Predicting Interference in Concurrent Multitasking

Menno Nijboer (m.nijboer@rug.nl)[1]
Jelmer P. Borst (jelmer@cmu.edu)[1,2]
Hedderik van Rijn (d.h.van.rijn@rug.nl) [3]
Niels A. Taatgen (niels@ai.rug.nl)[1]
[1]Department of Artificial Intelligence, University of Groningen
[2]Department of Psychology, Carnegie Mellon University
[3]Department of Experimental Psychology, University of Groningen

Abstract

Investigations into multitasking have shown that overlap in cognitive resource use between tasks leads to lower performance. Various computational models have been developed to explain this phenomenon. However, most concurrent multitasking models have been tested in dual-task situations only. In order to see if single-task models could explain dual-task performance, we developed a computational cognitive model that was fit on single-task performance. The model was used to generate a priori predictions of behavior for an experiment in which three tasks with various resource requirements were performed in isolation and in combination. The model predictions closely matched the behavioral data, indicating that models designed for single tasks can indeed account for performance degradation in dual-task situations. To further validate the model, a prediction of neuroimaging data in six regions of interest was generated, which was tested in an fMRI experiment. We achieved a partial fit for the neuroimaging data, indicating that some aspects of the BOLD response require further modeling.

Keywords: Multitasking; ACT-R; fMRI; Interference.

Introduction

Our society is one of avid multitaskers. We perform many actions concurrently, often without even realizing it. Multitasking is typically used to indicate the simultaneous performance of high-level tasks, such as driving a car while setting up the navigation software. What we often forget is that even something as simple as walking down the street while counting lampposts is also a valid example of multitasking. These tasks might not register as 'real' multitasking because we can perform them together easily.

But what if, during counting, your friend asks you how much money you owe him for lunch? Suddenly the situation becomes more complicated: now you have two activities, counting and algebra, which partly require the same cognitive resources. It would therefore be hard to do both activities at the same time. The interference in this example is probably caused by the overlap in the required resources of both tasks: a notion that was formalized in theories of multitasking (e.g., Pashler, 1994; Wickens, 2004). A recent theory, threaded cognition, also implemented those ideas as a computer simulation (Salvucci & Taatgen, 2008, 2011).

In threaded cognition each task is called a 'thread', several threads can be active at the same time. Each task can make use of several resources, such as vision or declarative memory. While these resources function in parallel, access to a resource is serial: if one task is using a resource, other tasks cannot use it. For instance, if a task requests a fact from memory, and the memory resource is already retrieving something for another task, the first task will be postponed until the current retrieval is complete.

Salvucci and Taatgen (2008, 2011) showed that a number of resources could lead to interference. These include obvious candidates such as perceptual and motor resources, but also declarative and procedural memory. Furthermore, the problem state, a one element working memory resource, was also found to be a source of significant interference.

Concurrent multitasking models using threaded cognition have focused almost solely on dual-task performance. In this paper we investigate whether task models that have been modeled for single-task performance can account for performance degradation, as well as neuroimaging data, when combined in dual-tasks. To this end, we designed an experiment in which three tasks with different resource requirements had to be performed alone or in pairs. We developed a computational cognitive model of these tasks, and generated a priori behavioral predictions, which were tested in the experiment. To further validate our results, we used the model to predict activation patterns in certain brain regions and tested these predictions in an fMRI experiment.

We will now first discuss the experiment, followed by the model and behavioral results. Afterwards, we will present the neuroimaging predictions and results.

Methods

Design

The experiment consisted of three tasks: visual tracking, tone counting, and n-back. Subjects were asked to perform the tasks both in isolation as well as in combination with the other tasks. The tasks were chosen in such a way that some combinations had mild or significant resource overlap, while other combinations did not overlap at all.

N-back: in the 2-back task a stream of letters was presented on screen. Each letter was shown for 1000 ms, followed by an inter-stimulus interval of 1500 ms. For each letter, the participant has to press a key with their left hand to indicate whether the letter was the same or different as two letters back. A response could be given up to 1500 ms after the letter was presented. The 2-back task was expected to require working memory and declarative memory, and to a lesser degree the manual and visual resources.

