

Huff, M., & Schwan, S. (in press). Integrating information from two pictorial animations: complexity and cognitive prerequisites influence performance. *Applied Cognitive Psychology*.

Integrating Information from Two Pictorial Animations: Complexity and Cognitive
Prerequisites Influence Performance

Markus Huff

University of Tübingen, Germany

Stephan Schwan

Knowledge Media Research Center, Tübingen, Germany

Author Note

Markus Huff, Department of Psychology, University of Tübingen, Germany; Stephan Schwan, Knowledge Media Research Center, Tübingen, Germany.

We thank Sandra Alpers for her help in conducting the experiments. We are also grateful to Paul Ayres and two anonymous reviewers for their helpful comments on an earlier version of this paper.

Correspondence concerning this article should be addressed to Markus Huff, Department of Psychology, University of Tübingen, Konrad-Adenauer-Str. 40, 72072 Tübingen, Germany.

E-mail: markus.huff@uni-tuebingen.de

Abstract

Dividing visual attention between spatially distinct sources of information could either be beneficial (if there is too much information for a single visualization) or detrimental (if interrelated information has to be mentally re-integrated) for learning. We present a new display technology allowing for the presentation of two distinct animations by avoiding split foci of visual attention: learners are able to switch between animations by moving their head. We examined how 84 naïve learners integrated information in three presentation modes: the “vexing-image” mode displaying two animations, participants being able to switch between them without shifting the visual focus, a classical “split-screen”, and an “overlaid” condition. Results showed that reduced complexity led to higher performance. Further, we showed that participants with high mental rotation abilities were best in the “split-screen” mode, whereas participants with low mental rotation abilities benefited most from the “vexing-image”. Theoretical and instructional consequences of these findings are discussed.

Keywords: multimedia learning, split attention, mental rotation, aptitude-treatment interaction effects

1. Introduction

One key characteristic of knowledge acquisition is the mental integration of information from different sources. Over the years, a vast body of research has addressed the role of different combinations of external representations – such as formulas, texts, pictures, and animations – in knowledge acquisition (Ainsworth, 2006). Closer inspection of the field shows that both text-picture combinations and audio-animation combinations have been extensively examined (e.g. Brünken, Plass, & Leutner, 2003; Huff & Schwan, 2008; Renkl, Atkinson, & Grosse, 2004; Schnotz, 2002). Aside from a few notable exceptions (Dutke & Rinck, 2006; Lee, Plass, & Homer, 2006), combinations of multiple pictorial external representations (both static and dynamic) have received scant attention. This is surprising because complex learning material often makes use of multiple pictures or multiple animations. For example, in medical testing, doctors have to visually compare two X-rays in order to monitor treatment success or in the field of chemical structures and processes, learners must attend to multiple animations simultaneously to get to the correct solution (Kozma, Chin, Russell, & Marx, 2000). In order to highlight certain aspects of the topic, designers of learning environments often place such visualizations side by side or present them in serial order. In this sense, the question arises whether established notions of multimedia learning can be applied to pairs of dynamic visualizations as well. This assumption was tested in terms of interrelations between design principles, complexity of the learning material, and learner characteristics. Further, our study investigated the value of innovative display technology as a means for presenting pairs of dynamic visualizations in an integrated form to model the vexing image representation displaying two animations such that participants are able to switch between them without shifting the visual focus. Instead they just have to move their head.

2. Related literature

2.1. Multimedia learning

Currently, there are two well-established models that describe knowledge acquisition with multiple external representations – namely the *Cognitive Load Theory* (CLT; Ayres & Van Gog, 2009; Chandler & Sweller, 1991; Sweller, van Merriënboer, & Paas, 1998) and the *Cognitive Theory of Multimedia Learning* (CTML; Mayer, 2001, 2006). Both are based on the assumption of a capacity-limited working memory, which is composed of multiple sub-systems. Thereby, the authors refer to three general cognitive assumptions: (1) learning materials are processed in two separate modality-specific channels (Baddeley, 1997; Paivio, 1986). “Modality” refers to the specific code information is presented in. Although written text and a picture share the same modality (the visual system) they are presented in different codes, namely verbal (written text) and pictorial code (picture). (2) The capacity of those channels is limited (Baddeley, 1997, Chandler & Sweller, 1991), and (3) learning means active processing by selecting incoming information and integrating it into existing knowledge structures (Mayer, 1997; Schnotz, 2002; Wittrock, 1989).

