A Cognitively Bounded Rational Analysis Model of Dual-Task Performance Trade-Offs

Christian P. Janssen, Duncan P. Brumby, John Dowell, and Nick Chater
c.janssen@ucl.ac.uk, brumby@cs.ucl.ac.uk, j.dowell@cs.ucl.ac.uk, n.chater@ucl.ac.uk
University College London, Gower Street, London WC1E 6BT, UK

Abstract

The process of interleaving two tasks can be described as making trade-offs between performance on each of the tasks. This can be captured in performance operating characteristic curves. However, these curves do not describe what, given the specific task circumstances, the optimal strategy is. In this paper we describe the results of a dual-task study in which participants performed a tracking and typing task under various experimental conditions. An objective payoff function was used to describe how participants should trade-off performance between the tasks. Results show that participants' dual-task interleaving strategy was sensitive to changes in the difficulty of the tracking task, and resulted in differences in overall task performance. To explain the observed behavior, a cognitively bounded rational analysis model was developed to understand participants' strategy selection. This analysis evaluated a variety of dual-task interleaving strategies against the same payoff function that participants were exposed to. The model demonstrated that in three out of four conditions human performance was optimal; that is, participants adopted dual-task strategies that maximized the payoff that was achieved.

Keywords: multitasking; performance operating characteristic; cognitively bounded rational analysis

Introduction

Multitasking behavior often involves trade-offs in performance (e.g., time, errors, extension, etc.) between the tasks. Such trade-offs can be described graphically with Performance Operating Characteristics, which show how the performance of separate tasks vary together systematically (Navon & Gopher, 1979; Norman & Bobrow, 1975). Trade-offs reflect strategic choices and can be modified, for example, in response to instructions to prioritize one task over another (e.g., Brumby, Salvucci, & Howes, 2009; Janssen & Brumby, in press).

Consideration of the strategic choices made in multitasking (i.e., of *why a specific* way of performing the tasks is chosen) naturally supposes some optimal trade-off. Why time is allocated differentially to the tasks, and why particular patterns of interleaving are adopted, must reference the relative success of those different strategies. In this paper we use an objective payoff function to integrate into a single score the performance rewards in a tracking-while-typing dual-task situation. Such payoff functions have been used before in multitask studies, but only to show that performance is sensitive to isolated factors such as changes in reward structure (e.g.,Wang, Proctor, & Pick, 2007). Objective payoff functions have not previously been used to support explanations of multitasking strategy choices, or to assess the optimality of strategies.

Combined with a cognitive model that can perform alternative multitasking strategies (i.e., alternatives for when to interleave and execute multiple tasks), a payoff function enables an evaluation of the success of each of the strategies (Howes, Lewis, & Vera, 2009). Strategies with the highest payoff can be determined and compared with human performance in experimental settings. This can be used to explain the strategic choices participants make.

We developed a tracking-while-typing dual-task to test the hypothesis that people can hone their dual-task behavior to maximize the payoff that is achieved. The task required participants to keep a randomly moving cursor inside a circular area and to type a string of digits. Tracking tasks have been used in several multitasking studies (e.g., Gopher, 1993; Hornof, Zhang, & Halverson, 2010; Kieras, Meyer, Ballas, & Lauber, 2000; Lallement & John, 1998; Salvucci & Taatgen, 2008). For example, Gopher (1993) showed that performance trade-offs in a tracking-while-typing task can be influenced by instructions to spend more time on one of the tasks. Within the cognitive modeling literature, the work by Lallement and John (1998) is interesting as it compares performance of models developed in several cognitive architectures on a tracking and choice task. We attempt to extend this work by showing how a payoff function enables us to bind normative cognitive models with experimental observations of multitasking behavior, and specifically, to show how multitasking strategy choice can be better explained by seeing it in relation to optimal performance.

Experiment

Method

Participants Eight participants (4 female) between 20 and 35 years of age (M = 23 years) from the subject pool at UCL participated for monetary compensation. Payment was based on performance (details are provided in the Materials section). The total payment achieved by participants ranged between £7.13 and £11.45 (M = £9.14).