Tracking: during the tracking task participants were asked to track a moving target using their right hand. The target only moved in a horizontal direction: keys for left and right were used to track it. Two green lines on either side of the target conveyed the maximal distance the cursor may be removed from the target. If the cursor crossed this threshold, the lines turn red. The tracking task uses the manual and visual resources.

Tone-counting: the final task was tone-counting, in which participants were presented with 20 tones at pseudo-random intervals. Tones could be either high (493.88 Hz) or low (261.63 Hz); participants had to count the high tones. The number of high tones in a trial varied randomly between 10 and 17, with the rest filled up by low tones. At the end of a trial an answer prompt was presented. The participant had 10 seconds to enter a response, which was given by pressing a key to increment the digits shown on screen. If tone-counting was combined with tracking the left hand was used to respond, in all other cases the right hand. Afterwards, feedback was given for 500 ms. The main resources used by the counting task are the aural, working memory, and declarative memory resources.

Participants

28 students participated in the behavioral experiment for monetary compensation. Because one participant did not follow instructions, and three performed the 2-back single-task at chance level, 24 participants were left for analysis (16 female, mean age 23, age range 19-30). Informed consent as approved by the Ethical Committee Psychology of the University of Groningen was obtained before testing.

Procedure

At the start of each trial an eight second fixation cross was followed by a two second presentation of the names of the tasks in the upcoming trial. In the single-task conditions, the task stimulus was presented in the middle of the screen. As no visual stimuli are used in the tone counting task, a fixation cross was presented instead. In the dual-task conditions the 2-back task was always presented on the left side of the screen, and the tracking task was always presented on the right side. The tone counting fixation cross was presented on whichever side was still empty. Each trial lasted 30 seconds, with an additional 10 seconds for trials with tone-counting for entering the response.

The experiment consisted of one practice block and two experimental blocks. In the practice block participants performed each single-task condition twice. The two experimental blocks consisted of 31 trials each, which were presented in a pseudo-random order: each condition appeared every six trials, resulting in 10 trials per condition. Every block also contained a fixation trial: during this trial participants did not need to perform a task.

Predicting Interference

To predict the performance in the experiment we developed a computational model in the ACT-R cognitive architecture (Anderson, 2007), which was also used to instantiate the threaded cognition theory (Salvucci & Taatgen, 2008). ACT-R consists of a set of resources that can be operated through production rules. The resources act in parallel, but each resource itself proceeds serially (Byrne & Anderson, 2001). The framework consists of peripheral resources (visual, manual, aural, and vocal), as well as cognitive resources (declarative memory, goal, and problem state). The problem state, or imaginal module, represents working memory in ACT-R, and can contain a single chunk of information. Threaded cognition adds the ability to ACT-R to run multiple goals, or tasks, at the same time by interleaving the production rules belonging to each goal.

In our model, each task was implemented as a separate thread. This means that the same task models were used in both the single-task and dual-task setup: dual-task performance was therefore a parameter-free prediction of the single-task models.

In the 2-back task, the model perceives the stream of letters, and maintains a problem state that contains the previous letter and a reference to the 2-back letter in declarative memory. When the model attends a new letter, the 2-back state is retrieved from declarative memory and compared against this new letter. The model then gives a response with its left hand for 'same' or 'different', depending on the retrieved letter. If the 2-back could not be retrieved because its activation fell below the retrieval threshold, the model guesses a response. Following the response, the current problem state is stored in memory – effectively storing the letter 2-back – and a new problem state is made that contains the attended letter and a reference to the old problem state.

When performing the tracking task, the model first perceived the positions of the cursor and the moving target dot. While tracking these positions, the distance between the target and cursor is observed. If this distance becomes too large, the model presses the key required to reposition the cursor on top of the target, using its right hand. Movements of the cursor will be rapid when it is far removed from the target, but slow when it is close to the target. This allows the model to make precise adjustments.

Finally, when the model is counting tones, it stores the count in a problem state. The model hears both high and low tones, but does not count the low tones. After a high tone is perceived, the model increments the current count n by retrieving $n+1$ from declarative memory. Sometimes a tone is heard while the current count is being incremented. This fires a production rule that stores the new tone in memory. Once the increment is completed, the tone is retrieved from memory to determine if it was low or high, and if needed a new increment operation is started.