Based on these models, a number of principles have been deduced that focus on how to design a multimedia learning environment that provides an optimal working memory workload and how to foster with it sustainable learning and understanding (Mayer, 2006). Several of these principles, for example the spatial and temporal contiguity principles that describe the importance of an integrated presentation of learning materials, have been shown to be empirically well grounded (Ginns, 2005, 2006). To date, the models mentioned above have concentrated on the integrative processing of simultaneously presented text and pictorial

information (Schnotz & Bannert, 2003). The question of multiple, simultaneously presented, pictorial, external representations has not been addressed explicitly. This is surprising as the presentation of multiple pictures, which requires the learner to apply a thorough inspection and comparison of both visual sources, abounds in educational material (Messaris, 1996; Oestermaier & Hesse, 2000). Typical examples are comparisons of x-rays of healthy and unhealthy types of cellular tissue in medical education, of paintings by different artists in art history, or of different visualizations of a given molecule structure in chemistry.

2.2. Processing of visual information

From the perspective of CLT and CTML, pictorial information is processed in the visual channel. Both theories argue that if interrelated pieces of information are presented to the same channel in a disconnected way, learners have to change their attentional focus frequently during the learning phase in order to integrate all relevant information. This split attention effect is related to a high working memory load thus hindering deeper processing (Kester, Kirschner, & van Merriënboer, 2005; Tarmizi & Sweller, 1988). Therefore, Mayer (2001) posed the principle of spatial contiguity, which aims at reducing split attention effects. Tarmizi & Sweller (1988) demonstrated in a first study on spatial contiguity that presenting visual information (charts and text) in an integrated way fosters learning and comprehension.

More specifically, in the case of presenting a picture and an accompanying text in a side-by-side manner, the main reason for the occurrence of a split attention effect is the necessity for the learner to establish point-to-point relations between parts of the picture and corresponding, but spatially distant, text segments. This is a resource demanding process, which requires cycles of identifying a certain part of a picture, memorizing it in the working memory, switching

attention to the text, searching for the corresponding text segment, and finally cognitively relating both pieces of information. In contrast, placing corresponding picture parts and text segments close to each other in a so-called “integrated format” minimizes the required effort for establishing point-to-point-relations. Empirical studies showed no clear result pattern regarding the preferred order of information extraction. Whereas Hegarty and Just (1993) showed that participants looked first at the verbal information, other studies found initial processing of the pictorial information (Rayner, Rotello, Steward, Keir, & Duffy, 2001; Underwood, Jebbett, & Roberts, 2004; Stone & Glock, 1981) suggesting different (task-dependent) strategies.

Similar demands should also hold for pairs of pictorial visualizations (pictures and/or animations) as learning material. Again, if both visualizations are placed side-by-side, learners are required to establish point-by-point relations between corresponding parts of the pictures. In line with the argumentation of CLT and CMTL (Ayres & Paas, 2007), research on comparative visual searches has shown that the distance between two related pictorial presentations exerts a strong influence on perceptual and cognitive processes (Hardiess, Gillner, & Mallot, 2008; Inamdar & Pomplun, 2003). In the comparative visual search paradigm, viewers are confronted with two pictures of object arrays (e.g. books on a shelf) that are similar except for one object (e.g. one particular book on the shelf), which the viewers have to detect. Research findings show that if the pictures are close to each other, viewers show a perception-based search strategy, which minimizes working memory load by brief glances at each picture in combination with frequent switches of focus between the two pictures (Pomplun, Sichelschmidt, Wagner, Clermont, Rickheit, & Ritter, 2001). In contrast, if the pictures are more distant from one another, viewers have to rely on a more resource-demanding, memory-based search strategy, which is characterized by extended gazes at each picture in combination with fewer switches of

focus between the pictures (Inamdar & Pomplun, 2003). From the findings of CLT and CTML, as well as from comparative visual search, it can be expected that the principle of spatial contiguity should also hold for pairs of visualizations. Therefore, learners should benefit from an integrated presentation.

