Concurrent multitasking: do people optimally use cognitive resources?

Pieter de Rooij, P.de.Rooij@student.rug.nl,
Supervisors: N. A. Taatgen*, M. Nijboer*

August 4, 2014

Abstract

How good are people at multitasking when given free choice in what tasks they perform? This question has yet to be answered in concurrent multitasking. Are they able to minimize load (lessen interference) to achieve the best performance? An experiment was conducted where participants were able to freely choose which combination of tasks they want to do at any time. The results show that people take on more tasks as the experiment progresses, because this maximizes their score. Also, the task combination predicted to have the least interference by the threaded cognition theory, is in fact the most preferred condition of the participants. The task combination which is expected to be disliked most does not significantly decrease over time compared to the most preferred one however. These findings indicate that the score system is likely to be a contributor to the intrinsic motivation of participants.

Keywords: multitasking, threaded cognition, interference, choice, motivation.

1 Introduction

1.1 Background

This paper is about the decisions people make in concurrent multitasking. Questions involved are: do people adapt their preference for tasks based on how easily these are combined? Do their decisions allow them to optimally use their cognitive resources? Performance plays an important role in these questions: how well do people perform and how does this compare to the theoretically optimal performance?

Some people seem to perform better in multitasking than others. A few show no decrement at all in dual-task performance compared to single-tasking. These are so-called 'supertaskers' (Watson & Strayer, 2010). This raises an interesting question: what affects multitasking performance? Some findings suggest that working memory, attention and fluid intelligence (Konig, Buhner, & Murling, 2005) reliably predict multitasking performance in the "Simultaneous capacity/Multitasking" (SIMKAP) scenario. The neural correlates of multitasking have been examined by Dux et al. (2009), who concluded that increased processing speed in the prefrontal cortex likely contributed to multitasking performance, rather than a functional reorganization of the brain circuits supporting multitasking. Such studies seem to point at a wide range of possible factors affecting multitasking performance. This raises the question whether all of them are truly independent, or that there is some underlying principle causing these observations? If the last is the case, then what underlies these findings and how can it all be united with physical observations?

A theory that has proven useful in explaining ob-

*University of Groningen, Department of Artificial Intelligence

served performance during multitasking is that of *threaded cognition* (D. Salvucci & Taatgen, 2010). It states that multitasking behaviour can be represented as multiple processes that can run simultaneously. Problems can occur between these threads as soon as they call on the same cognitive module or modules at the same time. For instance the processing of two images at the same time by the visual module, or the simultaneous performance of two different moves with both hands by the motor module. This is called interference. The cognitive modules thus limit information processing, becoming a bottleneck, as only a single thread is served at a time. Overlap in the cognitive demands of the performed tasks thus reduces multitasking ability.

An example that builds on these bottlenecks is the work by Borst, Taatgen, Stocco, and van Rijn (2010). It posits that people are only able to retain one problem state. The problem state module is used to maintain intermediate results, for example intermediate steps in algebra, or remembering the next turn in a route. The bottleneck becomes apparent as soon as multiple tasks demand this resource, like in the situation that one is rehearsing the route during the algebra. Borst et al. (2010) found that this problem state bottleneck is a plausible cause for the interference observed in their experiment and that it might be located in the intraparietal sulcus. This is just one of several studies that utilize threaded cognition (D. D. Salvucci & Taatgen, 2008; Taatgen, Juvina, Schipper, Borst, & Martens, 2009).

Quite some attention has been devoted to multitasking, but little is known about the decisions humans make in concurrent multitasking: when is a switch made between tasks that are performed simultaneously? Do people adapt to multitasking interference and if so, how? Are humans able to minimize interference and ultimately make optimal use of their cognitive resources?

A study attempting to answer whether humans adapt to multitasking interference is that of Nijboer, Taatgen, Brands, Borst, and van Rijn (2013). The study contained three experiments with different goals each: the first was to check whether the task combinations actually led to the expected interference pattern and preference, the second builds further on that by closer examination of the preferences and the third was to determine how much difference in interference between combinations is required to persuade participants into optimal conditions. The main conclusions drawn by Nijboer et al. (2013) are that the expected optimal task combinations indeed have the least interference: While a portion of the participants adapted to this combination, a considerable fraction only slowly converged to this solution, or not at all.

This research will continue on from those findings. Foremost, we address the reluctance in taking on more or different tasks. Our main concern will be motivating people into switching tasks. Should a method be found to accomplish this, it will allow us to investigate whether humans are capable of minimizing interference, given an intrinsic motivation.

1.2 Paradigm

To answer this question, a similar set-up as Nijboer et al. (2013) was used. In addition a score system is included to encourage subjects to take on as much tasks as possible. Besides this change, a fixed primary task might (implicitly) lead to that task being prioritized. For that reason there shall be no fixed task and subjects can freely choose which task to perform.