To interleave most resources in ACT-R, each task must check if a resource is busy and empty. Using the problem state resource is slightly more complicated. While the problem state itself can be checked whether it is busy, this does not provide any information about another task that might be using the problem state contents over the span of several production rules. As all threads have direct control of the contents of the problem state resource, overwriting its contents can cause problems: one thread might overwrite the problem state, causing another thread to be stuck forever as it requires its own problem state to be in the buffer. Therefore, problem states must be switched explicitly: at any step where the problem state is required the task must check if it has a state of the right task. If not, it retrieves a usable state, and replaces the current one. This is the only part of the models that would not ordinarily be required when just modeling single tasks. However, one could argue that a robust model would include it nonetheless.

A Priori Performance Predictions

The three task-models indicate that several resources could overlap and cause interference in the dual-task conditions. When 2-back is combined with tracking visual requirements overlap, as attention has to be divided between the letters on the left half of the screen and the tracking dot on the right. Although both tasks use the manual resource, there is no real overlap because the tasks use different hands.

In the 2-back and tone-counting condition, there is overlap in memory resources: both tasks require declarative memory and the problem state resource. The 2-back task requires the problem state to keep track of the last few letters, and needs declarative memory to retrieve the 2-back. The tone-counting task requires the problem state to keep track of the number of high tones, and declarative memory to determine the next increment.

Finally, in the tracking and tone counting condition there is no apparent overlap, as both tasks use different sensory modalities, only one task requires working memory, and only one task makes use of the manual resource during the trial.

Using our computational model we generated quantitative predictions of single-task performance and dual-task interference. These predictions are shown in Figure 1 (yellow squares)[1]. The largest overall performance loss was in the 2-back & tone-counting conditions, followed by 2-back & tracking. Both tracking and tone-counting performance suffer when combined with the 2-back task, but 2-back performance itself suffers most when combined with tone-counting. Hardly any performance difference is seen in the tracking & tone-counting condition.

Results

All *F*- and *p*-values are from repeated measure ANOVAs, and all accuracy data were transformed with a logit transformation before performing ANOVAs. All error bars depict 95% confidence intervals.

The white circles in Figure 1 show the error rate for each task (panels A, B, and D). An ANOVA showed a main effect for condition in the 2-back task accuracy ($F(2, 48) = 56.15$, $p < 0.001$). A *post-hoc* Tukey's test reveals that the difference between 2-back & tracking and 2-back & tone-counting was slightly less significant than the other

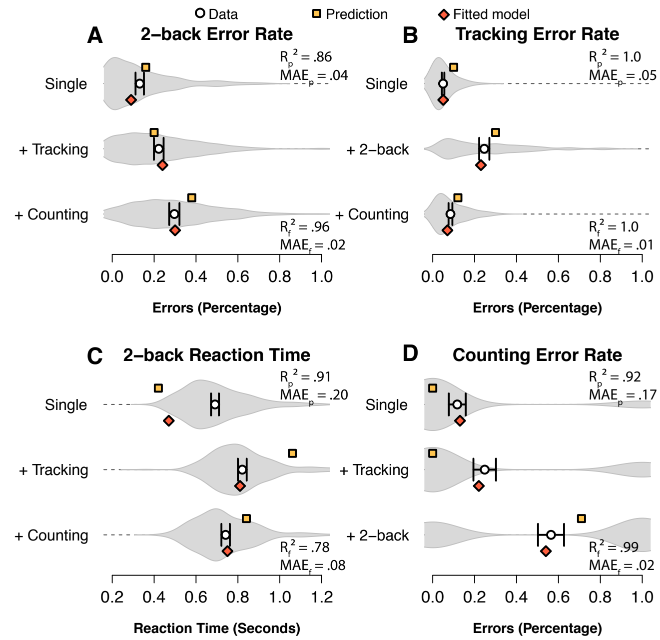

Figure 1. The behavioral results for each task. R_p^2 is the fit to the results of the a priori prediction, whereas R_f^2 is the fit for the posteriori model. In addition, the mean absolute error (MAE) is given. The gray background represents density estimates of the underlying data.

differences ($p < 0.01$ versus *p*s < 0.001). Performance is clearly worse in dual-task conditions, with 2-back & tone counting leading to the lowest performance, as predicted. Reaction times (panel C) for the 2-back did not mirror the accuracy data: an ANOVA showed a main effect for condition ($F(2, 48) = 31.34$, $p < 0.001$) with the single-task condition leading to the fastest times and the 2-back & tracking condition leading to the slowest times. The model predicted this pattern.