2.3. Integrated presentation of pictorial information

But what would an integrated presentation format of two corresponding visualizations look like? One obvious solution is to superimpose both pictures, thereby creating a compound visualization. By using this strategy, all corresponding elements of the two visualizations are exactly at the same spatial location, thereby avoiding split attention effects. However, on the other hand, the visual complexity of the display, defined as the absolute number of subcomponents of the image, is substantially increased (Donderi, 2006; Forsythe, Sheehy, & Sawey, 2003) thus resulting in high-complex displays. Cognitive load theory identified inherent complexity or element interactivity as one source for intrinsic cognitive load (Ginns, 2006; Sweller & Chandler, 1994).

Lee, Plass, and Homer (2006) recently addressed the trade-off between presenting information on a single vs. two displays and high vs. low complexity. By examining the influence of the design of a learning environment on learning outcome in terms of comprehension and transfer knowledge, Lee et al. (2006) found that reducing complexity by dividing the depicted information onto two screens increased learning performance for learners with high prior knowledge. These findings indicate that learning from integrated displays may be difficult due to a heightened degree of complexity, which imposes additional cognitive load on the learners. On the other hand, learners with low prior knowledge did not benefit from

separating information onto two screens, because the advantage of reduced visual complexity was nullified by the task of mentally (re)-integrating the relevant information. This is in line with meta-analytic results suggesting that for complex learning materials increasing spatial contiguity is beneficial for learning outcomes (Ginns, 2006).

New digital display technology may offer a solution to this dilemma. In particular, so-called “auto-stereoscopic displays” enable the creation of digital vexing images of two different visualizations. Which visualization is seen by a viewer depends on his or her viewing angle, which can be changed easily by moving the head position. In other words, with this type of technology, viewers can keep their visual focus on a certain point on the screen and nevertheless switch between two different visualizations by slight movements of their heads. As there is no need of relocating attention between two visualizations, this technology therefore avoids split attention, while keeping both visualizations separated, thus also avoiding an increase in complexity due to superimposition of the two visualizations.

Taken together, both theoretical deliberations and empirical results suggest that an integrated presentation of interrelated pictorial information obviates costly search and integration processes. As a result, the probability for simultaneous processing of related information in the working memory increases. As a consequence, the design of a multimedia learning environment should take into consideration the presentation of interrelated pictorial information in an integrated way. However, on the other hand, as the study by Lee et al. (2006) has shown, stimulus complexity and learners’ characteristics must also be considered as influencing factors. To address these issues in more detail, an experimental study was conducted in the area of molecular biology, which investigated the relationship between display format, complexity of the

content, and characteristics of the learner with regard to processing pairs of corresponding visualizations.

2.4. Experimental overview and hypothesis

In order to test influences of presentation mode and stimulus complexity on visual integration processes, we conducted an experiment that examined a classification task with low and high complex stimuli in three presentation modes. To examine split attention effects that are not influenced by any inherent redundancies it is important that both pictorial representations contribute essential information to understanding the content and that the task can only be solved by mentally integrating information from both representations (for an overview see Ayres & Sweller, 2005). In order to solve the presented task, participants had to identify, locate, integrate, and interpret information from two complementary and non-redundant dynamic representations of protein molecules.

