The principal objective of the present research has been to explore cross-cultural differences in multitasking, thereby addressing fundamental questions about the diversity of human information processing, heterogeneity of cognitive architectures, and adaptability of task-scheduling strategies. Taken overall, our results suggest that members of a collectivist society (i.e., Japanese culture) lean more toward simultaneous multitasking than do members of an individualistic society (i.e., American culture), who instead tend to favor sequential multitasking. This finding, which manifests distributed versus selective ặattention

135

to actionąś, extends prior discoveries about how cultural practices may influence attention to perception and cognition (Kitayama et al., 2003; Nisbett et al., 2001).

Our three experiments all showed that Japanese participants had less costs associated with simultaneously performing two tasks than did American participants. In Experiment 3, although only one task was performed at a time, participants could still show proclivity in performing two tasks simultaneously if they could hold two different rule sets in mind without interference. Comparing reaction times results for each culture in Experiments 1 and 2 showed that Japanese participants had less costs and thus less interference in time-sharing than did American participants. Clustering individual participants indicated that in both experiments people seemed to adopt one of four strategies: 1) A strategy that required little cost for both tasks, a parallel strategy 2) A strategy that required little cost for the tone task, but greater cost for the visual-manual task, a sequential strategy 3) A strategy that required little cost for the visual-manual task, but greater cost for the tone task, a sequential strategy, or 4) a strategy that required relatively large cost for the visual-manual and tone task. The first strategy was a parallel strategy that allowed both tasks to be processed simultaneously as much as possible. In both experiments, we found more Japanese engaged in this type of strategy, consistent with the notion of using a more holistic attentional strategy. The other two strategies are more sequential, indicating you first process one task and then do the other task. Not surprisingly if Americans allocate attention more analytically, Americans tended to disproportionately engage in these two strategies. We found a similar pattern in Experiment 3, as well, with Japanese participants more likely to engage in a strategy that resulted in neither switching or mixing-costs, hence indicating being able to hold both rule sets in mind with no interference.

It may be argued that Japanese advantages in simultaneous multitasking reflects a genetic difference from Americans (for example, task-switching has already been argued to be largely genetic, see Friedman et al., 2008). Although possible that Americans are more likely to have a cognitive response-selection "bottleneck" that constrains Americans to use

136

sequential multitasking strategies, (cf. Meyer & Kieras, 1997a,b), we find this unlikely. For one thing, in all our experiments we found that some Americans did indeed perform the two tasks simultaneously as well as any Japanese participant, and we found some Japanese participants performed the tasks sequentially. Another reason is that training can improve performance on simultaneous tasks and other forms of holistic attention (Gopher, 1993; Green & Bavelier, 2006; Polk & Farah, 1998). Further, the predilection for simultaneous multitasking does not seem to be unique to East Asians, but to other collectivist cultures as well (Chavay & Rogoff, 1999). Although to really ensure these differences are not genetic in nature, future research should examine Asian-Americans, especially 3rd or 4th generation Asian-Americans who should be more heavily influenced by American culture than Japanese culture and thus favor sequential strategies.

A possible criticism is that Japanese participants put forth more effort than did American participants. We included points earned in the results analyses to help refute this line of criticism. If Japanese participants were really trying harder, than they should have earned more points since that is what participants were striving to maximize. However, this clearly wasn't the case in Experiment 2 and 3. In Experiment 1, Japanese did earn a bit more points, but this effect disappeared when we examined participants in the sub-cluster. More importantly, the fact that the results in Experiment 1 were replicated in Experiment 2, which did not have a difference in points earned, indicates that motivation differences did not account for the behavioral differences. Further, the fact that the minimum cost on any given trial did not differ across cultures, as we predicted, is further evidence that participants were trying hard in both cultures.