The tasks used for this purpose are 1-back, tracking and tone counting. The cognitive resources each requires is believed to be as follows:

1-back The objective of 1-back is to respond whether the letter presented previously is equal to the current one. This task is expected to use working memory (WM) for remembering the last letter. Visual resources are also needed for processing the stimuli. Lastly motor resources are required for initiating a response.

Tracking A participant has to keep a circular cursor on top of a moving dot. This task is expected to require visual resources to process visual information and motor resources for controlling the circle.

Tone counting The goal of this task is to count high tones in a stream of either high or low tones. As soon as the count reaches five, this should be reported. We believe that this task involves WM for keeping track of the count, aural resources for interpreting the sounds and motor resources for reporting the fifth high tone.

Since the score system is expected to motivate people to do as many tasks as possible, supposedly most of the time two or all three of the tasks are performed concurrently. It is also expected that participants converge to a strategy where the interference is minimal. This is in the case where tone counting and tracking are done simultaneously. These tasks have the least overlap since tone counting is aural and tracking visual. Also, tone counting is the only task that requires WM in this combination. Should tone counting be combined with 1-back, there is overlap in WM demand. Should tracking be combined with 1-back, both demand visual resources. Therefore the other combinations are not optimal.

2 Methods

2.1 Participants and materials

For this experiment, 22 students agreed to participate with informed consent (12 male/10 female, age: $\mu = 21.4, \sigma = 2.5$). Each were granted 10 Euro for their participation. Only the data of 21 participants is used, as the scores of one participant were negative. This proved to be an indication of bad performance on the tasks. It is unclear whether this was caused by distractions or a lack of motivation or something else entirely.

At most three people performed the experiment simultaneously in a separate cabin of approximately four square metre. The monitor used is the BenQ XL2420T. The headset used is either the Sennheiser HD 201 or HD 280 pro.

Participants were allowed to do a few practice trials after having received instructions (Appendix A, page 12). The instructions contained information about how the experiment takes place and a high-score list was also included. High-scores were updated when these were achieved. The practice consisted of six trials, one for each single- (1-back, tracking and tone counting) and dual-task (1-back with tracking, 1-back with tone counting and tracking with tone counting) condition. Each trial had a duration of thirty seconds. After these trials the experiment initiates. Afterwards the subjects are asked to answer a questionnaire containing eleven questions, concerning difficulty and task preferences (a template can be found in the appendix, B on page 15).

The experiment code was written in MATLAB using Psychtoolbox (Brainard, 1997).

2.2 Tasks

Here follows a detailed description of how the three tasks appeared during the experiment.

2.2.1 1-Back

A letter was presented for one second. After that, a mask was displayed for one and a half seconds, after which the next letter was presented. The mask serves to make it less likely that participants used the mental image of the last letter to recollect whether it was the same as the previous. Specifically, a hash tag (#) was chosen for this purpose. Participants had to press the E key in case the current letter was the same as the previous and R if it was different. A response had to be given within half a second after the stimulus disappeared, otherwise it was recorded as a wrong response.

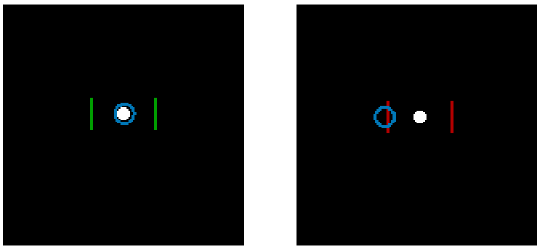
2.2.2 Tracking

The participant had to keep a circular cursor on top of a moving dot. The U and I keys moved the cursor left and right respectively. Two boundaries were drawn beside the moving dot, one to the left and one to the right of it. As long as the cursor was on top of the moving dot, these boundaries were green. Whenever the middle of the cursor exceeded a boundary, these turned red. Figure 1 shows the two different situations.

2.2.3 Tone counting

High and low tones were presented to the participant. As soon as five high tones were heard, the participant had to press the space bar. The pitch of the high tone is A4 ($\approx$ 440 Hz), the low tone is a C4 ($\approx$ 262 Hz). Feedback was provided aurally, where the negative feedback was a buzzer and the positive one sounded a bell twice.*

*https://code.google.com/p/correctwrong/



(a) Inside (b) Outside

Figure 1: Two examples in which the circular cursor is within limits or outside.

Figure 2: A typical arrangement of a dual-task condition. In this case, 1-back and tone-counting.

2.3 Display

The score display was always located slightly above the center of the screen. If only one task was performed, the task took place at the center of the display. If a participant performed two tasks at the same time, the tasks were arranged side by side. Whenever tone counting was part of the combination, a cross was drawn as fixation. For the triple-task condition 1-back and tracking were put side by side and tone counting was not provided visually at all. Figure 2 is an example of what was displayed whenever one performed 1-back and tone counting at the same time.

2.4 Score

The participants' score was calculated based on the number of tasks being performed at once and how long these tasks have been active.

2.4.1 Diminishing returns

Initially each of the three tasks yielded one point per second. This also implies that doing all three tasks together yielded three points per second.