Materials The dual-task setup required participants to perform a continuous tracking task and a discrete typing task, presented on a single monitor. Figure 1 shows the layout of the tasks on the display. The typing task was presented on the left side and the tracking task on the right. Each task was presented within a 450 x 450 pixels area, with a vertical separation of 127 pixels between the tasks.

The tracking task required participants to keep a square cursor that drifted about the display in a random fashion inside a target circle (see Figure 1). The cursor was 10 x 10 pixels and the target had a radius of either 80 (small target) or 120 pixels (large target). A random walk function was used to vary the position of the cursor in the display every 20 milliseconds. The rate at which the cursor drifted about the display was varied between different experimental conditions. In a low noise condition the random walk had a mean of zero and standard deviation of 3 pixels per update, while in a high noise condition the random walk had a mean of zero and standard deviation of 5 pixels per update.

Participants used a Logitech Extreme 3D Pro joystick with their right-hand to control the position of the cursor in the tracking display. The drift function of the cursor was suspended whenever the joystick angle was greater than 0.08 (the maximum angle was 1). The speed was scaled by the angle, with a maximum of 5 pixels per 20 milliseconds.

The typing task required participants to enter a string of twenty digits using a numeric keypad with their left-hand. The string was made up of the digits 1 to 3, where each digit occurred at least six times in a given sequence. Digits were presented in a random order with the constraint that no single digit was presented more than three times in a row in the sequence (e.g., “11233322132123132123” as in Figure 1). When a digit was entered correctly it was removed from the to-be-entered sequence. In this way, the left-most digit on the display was always the next one to be entered.

The study used a forced interleaving paradigm, in which only one of the two tasks was visible and could be worked on at any moment. By default the typing task was visible and the tracking task was covered by a gray square. In order to see and control the tracking task, participants had to hold down the trigger of the joystick. When the trigger was released, the tracking task was covered by a gray square and the typing task revealed.

Design The study manipulated aspects of the tracking task using a 2 (cursor noise: low vs. high) x 2 (target size: small vs. large) within-subjects design. The main dependent variables were the time required to complete the typing task and maximum distance of the cursor from the center of the target circle.

Participants were remunerated based on performance, as determined by an objective payoff function that was calculated for each dual-task trial. The function was designed to encourage fast completion of the typing task, while keeping the cursor inside the target. The payoff (in pounds) for a given trial was defined as:

Payoff = Gain + Digit Penalty + Tracking Penalty

The minimum payoff for a given trial was limited to £-0.20. The gain component was based on the total time required to complete a dual-task trial (in seconds):

$$\mathbf{Gain = 0.15 * e^{-1*TotalTrialTimeInSec/20 + 0.25}}$$

This function was determined using pilot studies, to make sure participants mostly gained money. To encourage accurate typing, a digit penalty deduced £0.01 from the total payoff whenever an incorrect digit was entered. To encourage participants to keep the cursor inside the target circle of the tracking task, a tracking penalty was applied:

$$\mathbf{Tracking\ Penalty = -0.1*e^{SecOutside/1.386 - 0.6931}}$$

This penalty was crafted such that £0.10 was lost when the cursor was outside of the radius for 0.5s, and £0.20 was lost when it was outside of the radius for 1s. In the remainder of this paper we will not look at the effect of digit penalty on payoff.

Procedure Participants were informed that they would be required to perform a series of dual-task trials and that they would be paid based on their performance. A participant’s payment was based on the cumulative payoff over the course of the study, in addition to their base payment of £3. Participants were told that they would gain more points by completing the typing task as quickly as possible, but that they would lose points if they made a typing error or if the cursor drifted outside of the target area in the tracking task. We chose not to give participants a formal description of the payoff function, but instead provided explicit feedback after every dual-task trial with the payoff score achieved.

After explaining how to perform each of the tasks participants performed two single-task training trials for each task and two dual-task practice trials. Participants were instructed that for dual-task trials only one of the two tasks would be visible and controllable at any moment in time, and they were instructed how to switch between tasks.

Participants then completed four blocks of experimental trials (one for each experimental condition). The order of conditions was randomized and counter-balanced across participants, with the exception that blocks of the same noise level were grouped together. The order of radius sizes was repeated across the first two blocks and the second two blocks. For each block, participants completed five single-task tracking trials, five single-task typing trials, and twenty dual-task trials. The dual-task trials were further grouped into sets of five trials, with a short pause between each set. The total procedure took about one hour to complete.