In fact, Experiment 3 offers further refutation of the motivation argument. We predicted a priori that Americans would be better at task-switching than Japanese, due to having more extensive practice in sequential strategies, and so should have less switch-costs than Japanese. Although overall we did not find this to be the case, close examination of the data revealed that was because so many Japanese discovered a parallel strategy we did not

consider. Our hypothesis did not factor in those participants, so instead we examined those participants who engaged in a sequential strategy. There we found Americans had lower switch costs and earned more points than Japanese participants who engaged in the same strategy. Although the points did not significantly differ, the fact that Americans had a trend to earn more certainly refutes any hypothesis that effort differed across cultures.

More research is needed to establish how general our results are. Higher-level multitasking, such as driving and using a cell phone, should be investigated, along with lower-level multitasking as in the Psychological-Refractory Period (PRP) procedure. Further, more tasks should be investigated and designed to favor sequential strategies over simultaneous strategies, to both clarify Western participants' advantages in multitasking, and to help examine the extent to which training can affect how people choose their multitasking strategy. It would be interesting to design experiments such that simultaneous or sequential strategies were clearly dominant, and then see if individuals regardless of culture attempted the corresponding proper strategy. Also, as mentioned above, it would be interesting to examine Asian-Americans and Westerners who have lived in Japan over a period of many years, to help refute possible genetic accounts.

Chapter 5

Cultural Differences in Optimal Adaptations in Ordered Responses

In Chapter 2, we examined how individuals adapt in a simple-RT task, and we found that participants generally adapted in a mathematically-optimal way to maximize their total points. However, in Chapter 4, we discovered that culture influenced people's strategic choices. Although the tasks in Chapter 4 involved relatively complicated multitasking, and we had a priori reasons to assume that multitasking strategies may differ across cultures, it could be that culture plays a role in cognitive experiments in general.

Hence the data from Chapter 2 provide us with a control for this hypothesis. When we ran the experiments on optimal adaptations in ordered responses, we did not consider culture, but we serendipitously included participants from different cultures. After finding out there were cultural differences in multitasking, we decided to reexamine our results from Experiments 2 and 3 of Chapter 2 in terms of cultural differences. We had no reason to assume there would be cultural differences, and in fact we expected that people from all cultures should adapt equally well.

Despite the fact that we expect a null result, a significant difference or even trend across cultures would be an interesting finding. Depending on the nature of the findings, it may

affect our interpretations of Chapter 4. It would also show other experimenters that taking culture into account is important for analyzing results from RT experiments. If we do not find any differences across cultures, this would indicate that we did a good job of constraining the strategy space so that there was one clear strategy regardless of cultural background.

Unfortunately, our original experimental design only allowed us to do a limited set of statistical tests for cultural differences. In Chapter 4, we looked at only two cultures: European-American and East Asian. Here, though, we had to further subdivide the participants since we also had African-American and South Asian participants. South Asians are not from a collectivist culture, so we did not want to aggregate them with East Asians. Likewise, we did not want to combine African-American participants with European-American participants, as we did not test African-Americans in our previous multitasking experiments.

We found no evidence for cultural differences in these experiments.

5.0.1 Method

The method was the same as in Experiments 2 and 3 from Chapter 2.

5.0.2 Results

We combined the data from Experiments 2 and 3 in order to ensure that we had enough participants from each culture. We ended up with 11 East Asian participants, 11 African-American participants, 18 European-Americans, 6 South Asians, and 1 Hispanic (who was removed from subsequent analyses due to having too small a sample size). Note that the participants from each culture were not separated equally across payoffs and external-noise levels, so this results in potential confounds we have to be wary of when examining our results.

We first examined how far from optimal each culture was on average. Americans averaged 12.2 msec too short for optimal, East Asians averaged 4.5 msec too short, African Americans averaged averaged 1.7 msec too long, and South Asians averaged 0.4 msec too long. The difference between Americans and East Asians was the critical difference, and this was not significant, $t(27) = 0.63$, $p > 0.5$. Because Americans and East Asians may not have been represented equally across our different groups, any cultural difference would be confounded with the level of total noise, which we found to have the biggest effect on how far away participants were from optimal, we ran an ANCOVA where we controlled for total noise. We again found that after controlling for total noise, culture was still not significant, $t(26) = 0.58$, $p > 0.5$. If we included African-American and South Asian cultures in our analysis, our results did not change.