To encourage participants to switch, the obtainable points for the current task combination decreased every second. Equation 2.2 on page 5 is the formula used in determining the new division of obtainable points.

In equation 2.2 n is the number of active tasks. *decrements* is the number of times points were deducted from active tasks: if a task reaches the minimum of 0.25 points, no further decrements are applied to it. If this is not accounted for, the zero-sum explained in section 2.4.2 is no longer true. For this reason, the added obtainable points for inactive tasks depends on the number of decrements. It is also clear that the minimum of obtainable points for a task is 0.25. If the second requirement (current gain > 0.25 or decrements > 0) is not fulfilled, *gain* stays equal to *current gain*.

2.4.2 Zero-sum

The total of three points per second for all tasks holds at all times. The score system therefore is a zero-sum game. As a formula:

$$\sum_{i=1}^{3} tp_i = 3 \tag{2.1}$$

Where tp_i is the amount of points awarded to task i. Initially it is true that $tp_i = 1$ for each task i. This means that a decrement in obtainable points resulted in an (equal) increase in obtainable points for the other task or tasks.

2.4.3 Penalty

Whenever a participant gave a wrong response, a deduction was made from their score. The deduction is not equal for all tasks, it depends on the task in which it was made. The criteria of an error and the score penalty applied for each task are:

1-Back Giving the wrong response or none at all. In these events, two points were subtracted.

Tracking Every 1000 ms that the circular cursor was considered outside the boundaries, one

$$\text{gain} = \begin{cases} \text{current gain} - 0.04/n & \text{if task is active and current gain} > 0.25 \\ \text{current gain} + \text{decrements} \cdot \frac{0.04/n}{3-n} & \text{if task is inactive and decrements} > 0 \end{cases} \quad (2.2)$$

point was subtracted from the score. The time is cumulative: that is, when the total time the cursor was outside the boundaries reached 1000 ms, the penalty was applied.

Tone counting Pressing the space bar before the count reached five was punished with a decrement of the score by three points. If the count reached six or higher, the score was decremented by three points for every time the count reached a multiple of five plus one. So the decrement was applied if the count reached six, eleven, sixteen, and onward. If the participant pressed the space bar somewhere during this time, no further points were subtracted.

The penalty system is also summarized in Table 1. Given this scoring system, if the subject just randomly chose responses, the score balanced around zero points.

task	penalty	applied if
dot tracking	1	1000 ms outside
1-back	2	wrong or no response
tone counting	3	count not five

Table 1: Table of penalties showing the deducted points per error and their criteria.

3 Results

3.1 Multitasking choices

The following plot (Figure 3, page 6) provides a brief overview of the conditions participants chose over time.

Looking at Figure 3 it is clear that the triple-task is the only task that gained popularity greatly over the duration of the experiment. Furthermore, it seems that dot tracking and tone counting together was performed most out of all combinations, and it decreased the least of all dual-tasks.

The slopes of every condition tell something about either an increase or decrease in preference over time. A comparison between them gives information about changes in preference of one condition against the other. Slopes were calculated for each condition using linear models for each participant. Table 2 gives an overview of the results of one-way analyses of variance on the change in popularity over time.

conditions	F	p
all	15.48	< 0.001
excluding triple-task	4.48	< 0.001
only dual-tasks	0.29	> 0.5
only single-tasks	0.16	> 0.5

Table 2: Results of a one-way ANOVA on the condition slopes.

A one-way analysis of variance comparing slopes of all conditions turns out to be significant ($F(6, 140) \approx 15.48$, $p < 0.001$). This implies that there are differences in the slopes of all conditions. Since the triple-task is the only one increasing continually, this result might be a consequence of including this combination. Excluding the slopes of the triple-task condition from the data does indeed reduce the F-value, but the result is still significant ($F(5, 120) \approx 4.48$, $p < 0.001$). This indicates that there is a difference in the slopes of the remaining tasks. The same test applied only to the slopes of the dual tasks proves not to be significant ($F(2, 60) \approx 0.29$, $p \approx 0.750$), which holds for the single tasks as well ($F(2, 60) \approx 0.16$, $p \approx 0.851$). This shows that the significant difference is between these collections rather than within them.

While the slopes are an indication of changes in the preference for conditions over time, the total time spent in them is an average of the preference. The sum of the time all participants spent in each condition can be found in Table 4. Comparing these times gives information about differences between the preference for one condition over the other over the duration of the experiment. Two sample, two-sided paired t-tests were used to analyse the time each participant spent in every condition. The comparison between the triple-task and