Tracking accuracy showed a main effect of condition ($F(2, 48) = 105.1$, $p < 0.001$): A *post-hoc* Tukey's test shows that all conditions were significantly different from each other (*p*s < 0.001). As seen in panel B, the accuracy only decreased by a meaningful amount when combined with 2-back. The model captures the same effect, but predicts slightly lower overall accuracy.

Finally, tone counting accuracy also shows a main effect for condition ($F(2, 48) = 50.64$, $p < 0.001$). Again, all conditions differ significantly from each other, with the difference between tone counting and tone-counting & tracking being slightly less significant than the other comparisons ($p < 0.01$ versus *p*s < 0.001). The data reveals two unexpected effects: in the tone-counting + tracking condition the accuracy is lower compared to the single-task condition. It seems there is still some interference, even though the tasks do not seem to share any resources. The condition with the lowest accuracy is clearly 2-back & tone-counting, as predicted by the model. However, the model overestimated performance in the single-task and tracking conditions, and underestimated performance in the 2-back condition.

[1] This prediction was generated before we ran the experiment; it was posted on the ACT-R mailing list on October 29, 2012.

Posteriori Model Fit

We adapted the model to fit the data of the experiment. To capture the low performance on the tone-counting single-task, the model was altered to sometimes mistake a low tone for a high tone. This was accomplished by adding an extra production rule that competes with the correct rule for handling low tones. In addition, we fitted the model parameters to create a better quantitative fit.

The red diamonds in Figure 1 show that the fitted model has a good quantitative fit to the data. Even the unexpected effect in the tracking & tone-counting condition is captured. In model terms, the explanation for this effect is that concentrated effort on the tracking task can sometimes cause tones to expire in the auditory buffer before they have a chance to be processed by the counting task: the tones are heard, but not consciously processed.

A Priori BOLD Predictions

As the cognitive resources used by the model have been mapped onto brain regions (e.g., Anderson, 2007), it is possible to generate predictions of the Blood-Oxygenation Level-Dependent (BOLD) response for those regions. The predictions were generated by convolving resource activity with the hemodynamic response function, and can be compared against the neuroimaging data to provide further validation of the model. For a detailed overview of the method see Anderson (2007) and Borst, Taatgen, Stocco, & Van Rijn (2010).

Using the fitted model we computed BOLD functions for five resources: the visual resource, the manual resource, the aural resource, the declarative memory resource, and the problem state resource. We split the manual resource prediction into left and right, as each task had different manual requirements. The predictions are presented in Figure 2 (odd rows)[2]. Each panel shows the predicted BOLD response over one trial: each scan is 2 seconds.

The aural prediction is presented in Figure 2A. It shows the same amount of activity for all three tone-counting conditions. This is a consequence of tone-counting being the only task that uses auditory cues. As such, no activity is predicted for any of the remaining conditions.

The prediction for the visual resource is shown in Figure 2B: the model predicts the highest activities for conditions that include the tracking task. Less activation is predicted for the remaining conditions that include the 2-back, and for the tone-counting single-task condition activation is only predicted at the end of the trial, when the answer needs to be entered. We see a similar spike in the 2-back & counting condition. These predictions align with our intuition: more visual information on the screen results in more visual resource activity. One might expect the 2-back & tracking condition to result in the highest activation, given that both tasks have a strong visual component. However, the model does not reflect this. The reason is that the time spent switching between the two tasks does not result in activity in the visual resource[3].