The main reason that Americans seemed to differ from East Asians is because Americans had 12 participants in the accuracy payoff (where participants tended to be too fast) and only six participants in the speed payoff (where, if different from optimal, participants tended to be too slow). East Asians did not have this skew, with five participants in accuracy and six in the speed payoff. Hence this disproportion probably accounts for why the cultures have a slight (albeit nonsignificant) difference in their deviation from optimal IRI.

As in Experiment 2, we examined difference from optimal IRI as a function of total noise, but this time took into account culture (see Figure 5.1). From this figure, we can see that there was a good spread of cultures in the different conditions and payoffs, although we ended up with four East Asians in the high-noise condition, but only two European-Americans. In all levels of noise, though, there are both East Asians and European-Americans, and they seem to be roughly equivalent in terms of distance from Optimal IRI.
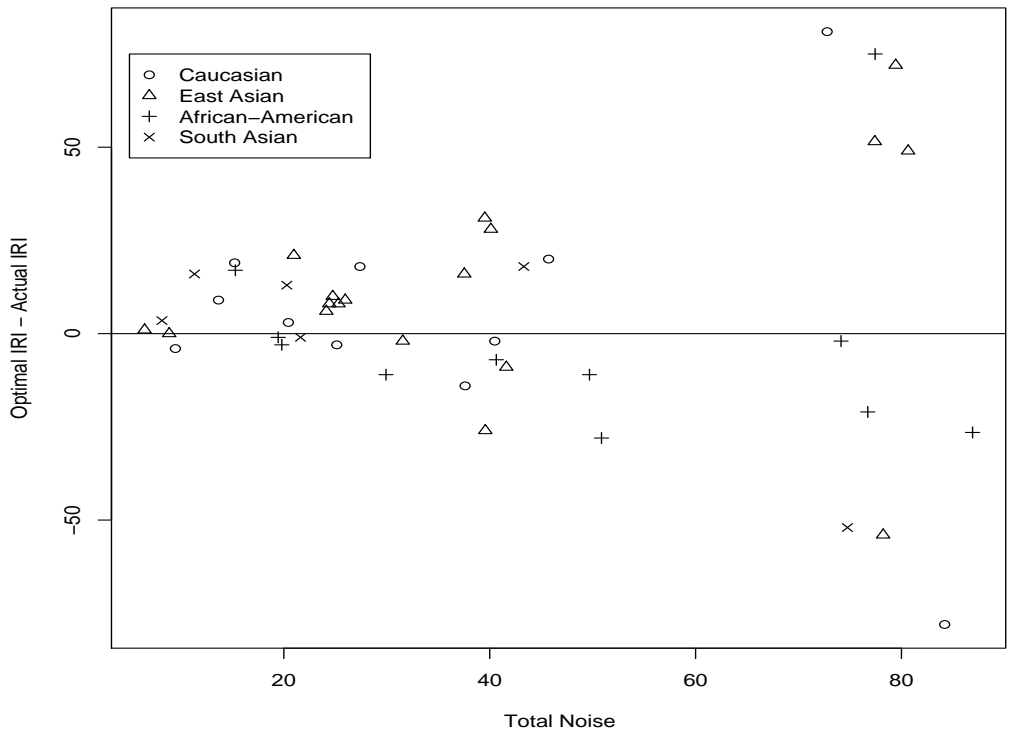
Figure 5.1. Individuals differences from optimal IRI, by culture

5.0.3 Discussion

Neither median RT nor the individual scatterplot showed a trend for cultural differences in making optimal adaptations for ordered responses. Nor was there any trend once we took into account total-noise level and payoff.

Hence this reanalysis of Experiment 2 showed that in many cases, culture may not affect results. This is not really surprising; we are all human after all. Nevertheless, it shows the importance of considering in what types of tasks culture matters and must be carefully controlled, and in what types of tasks culture does not matter.