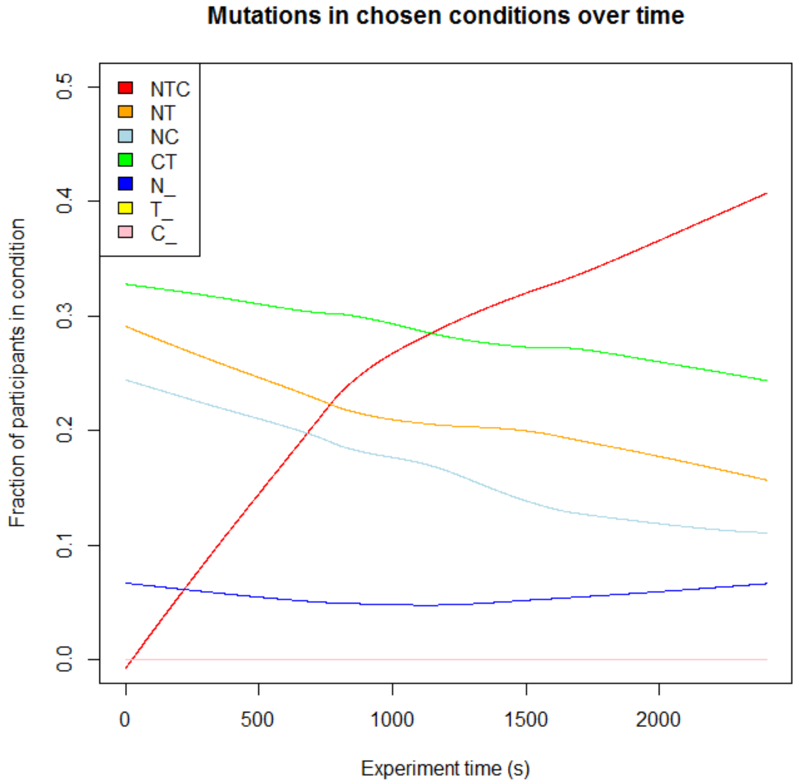


Figure 3: Fraction of participants in a certain condition at any time. Each (coloured) line represents a condition and the height is the percentage of participants who actually were in that condition at that time. The pink, yellow and darkblue line respectively represent the single-tasks tone counting (C_), dot tracking (T_) and 1-back (N_). Green, lightblue and orange represent the dual-tasks, respectively tracking and tone counting (CT), 1-back and tone counting (NC) and 1-back and tracking (NT). The red line is for the triple-task (NTC).

the dual-tasks is an exception: this is a Welch unpaired t-test since the distributions do not contain equally many observations. Note that *dual-tasks* and *single-tasks* are collections of conditions that qualify as such. An overview of the results can be found in Table 3.

conditions	t	p
triple-task vs dual-tasks	0.56	> 0.5
dual-tasks vs single-tasks	8.46	< 0.001
triple-task vs CT	-0.30	> 0.5
CT vs NC	2.98	< 0.01

Table 3: Results of two sample, two-sided paired t-tests between time spent on conditions. Dual-tasks and single-tasks are collections of those conditions.

Whether participants were rather multi-tasking than single-tasking can be analysed with the first two tests. The t-test between dual-tasks and single-tasks proves significant ($t(62) \approx 8.46$, $p < 0.001$), yet the one between the triple-task and the dual-tasks is not ($t(23.377) = 0.56$, $p \approx 0.580$). Given the positive t-value of the first test, this means that the time spent in dual-tasks is greater than in single-tasks. There is no indication though that either the triple-task or the dual-tasks are performed longer. This difference may however be obscured by the fact that the condition CT is a dual-task and it has the highest presence for some time. Only the triple-task exceeds CT in presence later in the experiment. Differences between the total time spent triple-tasking or dual-tasking may thus not be recognized because these conditions are similar in the time spent in them. This is confirmed by the non-significant result of the third test ($t(20) = -0.297$, $p \approx 0.769$). Lastly, to see whether participants at least show preference for the optimal combination (CT) over the worst combinable dual-task (NC), the last test is performed. The significant positive t-value ($t(20) = 2.98$, $p \approx 0.007$) shows that more time is spent in CT than NC.

3.2 Performance

Table 4 provides a summary of the performance at every condition. Indicated by this table is that the time spent single-tasking is relatively small.

condition	mean error	total time
N_-	0.030	2954
T_-	0.551	352
C_-	0.003	330
NT	0.441	10846
NC	0.110	8236
CT	0.322	14576
NTC	0.527	13145

Table 4: The mean error rate (errors/second) in a condition and the total time participants spent herein (seconds).

Excluding tracking, the multitasking conditions had higher mean errors than the single tasks and the triple-task had an even higher mean error than the dual-tasks.

A better means of comparison was to compare performance on each task in multitasking conditions. Figure 4 decomposes the performance on multitasking conditions per task. Table 5 are the results of one-way ANOVAs comparing the average error across conditions for each task. Again, the mean error for each task in the triple-task was higher than those in the dual-tasks, which in turn were higher than when people were single-tasking (excluding tracking).

task	F	p
N	4.26	< 0.01
T	1.62	> 0.1
C	37.63	< 0.001

Table 5: Results of one-way analyses of variance on the average error across conditions for each task. Conditions containing the task specified are compared to each other. The result of N for instance is a one-way ANOVA comparing the time spent in the conditions N, NT, NC and NTC.