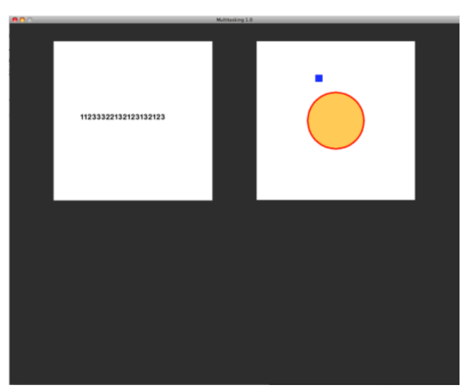

Figure 1: Position of the two tasks in the interface

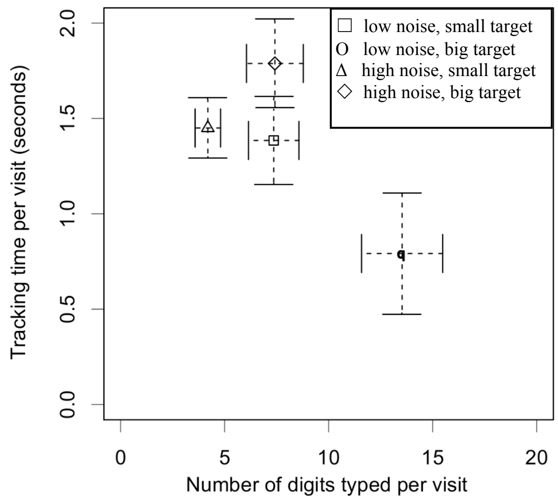

Figure 2: Number of digits typed and tracking time, both per visit. Error bars depict standard errors.

Results

We focus on performance during the last five dual-task trials of each experimental condition, as these reflect a period during which the participant has had time to adapt behavior to the objective payoff function. A 2 (cursor noise) x 2 (target size) analysis of variance (ANOVA) was used for all statistical analysis with a significance level of .05.

Overall performance We first consider the effect of varying aspects of the tracking task on the time required to complete the typing task and the maximum distance of the cursor from the center of the target circle in the tracking task. It was found that trial time was significantly longer when there was greater noise in the tracking task (M = 11.17s, SD = 4.32s) than when there was a lower level of noise in the tracking task (M = 7.51s, SD = 2.00s), $F(1, 7) = 15.07$, $p < .01$. Trials were also longer when the target in the tracking task was smaller (M = 10.59s, SD = 4.01s) than when it was larger (M = 8.09s, SD = 3.22s), $F(1, 7) = 11.84$, $p = .01$. There was no significant interaction, $F(1, 7) = 0.22$.

In the tracking task we consider the maximum distance of the cursor from the center of the target over the course of a trial. It was found that the cursor drifted more when there was a higher level of noise (M = 95 pixels, SD = 15 pixels) than when there was a lower level of noise (M = 61 pixels, SD = 8 pixels), $F(1,7)=33.42$, $p < .001$. There was no effect of target size on the maximum distance that the cursor drifted over a trial ($F(1,7) = 1.19$, $p = .31$), nor was the interaction effect significant ($F(1,7) = 0.07$).

These differences in overall task performance between conditions are somewhat expected and unsurprising because they partly reflect differences in the difficulty of the tracking task. We were far more interested in how this performance was achieved. We next consider the dual-task interleaving strategy that was adopted in each condition.

Strategies Two aspects determine a strategy: (1) the number of digits typed during each visit to the typing window and (2) the amount of time spent in the tracking window per visit to this window. Figure 2 shows these two basic strategy dimensions for each of the four conditions. It can be seen that for each experimental condition there is a unique point in this strategy space – strategies differ between conditions. The number of digits entered per visit increased with an increase in target size ($F(1, 7) = 17.4$, $p < .01$), and it also increased with a decrease in cursor noise (that is, more digits were typed when it took longer for the cursor to cross the boundary; $F(1, 7) = 15.18$, $p < .01$). There was no significant interaction ($F(1, 7) = 3.24$, $p = .12$).