If the experimenters manage to carefully control the strategy space, then culture does not matter, as all participants will opt for the same strategy. If, however, the strategy space is broader so there are multiple optimal strategies from which to choose, culture and other higher-level constructs like personality may matter. In multitasking, we know culture indeed does matter. Independent cultures favor more sequential strategies, while interdependent cultures favor more simultaneous strategies. In simple-RT experiments, though, both cultures favor the same strategy, the one that maximizes their payoff.

Chapter 6

Conclusion

The importance of strategy recurs throughout cognitive psychology. Even if the research question at first glance seems to have nothing to do with strategy, the strategy people opt for influences what conclusions an experimenter can and will deduce. It is thus critical to study strategy not just for its own sake, but to allow other researchers to control for strategy when interpreting their own experiments. For example, neuroimaging techniques such as fMRI and ERP are often used to try to answer some of the fundamental questions about cognition and about its limitations. However, a confound exists because those neuroimaging techniques will measure brain activity not only associated with the effect of interest, but also brain activity associated with choosing and implementing a strategy. If the data is analyzed at the individual level, it is possible that different strategies could be ascertained (e.g. Schumacher et al., 2001), and then possibly strategies could be controlled for so that brain regions associated with the effect of interest may be differentiable from the given strategic decision. However, this typically results in too few data points per strategy to yield meaningful results, which is why this is not typically done in neuroimaging studies.

Over the course of this dissertation, we first determined what guides a person's strategy, namely the maximization of points they earn (Chapter 2). We made use of this in two ways in Chapter 3. On the one hand we spent considerable effort in designing a payoff that

encouraged participants to prioritize Task 2 while still trying to quickly perform Task 1, because we knew that participants would adapt to the payoff as well as they could. On the other hand, we also chose which strategies in our computational models were optimal by comparing payoffs across those strategies. Again, we felt confident that participants were trying to maximize their points, so we felt that comparing points was the most fair method to compare different strategies. In Chapter 3 we expanded the application of the Strategic Response Deferment (SRD) model (Meyer & Kieras, 1997a) by finding that it was a good fit not just for Psychological Refractory Period tasks, but also for novel task-interruption tasks. Although our results were inconclusive about whether or not there was a Response-Selection Bottleneck, our best fit was via a strategy-based model.

In Chapter 4, we attempted to discern some of the factors that fostered some of the individual differences we find in selecting strategies, and we found strong evidence that culture heavily influences what strategy individuals select. To interpret our results, we made use of the fact that participants chose the strategy that maximized the points they earned (Chapter 2) and that people experimented with many different strategies in multitasking studies (Chapter 3). In Chapter 5, we found that when a reaction time task was designed so that optimal strategy was the same for every participant, cultural differences in strategy selection disappeared and behavioral results no longer differed.

Integration of these chapters suggests many implications for understanding how people select their strategies for reaction-time tasks. First, we demonstrated how important it is to give people clear goals, e.g. to maximize the number of points they can earn. Then we showed ways of analyzing strategies in a complex task to help disentangle the complicated interactions of strategies and structural limitations, in order to understand the true limitations of the mind. We also found further evidence that multitasking results are heavily influenced by strategic decisions on where to set up locking points to prevent response-reversal (or order) errors. Finally, we showed that in complicated tasks where different strategies can be optimal, culture influences the strategies people choose.

145

It should be clear that to help discern speed-accuracy tradeoffs, we need to consider all of these facets. Payoff matters not just in psychology but in real-world tasks as well. There are likely more crashes per mile in autoracing than in routine driving, despite the fact that race-car drivers are highly attuned to their driving, are very skilled, and have great reflexes, all conditions that are missing in many real-world drivers. Yet, obviously, race-car drivers emphasize speed over accuracy, which explains why they are in a disproportionate number of accidents. The first step in understanding speed-accuracy tradeoffs is to make sure that we understand the real implications of the payoff that is guiding the participant. Without seriously considering the payoff, we cannot be sure if the results we obtain are caused by poor performance, cognitive constraints on that task, or strategic considerations.