Table 5 shows that the number of errors made per second significantly differs for the tasks 1-back ($F(3, 408) = 4.26$, $p \approx 0.006$) and tone counting ($F(3, 419) = 37.63$, $p < 0.001$). The difference in heights of the bars for these tasks is thus significant. This does not hold for dot tracking though ($F(3, 464) = 1.62$, $p \approx 0.184$).

3.3 Questionnaire

The top answers to a few questions is displayed in Table 6. A complete overview is included in the appendix (Table 7, page 19).

Participants considered dot tracking the easiest (single-)task (15 out of 21). Most participants showed a preference for the optimal condition, the combination tracking and tone counting (15 out of 21). Most also had the greatest aversion for the worst condition, the combination 1-back and tone counting (11 out of 21). Lastly, a point maximization strategy was employed most often (15 out of 21).

4 Discussion

4.1 Multitasking choices

The difference between the slopes of the conditions (Table 2) indicates that people's preference for combinations develops over time. This is an indication that people adapt to interference as suggested by Nijboer et al. (2013): people start to prefer different conditions, because the smaller interference of these makes them easier to combine than the conditions they chose first. However, no

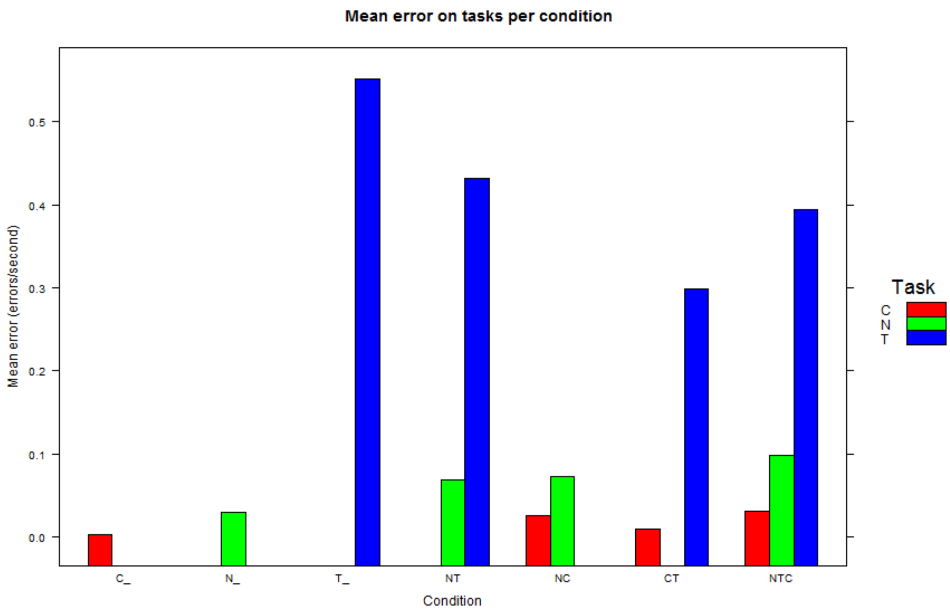


Figure 4: Decomposition of the mean error rate (errors/second) of a certain task in a condition.

question	top answer	%
What task do you think was easiest?	tracking	68
Which combination did you prefer most?	tracking and tone counting	68
Which combination did you try to avoid most?	1-back and tone counting	52
What strategy did you employ during the experiment?	maximizing points	68

Table 6: The top answers to some of the questions on the questionnaire. The last column (%) contains the percentage of participants that gave that specific answer (out of 21).

differences between the slopes of dual-tasks are recognized, suggesting that the change of preferences does not lie in the severity of interference in task combinations. People do not necessarily change their preference for combinations, rather for the number of tasks combined: the differences were between the single- and dual-tasks (and triple-task), rather than within them.

That dual-tasks are performed more often than single-tasks (Table 3) shows that participants preferred multitasking conditions over single-tasks, which might be a consequence of the score system. Another explanation is that the single-tasks were simply not engaging enough. There does not seem to be an indication that the triple task is done more often than dual tasks, but this might be because it is similar to the combination tracking and tone counting (in time spent at least). Looking back at figure 3 the difference between these two conditions seems to be how often they are chosen over time: NTC starts low, but ends high, while CT starts highest and decreases a little toward the end of the experiment. The time spent in conditions is comparable to the area underneath a curve, so this measure can become high for several reasons: the curve is steady (horizontal) starting high and ending high, the curve starts low but becomes very high or it starts very high and ends low. Therefore, the total time spent in conditions (considered as preference for conditions) seems not to be a reliable measurement for participants' overall preference for one of these conditions. If the slopes cross, as in the case of these two conditions, one condition was preferred at the start while the other was preferred at the end.

The theory of threaded cognition predicts that dot tracking and tone counting together is the dual-task with the least interference in this experiment. Looking at figure 3, this combination seems to have the highest precedence of all dual-tasks and it seems to decrease the least. This suggests that participants show more preference for this condition in the experiment than for the other dual-task conditions. This can also be explained by the theory, as for instance 1-back and tone counting together introduces interference as both call on memory, while dot tracking and tone counting have no overlap in cognitive modules. This preference is also supported by the fact that 15 subjects expressed their preference for dot tracking and tone counting in the questionnaire. Moreover, the majority of the subjects ($n = 11$) dislikes 1-back and tone counting most of all combinations (Table 6).