It can also be seen in Figure 2 that the time spent in the tracker window per visit increased with an increase in the noise associated with the cursors movement ($F(1,7)=14.98$, $p = .01$). An interaction effect was present as visit time was particularly short in the low noise, large radius condition ($F(1,7)=11.55$, $p = .01$). There was no significant effect of radius ($F(1,7)=0.54$).

A CBRA Model of Tracking-while-Typing

The results show that participants adapted their dual-task behavior to changes in the difficulty of the tracking task. However, what these results do not show is whether participants were adopting a strategy that is *optimal* in terms of maximizing the expected payoff that could be achieved in each condition. To answer this question we developed a cognitively bounded rational analysis model (Howes, et al., 2009). This framework is particularly useful for comparing the performance of alternative strategies, allowing strategies to be discriminated based on the payoff achieved. The model developed here is inspired by our previous work in developing models of a dialing-while-driving dual-task set-up (e.g., Brumby, Salvucci, & Howes, 2007; Brumby, et al., 2009; Janssen & Brumby, in press). Both dual-task environments share core characteristics, but the current work differs in that it incorporates an explicit payoff function against which various dual-task interleaving strategies can be evaluated. In the next section, we use a model to determine whether people were selecting strategies that would maximize the financial payout that could be achieved in each condition.

Model Development

Tracking Model The crucial question for developing a model of the tracking task was at what angle participants held the joystick given their current distance from the center of the target. Figure 3 shows the mean values for discrete bins of 5 pixels for the horizontal axes (vertical data is similar). We fitted a linear function (shown as a dotted line):

Angle = -0.01 * current distance from target

The joystick had a maximum angle of (-)1. As in the experiment, the speed of the cursor is calculated by multiplying the angle of the joystick with a value of 5 pixels. Speed is calculated once every 250 milliseconds of tracking, and the cursor position is updated every 20 milliseconds based on this speed value. As in the

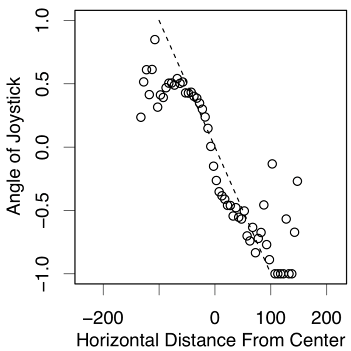

Figure 3: Angle of the joystick as a function of distance from the target. The dashed line shows a fitted function.

experiment, the cursor could only be controlled when the tracking window was open. The total time spent tracking in dual-task was varied as part of the strategy (see below).

Typing Model To model the typing task we fitted model performance to human single-task typing performance data. The time taken to type a digit (260 milliseconds) is identical to the mean inter-keypress interval measured in single-task.

Dual-Task Model The dual-task model works as followed. The model starts of with typing a series of digits (the length of which is varied as a strategy). For switching between typing and tracking a switch cost of 250 milliseconds is incurred, based on experimental data (time between last key press and pressing the trigger on the joystick: 247 milliseconds). The model then tracks the cursor for a designated amount of time (varied between runs as a strategy aspect). When it switches back to typing, a switch cost of 180 milliseconds is incurred (time between releasing the trigger and pressing the first key press minus single task inter-keypress interval: 185 milliseconds). Noteworthy, switch cost values are close to those in ACT-R models (e.g., Borst, Taatgen, & Van Rijn, 2010) and in the Cognitively Bounded Rational Analysis driving models.

Strategies We used this basic model to explore performance of a variety of strategies. A strategy is determined by the number of digits that are typed in sequence during a visit to the target window. We consider only a subset of twenty simple strategies that placed a consistent number of digits during each visit (between 1 and 20), with the exception of the last visit during which the remaining digits were placed (e.g., strategy 6-track-6-track-6-track-2, but not 6-track-4-track-6-track-4). In addition, for each visit to the tracking task, more or less time can be spent on tracking. We systematically explored performance for models that spent between 250 to 3000 milliseconds on tracking during each visit to the tracking window, using steps of 250 milliseconds (i.e., 12 alternatives). This gave a total of 229 (20 x 12 – 11) strategy alternatives.