Although the material consisted of molecules, the task itself did not require prior knowledge of chemistry, but did require some basic cognitive competencies, such as mental rotation abilities (e.g., Pillay, 1994). In the “split screen” mode, the two representations were displayed in a side-by-side manner. Consequently, participants needed to move their visual focus in order to identify corresponding information. In contrast, the “integrated” presentation modes – “vexing image” and overlaid -, in which corresponding visualizations were presented close to each other, make large shifts of the visual focus, accompanied by resource-consuming memorization processes, unnecessary. In the “vexing image” presentation mode, learners were able to switch between two dynamic representations by head movement, without changing their visual focus. In the “overlaid” mode, both representations were presented in a single

presentation, which was based on the superimposition of both visualizations. In the latter presentation mode, the complexity of the material as defined by the number of identifiable subcomponents was higher than in the other two conditions.

We hypothesized that presenting molecules in a low complex presentation mode (“vexing image” and “split screen” modes) leads to higher classification rates and lower classification times than presenting them in a high complex presentation mode (“overlaid” mode; see also Table 1). Further, we addressed the influence of spatial contiguity in terms of split attention on the classification task. Here we predicted that low spatial contiguity (as it occurs in the “split screen” condition) would be detrimental for this task as frequent shifts of the visual focus hinders deeper processing (e.g., Kester et al., 2005). Consequently, classification rates should be higher and classification times lower in the condition with having no need for split attention, particularly in the “vexing image” presentation mode.

As inherent complexity of the learning material is related with cognitive load (Ginns, 2006; Sweller & Chandler, 1994) we further hypothesized that these findings should be more pronounced for high complex molecules than for low complex molecules.

(Table 1 about here)

A further research question addressed in this study was the interplay of cognitive characteristics, presentation mode, and learning outcomes (classification performance and classification time). Learner related factors like prior knowledge and spatial abilities are important factors in multimedia learning and are known to influence learning outcomes (for an overview see Mayer, 2001). Therefore, we measured mental rotation abilities (Vandenberg & Kuse, 1978) of learners.

We hypothesized that especially learners with low mental rotation abilities will profit from the integrated presentation modes (both “vexing image” and “overlaid” modes), because they are not forced to move their visual focus between two representations. In comparison, learners with high mental rotation abilities should be able to integrate visual information in the split screen mode. Moreover, those participants should benefit from reduced complexity in the “split screen” presentation mode (Lee et al., 2006).

3. Method

3.1. Participants and Design

Subjects were 84 students (27 male) of the University of Tübingen, Germany. They were naïve with regard to the objective of the study. Average age was 25 years ($SD = 2.8$). Participants received compensation for their participation and were randomly assigned to one of three cells of the between-subjects factor “presentation mode”, which described whether information was presented in a split-screen mode, in a vexing image mode, or in an overlaid mode. Additionally, stimulus complexity was manipulated as a within-subjects factor.

3.2. Material and Apparatus

For each participant, the materials consisted of one questionnaire for demographic information (paper-based) and a computer-based learning environment containing the instructions for the classification task. The computer-based learning environment consisted of 22 slides, which introduced the subsequent classification task. The learning environment was divided into three sections and it took approximately 20 minutes to complete. First, participants were presented with the various forms of molecule visualizations. Beginning with the classic

wireframe and sticks-and-dots models, the advantages of modern computer visualization with its models (e.g. cartoon and trace models) are displayed. Next, the molecular structure of proteins was introduced. Since the subsequent testing phase focused on the secondary structure of proteins, the primary, tertiary, and quaternary structure of proteins were only discussed briefly. The secondary structure was introduced by explaining the alpha-helices and beta-sheets. Subsequently, the learners' task, namely the classification of whether a displayed beta-sheet is parallel or anti-parallel, was described.

The non-interactive stimuli were composed of two animations, each depicting the same protein structure. An example of the stimuli is available on our webpage (<http://www.iwm-kmrc.de/cybermedia/attentional-processes/>). The first animation was the trace model and the second a sticks-and-dots visualization (see Figure 1). The animation lasted 60 seconds, during which the whole structure performed a rotation of 360 degrees around its vertical axis. Due to the high complexity of the visualizations, it was necessary to rotate the stimuli during the classification phase; otherwise, the internal structure of the molecules would not become evident. There were 10 low and 10 high complex molecules: If a molecule consisted of less than 101 amino acids each represented by four balls in the sticks-and-dot visualization, it was characterized as "low complex" (96 – 100 amino acids; $M = 98.3$; $SD = 1.3$); if there were more than 100 amino acids, it was labeled as "high complex" (101 – 128 amino acids; $M = 106.8$; $SD = 7.9$).