The statistical results do not fully agree with this though: no significant differences were found between the slopes of the dual-tasks (Table 2). This contradicts the expectation that the optimal combination (dot tracking and tone counting together) undergoes the least decrease. The t-test (Table 3) between the best (CT) and worst (NC) dual-task combinations predicted by threaded cognition does show that significantly more time is spent in the best than the worst. This confirms the expectation that the optimal combination has the highest prevalence.

The hypothesis that people will converge to a strategy with the least interference is thus partially true. People immediately have a preference for the optimal condition, rather than gaining preference over time.

The triple-task is odd in this respect. The increase in triple-taskers is not readily explained by the theory of threaded cognition. Compared to any dual-task in this experiment, doing these three tasks obviously requires all modules of that dual-task plus those required by the remaining task. This can only lead to more interference. Since that is the case, it would be expected that people try to avoid this combination.

A reasonable explanation can be given however if this increase is attributed to the score system. Since this combination yields the most points per second (at least .25 point more compared to dual-tasking), it may have encouraged participants to engage in this condition anyway. Provided a participant wants to maximize his or her score, it is reasonable to assume they will at least try this combination, with the possibility of taking it on more often if it seems to be advantageous to their score. On top of the score calculation itself, especially the list of high scores presented at the start may have triggered them to utilize this strategy. If this reasoning is true, it is not unthinkable that the displayed increase in the triple-task comes from the point maximization strategy, because fifteen subjects indicated that they utilized this strategy.

In the study by Nijboer et al. (2013), one of the main problems was that the participants did not discover the optimal combination, because they did not switch often enough. In this experiment, the score system was included to encourage subjects to also switch tasks, in order to find the combination where they got as much points as possible. Looking at the prevalence of the triple-task, it can be defended that this system proved to be effective, persuading people to do their best multitasking. If this is all true, it offers yet another account to intrinsic motivation (Bnabou & Tirole, 2003). Apparently, such a score system is able to intrinsically motivate people to try their hardest, without the need of other incentives.

4.2 Performance

The performance of participants on conditions shows an unexpected discrepancy. Tracking is considered the 'easiest' task, as 15 participants indicated (Table 6). This is however also the task with the greatest average error rate. On closer inspection this seems to be caused by the system of penalties. An 'error' is just defined as some mistake and their criteria are defined in section 2. There is however not a one on one relation with the points deducted. The penalty applied actually depends on the task for which the error is made. Also stated in section 2 is that these penalties are determined in such a way that someone's score balances around zero points if they randomly give their responses. As such, it is not strange that most errors are recorded per second in the easiest task, as the criterion of an error is easily satisfied (one second outside boundaries). As a compensation, the severity of the penalty applied is mild (one point deducted per error). On the contrary, tone-counting has an error criterion that might take quite a while to fulfil (did not press at a count of five), but then the penalty is rather severe (three points per error). This puts the performance results of Table 4 and Figure 4 in perspective.

Table 4 yields a better perspective into how the performance during conditions compare to each other. The mean error on tasks is highest in the triple-task and higher in the dual-tasks than the single-tasks. This points out that people experience more difficulty (thus make more errors) doing more tasks. Nevertheless the error rate on dot tracking is odd in these respects. When it is done alone, most errors are made. Since the single-tasks are not performed much though, these numbers may not be entirely reliable. Also peculiar is the fact that the error on tracking is next highest in condition NT, rather than NTC. The most reasonable explanation for this phenomenon is that participants prioritize tracking more in the triple-task than in the dual-task. It might also be a result of practice. Since triple-tasking is done more often in the end, the practice people had beforehand may be beneficial for their proficiency on the tasks. Another possibility is that the triple-taskers are simply more capable of combining tasks, so this result just shows differences between dual- and triple-taskers.

4.3 Implications

A less welcome implication of this observation is that people might be persuaded to take on tasks that do not necessarily combine well. This notion can be connected to utility theory (Fishburn, 1970). The advantage of triple-tasking in this experiment is that a participants' score rises as much as possible, provided one does not make many more errors. This option can thus have a higher utility than combining 'easier' tasks. This effect may seem harmless at first, but can have major impacts in society. For instance the driving examples stated at the beginning of this article: driving while handling a cellphone can lead to dangerous situations. Although it is good to know that certain multitasking conditions are detrimental, it will not necessarily stop people from multitasking as for some reason the utility of doing this concurrently seems greater at the moment. A different situation is on the work-floor. Deadlines are known to motivate people for tasks (Amabile, DeJong, & Lepper, 1976). This is also used to increase efficacy of a company. Deadlines however just prioritize tasks that should be finished first, not necessarily those that combine best. This consequence may actually prove to be a bad thing for efficiency. When deadlines force someone into multitasking, this may make them combine jobs that are not done together easily, probably resulting in making more errors in the process. In the best case, this just results in an increase in completion time of the jobs compared to when these are done separately or in another combination. In the worst case however, this may lead to more resources being wasted in the form of faulty, or even hazardous products.