The objective function for rating performance is similar as in the experiment with the exception that the model does not make typing errors. For each strategy alternative 100 runs were performed. Mean performance is reported.

Model Results

The first question of interest was whether the model would fit the experimental data. In particular, if we hardcode a strategy that types the same number of digits per visit and spends about the same amount of time tracking as participants did in each condition (with both measures lying within two standard errors of human means), does this then

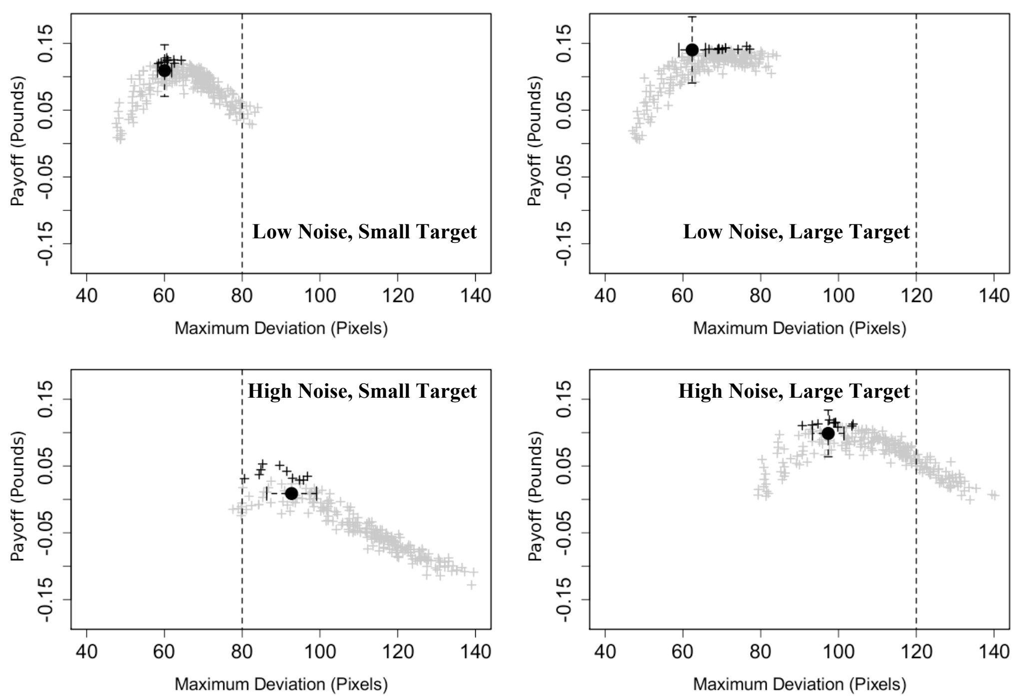

Figure 4: Maximum deviation versus predicted payoff per trial for the ten best (black crosses), and other strategies (gray crosses) per condition. Human results are shown as circles with standard error. The dotted line shows the target boundary.

result in similar total trial time and maximum deviation in each experimental condition (again with performance within two standard errors of the mean)? This is important so as to know that we have a reasonable calibration of the model's performance relative to the human data. This was the case.

Given that we can be confident that the model is reasonably calibrated to the human data on the observed strategy, we can now use the model to evaluate the payoff achieved by different (unobserved) dual-task interleaving strategies. Figure 4 shows a plot of the average maximum deviation versus payoff. We plotted the ten highest scoring strategies with black crosses, and the other strategies with gray crosses. In each condition there is a strong peak, though the shape of the distribution of scores differs between experimental conditions. In three out of four conditions the human data (black circles) lies in the region of maximum deviations that can achieve the highest performance. In each figure a vertical line shows the boundary of the target. Note that in the low noise, large radius condition participants could have let the cursor drift more to improve their score slightly (they would never cross the target boundary). Due to space limitations, we omitted a plot of total time data versus score; the pattern is similar.