After we have considered the implications of the payoffs, we have to consider how to analyze the strategies used in relatively complicated tasks. For both low and high-level tasks, this can involve breaking the task down into stages (Sternberg, 1969a,b). In higher-level tasks, this may require a flowchart. We then have to manipulate the stages to really understand what affects performance. For example, in Chapter 3, changing the stimulus-encoding time of our Task 1 stimulus did not affect performance in Task 2, but changing the response-selection difficulty of Task 1 did. Understanding how to parse tasks allows us to better understand what strategies are being chosen and what factors cannot be strategically varied.

Once we have thought about how to analyze strategies, we have to be aware of the vast number of unconscious influences that affect what strategy people choose for common tasks. Culture is rarely salient to most of us, except when we leave our own culture. As a result, often neither experimenters nor participants consider how much culture affects people. Yet growing evidence finds that culture influences everything from perceptual tasks (e.g., Chua et al., 2005; Miyamoto et al., 2006) to high-level decision making (Yates et al., 2002). It is therefore not surprising that culture influences strategic choices on basic cognitive tasks; culture needs to be considered by cognitive psychologists when they attempt

146

to analyze and design their own experiments.

There are, of course, many factors besides culture that affect what strategies an individual chooses. Personality has been shown to have an effect on how a person selects a choice along the speed-accuracy tradeoff (Dickman & Meyer, 1988). It seems likely that gender should affect strategies in certain tasks, although we found no evidence of gender effects in our experiments. Videogame experience has been shown to affect visuospatial attention (Green & Bavelier, 2006), so video-game experience is likely to affect what strategies would be optimally chosen on many tasks. It may be that being from an urban, suburban or rural setting may also affect what strategies you choose on many given tasks. Ultimately, researchers have to be aware of these factors, if only to control for different strategies so that we can progress toward a meaningful, unified theory of cognition.

Our findings also have broad real-world applications. As discussed earlier, a warning system for people in certain jobs where both tasks are critical may cause a sharp detriment in performance on the initial task and make the person slow to respond. To ensure a quick response, the warning system must be well designed to result in no motor interference, as well as only interrupt when the initial task is not in immediate mode. Cultural differences may guide how we design things to maximize efficiency to take advantage of people's preferred strategies, i.e., design products that favor sequential strategies for Americans and simultaneous strategies for East Asians. And even our basic cognitive experiment may have broad real-world applications: the importance of guiding strategies and giving people proper feedback about how they are doing, as we did via points. It may, for instance, be better to teach people musical instruments via a points system (as in the popular video game Guitar Hero™) so they can measure their improvement.

Bibliography

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.

Allport, D. A., & Wylie, G. (1999). Task-switching: Positive and negative priming of task-set. In G. W. Humphreys, J. Duncan, & A. M. Treisman (Eds.), *Attention, space and action: Studies in cognitive neuroscience* (pp. 273–296). Oxford, England: Oxford University Press.

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science*, *26*, 39–83.

Altmann, E. M., & Trafton, J. G. (2004). Task interruption: Resumption lag and the role of cues. In *Proceedings of the 26th Annual Conference of The Cognitive Science Society*.

Altmann, E. M., & Trafton, J. G. (2007). Timecourse of recover from task interruption: Data and a model. *Psychonomic Bulletin & Review*, *14(6)*, 1079–1084.

Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought.*. Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Taatgen, N. A., & Byrne, M. (2005). Learning to achieve perfect time shar-

ing: Architectural implications of Hazeltine, Teague & Ivry. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 749–761.

Byrne, M., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review*, *108*, 847–869.

Caroselli, J. S., Hiscock, M., & Roebuck, T. (1997). Asymmetric interference between concurrent tasks: An evaluation of competing explanatory models. *Neuropsychologia*, *35*, 457–469.

Chavay, P., & Rogoff, B. (1999). Cultural variation in management of attention by children and their caregivers. *Developmental Psychology*, *35*, 1079–1090.

Chisholm, C. D., Dornfeld, A. M., Nelson, D. R., & Cordell, W. H. (2001). Work interrupted: A comparison of workplace interruptions in emergency departments and primary care offices. *Annals of Emergency Medicine*, *38*, 146–151.