Another implication of these findings is that people might just immediately have a feel for what is best to combine, rather than to learn this over time. This can be either positive or negative. It is positive if the combination they choose is indeed optimal in some sense. In that case people indeed optimally employ their resources. If however the combination is sub-optimal at best, resources could be utilized more efficiently. This is still not bad, unless people just stick to their current preference, in spite of other evidence. This way people will never be optimally effective, no matter how much time and educational opportunities have passed.

4.4 Impurities

Certain aspects of this experiment can be revised. First of all, there might be a selection bias in the pool of participants. All subjects are students and almost all studied Artificial Intelligence.

Another thing to note is that the experiment conditions were not always optimal. Since the lab has to be shared with other experimenters and subjects, some participants may have experienced hinder from others.

Besides that, equipment failure such as an (almost) empty battery or a malfunctioning keyboard probably are major distractions. In these cases it was possible to continue data collection, nevertheless participants could be 'off' for at least a small period of time.

As noted, the error rate measured is not a good indication of performance on conditions. For this reason, an assessment of the difficulty of each combination compared to each other is in place. As the threaded cognition theory predicts dot tracking and tone counting combined has the least interference of all, this should be traceable in the performance of participants. Further examination of this is thus justified.

4.5 Further topics

Further research accounting for the aforementioned problems is thus justified. To ensure that these findings hold generally, a more diverse pool of participants is recommended. Besides that, it may be interesting to know what effect a less noisy, or even a noisier environment, has on multitasking performance.

It is also interesting to know to what extent these conclusions hold in real-life settings. One of the implications discussed is the possibility of people making the wrong decisions given certain incentives. The consequences can be quite profound and thus require verification.

Besides that, the subjects know they are part of an experiment. This fact might have influenced their behaviour (Aarts & Dijksterhuis, 2003). Again a verification of these findings in a more natural setting is in place.

Although there does not seem to be significant differences between participants' preferences over time for the dual-tasks (as indicated by the slopes), Figure 3 seems to suggest otherwise. The increasing popularity of the triple-task may have obscured such differences. Therefore, an experiment with probably a slightly different set-up could examine this further. The outcome would be very interesting as this may confirm or contradict that people are able to learn what the optimal combination is.

References

Aarts, H., & Dijksterhuis, A. (2003). The silence of the library: environment, situational norm, and social behavior. *Journal of personality and social psychology*, *84*(1), 18.

Amabile, T. M., DeJong, W., & Lepper, M. R. (1976). Effects of externally imposed deadlines on subsequent intrinsic motivation. *Journal of personality and social psychology*, *34*(1), 92.

Bnabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, *70*(3), 489-520. doi: 10.1111/1467-937X.00253

Borst, J. P., Taatgen, N. A., Stocco, A., & van Rijn, H. (2010, 09). The neural correlates of problem states: Testing fmri predictions of a computational model of multitasking. *PLoS ONE*, *5*(9), e12966. doi: 10.1371/journal.pone.0012966

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436.

Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron*, *63*(1), 127–138.

Fishburn, P. C. (1970). *Utility theory for decision making* (Tech. Rep.). DTIC Document.

Konig, C. J., Buhner, M., & Murling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance*, *18*(3), 243-266. doi: 10.1207/s15327043hup1803_3

Nijboer, M., Taatgen, N. A., Brands, A., Borst, J. P., & van Rijn, H. (2013, 11). Decision making in concurrent multitasking: Do people adapt to task interference? *PLoS ONE*, *8*(11), e79583. doi: 10.1371/journal.pone.0079583

Salvucci, D., & Taatgen, N. (2010). *The multitasking mind.* Oxford University Press, USA.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological review*, *115*(1), 101.

Taatgen, N. A., Juvina, I., Schipper, M., Borst, J. P., & Martens, S. (2009). Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive Psychology*, *59*(1), 1 - 29. doi: http://dx.doi.org/10.1016/j.cogpsych.2008.12.002

Watson, J. M., & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. *Psychonomic Bulletin & Review*, *17*(4), 479-485. doi: 10.3758/PBR.17.4.479

A Briefing

Instructions

1 Introduction

Note: Instead of reading the instructions on this paper, it is also possible to just follow the instructions that accompany the experiment.

The purpose of this experiment is to investigate how good humans are at multitasking. It is primarily aimed at task preference in multitasking conditions. For this purpose you may do three tasks (see section 4 for further instructions). You are free to switch between combinations of these at any time.

2 High scores

Before continuing, here is the top five of high scores of other participants:

#	Score
1	4782
2	4704
3	4619
4	4316
5	4107

Just try and beat that!

3 Score system

3.1 gain

Initially, you can earn the same amount of points for every task, so it does not matter which task you choose at first. However, the longer you perform a certain task the less points you gain from it. The tasks you are not performing however will gain you more and more points. Therefore, it might be wise to switch tasks every now and then.