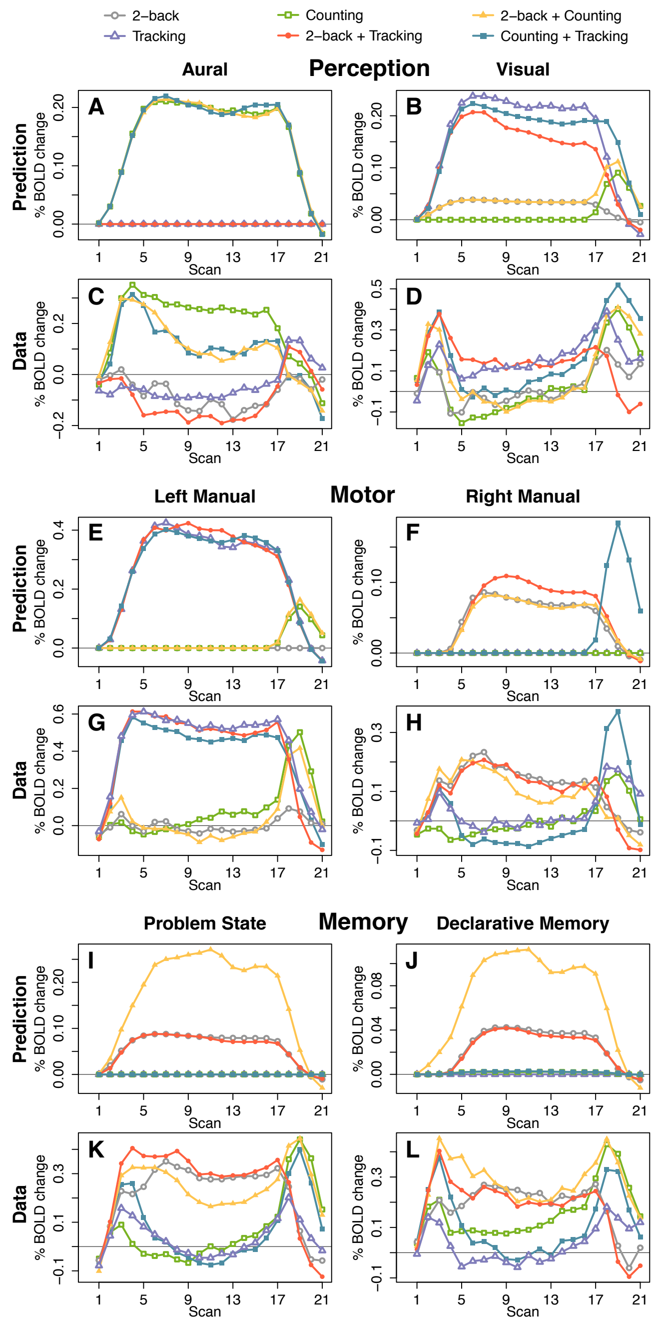

Figure 2. BOLD response prediction (odd rows) and neuroimaging results (even rows) for six ROIs.

Figure 2E and 2F show the BOLD predictions for the left and right manual resources. The left manual prediction is similar to the visual prediction: it shows large activations for all three tracking conditions. This is expected, as the right hand is used for the tracking task. The activation at the end of the tone-counting and 2-back & single tone-counting

[2] This prediction was also posted on the ACT-R mailing list on January 16, 2013.

[3] It does result in activity in the visual-location resource, which we do not cover, as there is no mapping of that resource to a specific brain region in ACT-R.

trials is present as well: the right hand is used to enter the tone count in these conditions. As 2-back is performed with the left hand, it shows no activation in the single-task condition.

All 2-back conditions show a clear response in the right manual prediction. The magnitude of the BOLD change is smaller than seen in the left manual prediction, as the 2-back task requires much less manual input than the tracking task. The activation at the end of the trial in the tone-counting & tracking condition is due to the tone counting response being made with the left hand in that condition.

Figure 2I shows the problem state prediction. The 2-back & tone-counting condition generates by far the largest BOLD response during the trial. This is due to the interference between the tasks: Both tasks require a problem state, and given that only one state can be active at any particular time, states are switched in and out of the resource repeatedly. This switching means that the problem state resource is very active, which results in the large BOLD response. The remaining 2-back conditions show an intermediate level of activation: even without an interfering task, the 2-back is still quite taxing on memory. When a new letter has to be stored, the old problem state is put in declarative memory, and a new state is created, which causes the predicted activity. The tone-counting single task shows a low level of activation: the problem state is updated during a trial, but not replaced with a new state. As such, it is less intensive than the 2-back task. As tracking does not use the problem state, it shows no activation.