Traditionally, differences in dual-task performance are plotted in Performance Operating Characteristics (POCs), in which performance on one measure or task is shown against performance on the other measure or task (Navon & Gopher, 1979; Norman & Bobrow, 1975). In Figure 5 we show the POC of total time and maximum deviation for the model and human data. The ten best performing strategy alternatives are again plotted with black crosses. There are a couple of interesting aspects to these graphs. First, the best performing models lie on the outer edge (left side, and bottom side) of the strategy space: the trade-off curve. That is, the best strategies make an optimal trade-off between performance on the two tasks. Furthermore, the position of the optimum strategies is at a *different section* (e.g., top left, or bottom right) of this curve for each condition. The model is essential for this assessment, as traditional POCs cannot predict optimal regions by themselves.

Human data again lies in the region of optimum payoff for three out of four conditions. Only in the low noise large target condition could participants have scored better by spending less time on the tracking task (increasing maximum deviation, but decreasing trial time). In all other conditions, the model illustrates that participants made good performance trade-offs to optimize their payoff.

General Discussion

In this paper we have presented an experiment and a model of a tracking-while-typing dual-task setup. A good feature of the task environment, in which participants need to track a cursor and type in digits, is that it translates performance on both tasks into a single performance score. Due to this feature, we were able to move beyond observations that participants trade-off performance in tasks, as done in classical dual-task research (Navon & Gopher, 1979; Norman & Bobrow, 1975) and in research on dual-task driving behavior (e.g., Janssen & Brumby, in press). Here, we were able to demonstrate that participants mostly made performance trade-offs in an optimal manner, so as to maximize pay-off (cf. Howes et al., 2009).

These claims are possible because of the use of a payoff

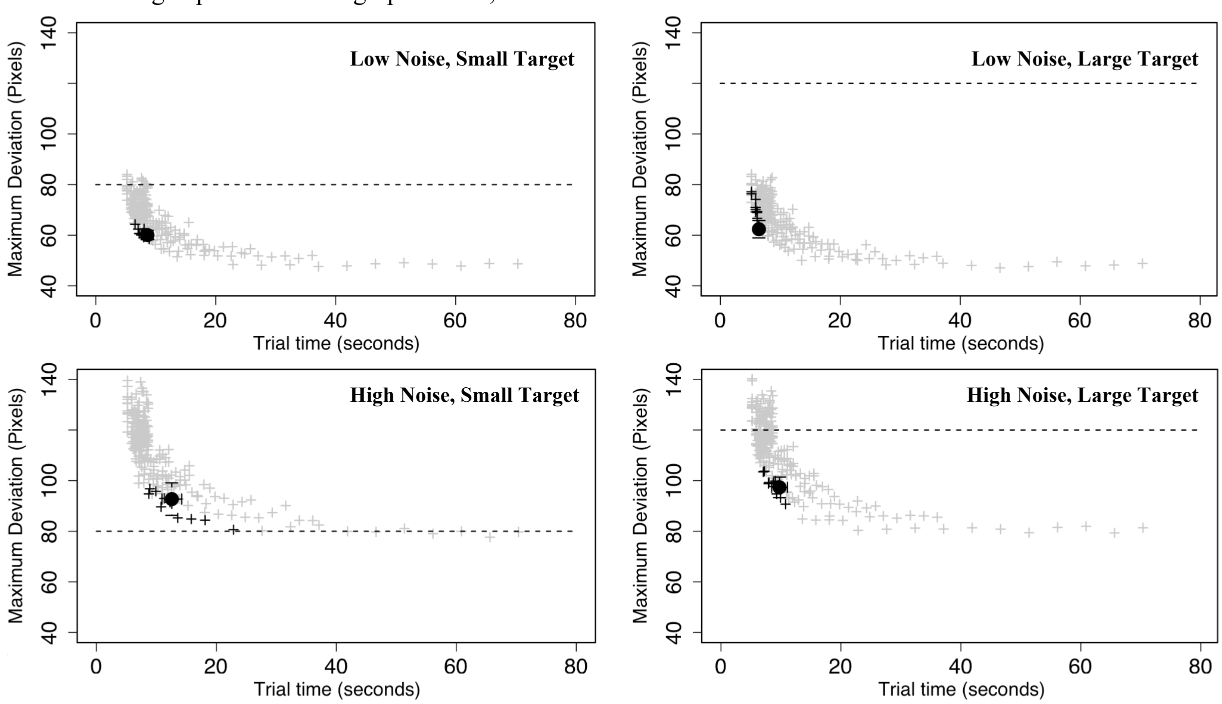

Figure 5: POCs of model performances for the ten best (black crosses), and other strategies (gray crosses) per condition. Human results are shown as circles with standard error. The dotted line shows the target boundary.