The display containing your score will also change color based on how much points are gained (green is good, red is bad). This should give an indication on how fast your score is increasing.

3.2 penalty

If you make mistakes, a penalty is deducted from your score. Try to perform the tasks at your very best to avoid nasty penalties!

4 Tasks

4.1 1-back

A stream of letters is presented, masked with a #. You have to indicate whether the current one is the same as the last you saw (press E for yes, R for no).

4.2 tracking

You have to keep a moving dot inside a circle. With U and I you can move the circle left and right respectively.

4.3 tone-counting

Randomly either a high or a low tone is played. Count the number of high tones and press the spacebar as soon as you hear the fifth high tone.

5 Switching

You may switch at any time by pressing P. This will bring up a task selection screen where you can select which tasks to do next by pressing E, R or U.

6 Good luck

There is not much left to say. Good luck in the experiment!

B Questionnaire

Debriefing

1. Finishing up

Before leaving, would you like to answer a few questions about the experiment? In the mean time I will fetch your reward. Please tick only one box per question. If you make a mistake, place a circle around the answer you intended.

Name:
Date:
male / female

- Which task do you think was easiest?
  - ☐ 1-back
  - ☐ tracking
  - ☐ tone counting
  - ☐ none specifically

- Which task do you think was hardest?
  - ☐ 1-back
  - ☐ tracking
  - ☐ tone counting
  - ☐ none specifically

- Which combination did you prefer most?
  - ☐ 1-back and tracking
  - ☐ 1-back and tone counting
  - ☐ tracking and tone counting
  - ☐ all three together
  - ☐ none specifically
  - ☐ otherwise, namely …

- Which combination did you try to avoid most?
  - ☐ 1-back and tracking
  - ☐ 1-back and tone counting
  - ☐ tracking and tone counting
  - ☐ all three together
  - ☐ none specifically
  - ☐ otherwise, namely …

- *1-back:* Did you use mental rehearsal to remember the letters?
  - ☐ yes
  - ☐ no
  - ☐ don't know

- *1-back:* Did you visualize the letters, or verbalize them (repeat them)?
  - ☐ visualize
  - ☐ verbalize
  - ☐ both
  - ☐ otherwise, namely …
  - ☐ don't know

- *Tracking:* Did you anticipate movement of the dot?
  - ☐ yes
  - ☐ no
  - ☐ don't know

- *Tone counting:* Did you use mental rehearsal to keep track of the count?
  - ☐ yes
  - ☐ no
  - ☐ don't know

- What strategy did you employ during the experiment?
  - ☐ maximizing points
  - ☐ making it easy for myself
  - ☐ making it hard for myself
  - ☐ otherwise, namely …
  - ☐ none specifically

- Did you enjoy the experiment?
  - ☐ yes
  - ☐ no
  - ☐ don't know
  - ☐ Gosh, did you really just ask this?

- Any further remarks?

  ………
  ………
  ………
  ………
  ………
  ………

2. Expectation

As stated, we are investigating how good humans are at multitasking. When it comes to task preference we expect that tone counting and tracking go together best. These tasks have the least overlap in cognitive modules compared to the other combinations, so they should be easiest to do together.

The score system was put in place to encourage you to switch tasks and as such increase the probability of finding this optimal combination. To answer the question whether humans are good at multitasking, the combinations done over time are compared. We think humans are capable of finding this solution, so they will perform this combination more often later on.

3. Thanks again!

Thank you for your participation and have a nice day!

question	answer	n
What task do you think was easiest?	1-back	2
	tracking	15
	tone counting	2
	none specifically	2
What task do you think was hardest?	1-back	9
	tracking	4
	tone counting	7
	none specifically	1
Which combination did you prefer most?	1-back and tracking	2
	1-back and tone counting	0
	tracking and tone counting	15
	all three together	4
	none specifically	0
	otherwise, namely ...	0
Which combination did you try to avoid most?	1-back and tracking	3
	1-back and tone counting	11
	tracking and tone counting	0
	all three together	5
	none specifically	2
	otherwise, namely ...	0
1-back: Did you use mental rehearsal to remember the letters?	yes	11
	no	8
	don't know	2
1-back Did you visualize the letters, or verbalize them (repeat them)?	visualize	6
	verbalize	8
	both	5
	otherwise, namely ...	0
	don't know	2
Tracking: Did you anticipate movement of the dot?	yes	19
	no	2
	don't know	0
Tone counting: Did you use mental rehearsal to keep track of the count?	yes	19
	no	2
	don't know	0
What strategy did you employ during the experiment?	maximizing points	15
	making it easy for myself	2
	making it hard for myself	2
	otherwise namely combination of points, easy and hard, just a bit of different things combined	1
	none specifically	1
Did you enjoy the experiment?	yes	11
	no	2
	don't know	2
	Gosh, did you really just ask this?	6

Table 7: The given answers to the questions posed in the questionnaire. n is the amount of participants that gave that specific answer (out of 21).