Chua, H., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of The National Academy of Sciences*, *102*, 12629–12633.

de Jong, R., Coles, M. G., & Logan, G. D. (1990). In search of the point of no return: The control of response processes. *Journal of Experimental Psychology: Human Perception and Performance*, *16(1)*, 164–182.

Dickman, S. J., & Meyer, D. E. (1988). Impulsivity and speed-accuracy tradeoffs in information processing. *Journal of Personality and Social Psychology*, *54*, 274–290.

Edwards, D. C., & Metz, C. E. (2006). Analysis of proposed three-class classification decision rules in terms of the ideal observer decsion rule. *Journal of Mathematical Psychology*, *50(5)*, 478–487.

Fitts, P., & Seeger, C. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*, 199–210.

Friedman, N. P., Miyake, A., & Young, S. E. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137(2)*, 201–225.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*, 267–314.

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 299–322). Cambridge, MA: The MIT Press.

Gray, W. D., Sims, C. R., Fu, W. T., & Scholles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113(3)*, 461–482.

Green, C. S., & Bavelier, D. (2006). Effect of action video games on spatial distribution of visualspatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1465–1478.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: John Wiley and Sons.

Hedden, T., Ketay, S., & Aron, A. (2008). Cultural influences on neural substrates of attentional control. *Psychological Science*, *19(1)*, 12–17.

Howes, A., Lewis, R. L., & Vera, A. (2007). Bounding rational analysis: Constraints on asymptotic performance. In W. D. Gray (Ed.), *Integrated models of cognitive systems.*. Oxford University Press.

Howes, A., Lewis, R. L., & Vera, A. (unpublished manuscript). Cognitively bounded rational analysis.

Howes, A., Lewis, R. L., Vera, A., & Richardson, J. (2005). Information-requirements grammar: A theory of the structure of competence for interaction. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*. Italy.

John, B. E., Prevas, K., Salvucci, D. D., & Koedinger, K. (2004). Predictive human performance modeling made easy. In *Proceedings of CHI 2004* (pp. 455–462). Vienna, Austria: ACM New York.

Kaufman, L., & Rousseuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. New York: John Wiley.

Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J. (2003). Perceiving an object and its context in different cultures: A cultural look at new look. *Psychological Science*, *14*, 201–206.

Kitayama, S., Duffy, S., & Uchida, U. (2008). Self as cultural mode of being. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 136–174). New York: Guilford Press.

Kitayama, S., Ishii, K., Imada, T., Takemura, K., & Ramaswamy, J. (2006). Voluntary settlement and the spirit of independence: Evidence from Japan's "northern frontier". *Journal of Personality and Social Psychology*, *91*, 369–384.

Knill, D. C. (1998). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Research*, *38*, 1683–1711.

Laird, J. E., & Rosenbloom, P. (1996). The evolution of the Soar cognitive architecture. In D. M. Steier & T. M. Mitchell (Eds.), *Mind Matters: A Tribute to Allen Newell* (pp.

1–50). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

URL `citeseer.ist.psu.edu/laird94evolution.html`

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91(3)*, 295–327.

Luce, R. D. (1986). *Response times*. New York: Oxford University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, NJ: Lawrence Erlbaum Associates, 2nd edition ed.

Maloney, L. T., Trommershäuser, J., & Landy, M. S. (2007). Questions without words: A comparison between decision making under risk and movement planning under risk. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 297–313). New York, NY: Oxford University Press.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, *81*, 922–934.

Meyer, D. E., Abrams, R. A., & Kornblum, S. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, *95(3)*, 340–370.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychology Review*, *104*, 3–65.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychology Review*, *104*, 749–791.

Meyer, D. E., & Kieras, D. E. (1999). Precis to a practical unified theory of cognition and action: Some lessons from EPIC computational models of human multiple-task performance. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII. Cognitive Regulation of Performance: Interaction of Theory and Application* (pp. 17–88). Cambridge, MA: MIT Press.