The declarative memory resource prediction, shown in Figure 2J, shows a pattern very similar to the problem state prediction. This makes sense, as declarative memory is primarily used to retrieve old problem states from memory.

fMRI Experiment

Eighteen people participated in the experiment for monetary compensation. Two participants were removed due to excessive head motion, leaving 16 datasets for analysis (4 male, mean age 22, age range 18-25, right-handed). Written consent as approved by the Medical Ethical Committee of the University Medical Center Groningen was given before the experiment.

Before the scanning session participants performed a 20-minute practice session with the paradigm outside of the scanner. The scanning session starts with a structural scan. During this scan participants perform a 10-minute practice block, to become accustomed to performing the tasks inside the scanner. Afterwards, participants performed six 10-minute blocks. Each block contained each condition twice, plus one fixation trial, for a total of 78 trials. The trial order within each block was randomized. The remaining details of the paradigm are identical to the behavioral experiment described earlier.

fMRI Procedures and Preprocessing

The neuroimaging data were obtained with a Philips Interna 3 Tesla scanner using a standard radio frequency head coil. Each functional volume consisted of 37 axial slices (3.5 mm thickness, 64x64 matrix, 3.5 mm x 3.5 mm per voxel), acquired using echo-planar imaging (2000 ms TR, 20 ms TE, 70° flip angle, 224 mm field of view, 0 mm slice gap, with AC-PC on the 19th slice from the bottom). Anatomical images were acquired using a T1-weighted spin-echo pulse sequence with the same parameters as the functional images, but with a higher resolution (1 mm thickness, 256x256 matrix, 1 mm x 1 mm per voxel).

The data were preprocessed using SPM8[4]. The steps consisted of realigning the functional images, coregistering them with the structural images, normalizing the images to the MNI (Montreal Neurological Institute) ICBM 152 template, and smoothing them with an 8 mm FWHM Gaussian kernel.

Imaging Data Analysis

The results of a region of interest (ROI) analysis performed on the imaging data are presented in Figure 2 (even rows). The location of the regions is based on a recent meta-study (Borst & Anderson, 2013), instead of the mapping provided by ACT-R (see Borst, Nijboer, Taatgen, & Anderson, submitted, for details). Results were averaged over the left and right regions, with the exception of the manual regions. The aural resource (left superior temporal gyrus), visible in Figure 3C, only showed activation for the conditions that include the tone-counting task. While the model predicted identical activation patterns for all three tone-counting conditions, there was clearly less activity in this region while the dual-tasks were performed.

The results for the visual resource (left middle occipital gyrus) are presented in Figure 2D. Apart from the large initial spike, the results partially resemble the model prediction: the three tracking conditions show the highest activation. However, the 2-back conditions show almost no activation. It is therefore likely that the visual ROI is more related to spatial attention than the detailed visual processing required in parsing the 2-back letters shown on the screen. Further evidence of this is that all the conditions that include tone-counting show a spike at the end of the trial, when the tone count input screen is presented. Hence, all the instances that require significant spatial attention show high activity in the visual ROI.

Figure 2G and 2H show the results of the manual resource region (the left and right precentral gyrus, respectively). The left region shows large, near identical activation for all three conditions that involve tracking. The right region shows almost identical activation patterns for the conditions that include the 2-back task. The figure also shows deactivation in the tracking single-task and the tracking & tone-counting dual task conditions. The counting single-task shows no activation during the trial.

The problem state resource (left inferior/superior parietal lobule, Figure 2K) showed strong activation in the

[4] Wellcome Trust Centre for Neuroimaging (www.fil.ion.ucl.ac.uk/spm/)

conditions in which the 2-back task was present. In contrast to the model prediction, we did not see an over-additive effect in the 2-back & tone-counting condition. In fact, of the three conditions that show high levels of activation, it is the least active. All conditions showed an activation spike at the start of the trial. This might be due to an attention shift: it has previously been shown that activation in the problem state region can reflect visual processing.