function that explicitly describes how participants ought to trade performance on each task to gain payment. The goal of this paper is not to argue that objective functions are the most prevalent aspect of performance in the real world. However, they make it possible to quantify how good performance is. This contrasts with previous work where verbal instructions on how to trade performance on each task is given (e.g., Gopher, 1993; Horrey, Wickens, & Consalus, 2006; Levy & Pashler, 2008), or where performance is sensitive to a change in payment (e.g., Wang, et al., 2007). In contrast, we can define *optimal* performance in terms of maximizing payoff.

There was however one condition (the low noise, large target condition) in which participants did not maximize the payoff that was achieved. In this condition, participants could have typed all of the digits in one sequence (i.e., without multitasking) to receive a slightly higher payoff than was actually observed. Two possible explanations for suboptimal performance are that participants overestimated the danger of the cursor crossing the boundary (which would give a severe penalty), or they were biased to switch between the two tasks (which is necessary in the other conditions). In this sense, participants not always adapt their behavior to maximize the payoff function. Further research is required to investigate such biases.

The model was developed with a minimal set of assumptions. This was already enough to demonstrate that people mostly adapt performance to an objective function. Further research can investigate how people adapt their behavior to different payoff functions, which, for instance, give greater weight to performance on one of the two tasks. Also, the model of the typing task might be refined to predict typing errors, and to predict the effect of the different times needed to type repeating digits versus non-repeating digits (cf. Janssen, Brumby, & Garnett, 2010). At a different level of analysis, the role of eye-movements can be considered to explore a wider variety of strategies (cf. Hornof, et al., 2010), such as strategies in which some visits to the typing task window are only spent on studying, and not typing digits.

Acknowledgments

This work was supported by EPSRC grant EP/G043507/1. We thank Julian Marewski and two anonymous reviewers for their valuable comments on this paper.

References

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, memory, and cognition, 36*, 363-382.

Brumby, D. P., Salvucci, D. D., & Howes, A. (2007). Dialing while driving? A bounded rational analysis of concurrent multi-task behavior. In *Proceedings of the 8th International Conference on Cognitive Modeling*

Brumby, D. P., Salvucci, D. D., & Howes, A. (2009). Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1629-1638).

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In *Attention and performance XIV* (pp. 299–322). Cambridge, MA: MIT Press.

Hornof, A. J., Zhang, Y., & Halverson, T. (2010). Knowing where and when to look in a time-critical multimodel dual task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied, 12*, 67-78.

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review, 116*, 717-751

Janssen, C. P., & Brumby, D. P. (in press). Strategic adaptation to performance objectives in a dual-task setting. *Cognitive Science*.

Janssen, C. P., Brumby, D. P., & Garnett, R. (2010). Natural break points: Utilizing motor cues when multitasking. In *Proceedings of the 54th annual meeting of the Human Factors and Ergonomics Society*. San Francisco, CA, USA: Human Factors and Ergonomics Society.

Kieras, D. E., Meyer, D. E., Ballas, J. A., & Lauber, E. J. (2000). Modern computational perspectives on executive mental processes and cognitive control: Where to from here? In *Attention and performance XVIII* (pp. 681-712). Cambridge, MA: MIT Press.

Lallement, Y., & John, B. (1998). Cognitive architecture and modeling idiom: An examination of three models of the wickens task. *Proceedings of the twentieth annual conference of the Cognitive Science Society*, 597-602.

Levy, J., & Pashler, H. (2008). Task prioritisation in multitasking during driving: Opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology, 22*, 507-525.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86*, 214-255.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44-64.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review, 115*, 101-130.

Wang, D. D., Proctor, R. W., & Pick, D. F. (2007). Acquisition and transfer of attention allocation strategies in a multiple-task work environment. *Human Factors, 49*, 995-1004.