Meyer, D. E., Kieras, D. E., Lauber, E. J., Schumacher, E. H., Glass, J. M., & Zurbriggen, E. (1995). Adaptive executive control: Flexible multiple-task performance without pervasive immutable response-selection bottlenecks. *Acta Psychological*, *90*, 163–190.

Miyamoto, Y., Nisbett, R. E., & Masuda, T. (2006). Culture and the physical environment: holistic versus analytic perceptual affordances. *Psychological Science*, *17*, 113–119.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, *67*, 949–971.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual Information Processing*. New York: Academic Press.

Newell, A. (1990). *Unified theories of cognition..* Cambridge, MA: Harvard University Press.

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently and why*. New York: Free Press.

Nisbett, R. E., & Masuda, T. (2003). Culture and point of view. *Proceedings of The National Academy of Sciences*, *100*, 11163–11170.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review*, *108*, 291–310.

Osman, A., Kornblum, S., & Meyer, D. E. (1986). The point of no return in choice reaction time: Controlled and ballistic stages of response preparation. *Journal of Experimental Psychology: Human Perception and Performance*, *12(3)*, 243–258.

Osman, A., Kornblum, S., & Meyer, D. E. (1990). Does motor programming necessitate response execution? *Journal of Experimental Psychology: Human Perception and Performance*, *16(1)*, 183–198.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). New York: Halstead Press.

Pashler, H. (1994). Dual-task interference in simple tasks. *Psychological Bulletin*, *116*, 220–244.

Peters, M. (1977). Simultaneous performance of two motor activities: The factor of timing. *Neuropsychologia*, *15*, 461–465.

Polk, T. A., & Farah, M. J. (1998). The neural development and organization of letter recognition: Evidence from functional neuroimaging, computational modeling, and behavioral studies. *Proceedings of The National Academy of Sciences*, *95*, 847–852.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107(2)*, 358–367.

Rogers, R., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124(2)*, 207–231.

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *27(4)*, 763–797.

Schumacher, E. H., Lauber, E. J., Glass, J. M., Zurbriggen, E., Gmeindl, L., Kieras, D. E., & Meyer, D. E. (1999). Concurrent response-selection processes in dual-task performance: Evidence for adaptive executive control of task scheduling. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 791–814.

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., & Meyer, D. E. (2001). Virtually perfect time-sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science*, *12(2)*, 101–108.

Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., & Yaniv, I. (1994). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R. Sternberg & J. Davidson (Eds.), *The nature of insight* (pp. 65–124). MA: MIT Press.

Shweder, R. A., & Bourne, E. J. (1982). Does the concept of person vary cross-culturally? In A. J. Marsella & G. M. White (Eds.), *Cultural conceptions of mental health and therapy* (pp. 130–204). London: Reidel.

Speier, C., Valacich, J. S., & Vessey, I. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences Journal*, *30(2)*, 337–360.

Sperling, G., & Dosher, B. A. (1986). Strategy optimization in human information processing. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Volume 1. Sensory processes and perception* (pp. 2–1 – 2–65). New York: John Wiley and Sons.

Sternberg, S. (1969a). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276–315). Amsterdam: North-Holland.

Sternberg, S. (1969b). On the discovery of processing stages: Some extensions of Donders method. *Acta Psychological*, *30*, 276–315.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcment Learning: An Introduction*. MIT Press/Bradford Books.

Trafton, J. G., Altmann, E. M., & Brock, D. P. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, *58(5)*, 583–603.

Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.

Vera, A. H., Howes, A., McCurdy, M., & Lewis, R. L. (2004). A constraint satisfaction approach to predicting skilled interactive performance. In *CHI'04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 121–128). New York, NY: ACM Press.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high speed performance. A review and a theory. *British Journal of Psychology*, *43*, 2–19.

Yates, J. F., Lee, J.-W., & Sieck, W. R. (2002). Probability judgment across cultures. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristic and biases: The psychology of intuitive judgment* (pp. 271–291). New York, NY: Cambridge University Press.