Finally, the declarative memory region (left inferior frontal gyrus) is presented in Figure 2L. The region showed a pattern very similar to the problem state region: the 2-back & tone-counting condition shows somewhat less activation here, and the tone-counting single-task shows slightly more activation.

In summary, the results in the manual regions followed the predictions closely. The prediction for the visual region has the conditions in roughly the right order, but does not capture the differences between the three conditions that contain the tracking task, and did not match predicted n-back activity. Looking at the aural region, the model and the results have a similar overall shape, but the model does not produce the deactivation found in all conditions except the tone-counting single-task. Finally, in the problem state and declarative memory regions the model does not capture the BOLD response for the 2-back & tone-counting condition.

Discussion

We developed a model to investigate interference effects in concurrent multitasking. To test this model, we performed a behavioral and an fMRI experiment. The behavioral predictions of the model were confirmed: the model predicted dual-task performance by combining the single-tasks models, without changing the single-task models. This results in a parsimonious account of concurrent multitasking, which does not rely on any additional systems such as a central executive (Baddeley, 1996; Kieras et al., 2000). Furthermore, having three tasks that show different performance profiles for each condition imposes strong constraints on the model, as the interaction between tasks limits the space of possible solutions.

The neuroimaging predictions were partly confirmed by the data. The greatest discrepancy between model and data occurred in the 2-back & tone-counting condition, which showed a poor fit for the problem state and declarative memory. In addition, the aural regions showed deactivation in the dual-task conditions that contained tone-counting. This might indicate that there is a limit on the amount of blood flow to various regions. As dual-task conditions do not only use certain regions more intensely, but also a greater number of regions, it could be that we are observing a ceiling effect, limiting the increase in the BOLD response. However, this does not explain why 2-back & tone-counting produces less activation than the 2-back single task.

Alternatively, it might be that participants changed their strategy in the dual-task condition. That is, they might have used a less memory-intensive strategy to perform the 2-back task in this condition, based on familiarity and timing (e.g., Iuvina & Taatgen, 2007). This would lead to lower activity in the regions-of-interest, but also to lower behavioral performance, matching our results. Finally, we might simply be looking in the wrong region for evidence that 2-back and tone-counting interfere. In conclusion, while the single-task models gave an excellent fit for dual-task behavioral data, the neuroimaging results pose a greater challenge.

Acknowledgments

This research was funded by ERC-StG grant 283597 awarded to Niels Taatgen.

References

Anderson, J. R. (2007). How can the human mind occur in the physical universe? USA: Oxford University Press.

Baddeley, A. (1996). Exploring the Central Executive. The Quarterly Journal of Experimental Psychology, (1), 5–28.

Borst, J. P., & Anderson, J. R. (2013). Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. Proceedings of the National Academy of Sciences USA, 110(5), 1628–1633.

Borst, J. P., Nijboer, M., Taatgen, N. A., & Anderson, J. R. (submitted). A data-driven mapping of five ACT-R modules on the brain. Proceedings of the 12th International Conference on Cognitive Modeling.

Borst, J. P., Taatgen, N. A., Stocco, A., & Van Rijn, H. (2010). The neural correlates of problem states: testing fMRI predictions of a computational model of multitasking. PloS ONE, 5(9), e12966.

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2011). Using a symbolic process model as input for model-based fMRI analysis: locating the neural correlates of problem state replacements. NeuroImage, 58(1), 137–147.

Byrne, M. D., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. Psychological Review, 108(4), 847–869.

Juvina, I. & Taatgen, N.A. (2007). Modeling control strategies in the N-back task. Proceedings of the eight International Conference on Cognitive Modeling (pp. 73-78). New York: Psychology Press.

Kieras, D. E., Meyer, D. E., Ballas, J., & Lauber, E. J. (2000). Modern computational perspectives on executive mental processes and cognitive control: Where to from here? In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII: Control of Cognitive Processes.* (pp. 681–712). Cambridge, MA: MIT Press.

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

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. Psychological Review, 115(1), 101–130.

Salvucci, D. D., & Taatgen, N. A. (2011). The multitasking mind. New York: Oxford University Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159–177.