There were three presentation modes: In the "split screen" mode, the animations were presented side by side. In contrast, the "vexing image" mode used an auto-stereoscopic display, which presented the two animations such that learners were able to alternate between them simply by shifting their head slightly, without the need to shift the visual focus. Finally, in the

“overlaid” mode, the two animations were presented superimposed on each other (see Figure 1). Hence, both vexing image and overlaid mode were characterized by high spatial contiguity of stimulus presentation, and both the vexing image and split screen mode were characterized by a low complexity of the presentation. As a result, the vexing image presentation mode combines the advantages of both low complexity with high spatial contiguity of information.

(Figure 1 about here)

The classification task required participants to relate information from both animations (see also Figure 2). First, they had to identify the location of a beta-sheet on the trace model. Then they were required to count the number of the atoms between two H-bonds in the sticks-and-dots visualization. A beta-sheet is classified as “anti-parallel” if the number of atoms is the same on both strings of the beta-sheet. If the number of the atoms on both strings is different, the beta-sheet is classified as “parallel”. As we wanted to exclude possible influences of prior knowledge, we instead addressed more generic cognitive functions and focused on this specific molecular structure. This is not part of the German (i.e., primary and secondary) school curriculum (Linder, 2005) and no participant was familiar with these concepts.

(Figure 2 about here)

Cognitive load was assessed with the self-report questionnaire NASA-TLX (Hart & Staveland, 1988). The mental rotation test (MRT, Vandenberg & Kuse, 1978) was employed to measure participants’ speed and accuracy in matching two figures that were rotated.

With the post-test, we controlled for learners' (prior) knowledge of the classification test. It contained items that were directly related to the classification task (e.g. "If a beta sheet has the same number of carbon atoms on both strings, is it classified as 'anti parallel' or 'parallel'?"), as well as items that asked for alternative solution procedures. This was necessary, as there was an alternative way to solve the classification task, which did not require the participants to integrate information from both dynamic representations. Instead, information depicted in just one animation could be sufficient to solve the task. Hence, the observed classification behavior would not reflect the task of integrating information from two pictorial animations. However, no participant knew of an alternative strategy. This test was given at the end of the experiment, as we wanted to eliminate the possibility that it influenced participants' classification strategies during the experiment.

The experimental procedures were controlled by a PC and programmed using Visual Basic (<http://www.microsoft.com/vbasic/>). The learning environment and the stimuli of the classification task were presented on a 24" display. In all presentation modes, the visual angles of the animations were comparable.

3.3. Procedure

All participants were tested individually and received instructions via the computer monitor. The experiment lasted approximately 90 minutes. At the beginning, participants completed the learning environment. The subsequent classification task began with a training phase consisting of four proteins that had to be classified as "parallel" or "anti-parallel"; these exercise trials were excluded from analysis. The subsequent experimental phase contained 20

protein structures. Finally, the cognitive load assessment questionnaire, the mental-rotation test, and the post test were filled out by the participants.

In the classification task, the learners' task was to find and to classify a beta sheet within the 60 seconds of stimulus presentation. To this end, they first needed to localize the beta sheet on the amino acid in the trace model visualization. Consequently, they had to switch to the other representation depicting the same molecule in the same orientation as a sticks-and-dots model. Here learners were asked to count the number of carbon atoms between two red-colored H-bonds on both strands of the molecule. If the number of carbon atoms on both strands was the same, then the beta sheet was anti-parallel. If there were a different number of carbon atoms between two H-bonds on both strands, then it was parallel. Participants were asked to press one of two buttons on the keyboard labelled "anti-parallel" and "parallel" respectively.

4. Results

4.1. Prior Knowledge

Asking participants if they knew the domain of beta-sheets and possible alternative strategies of solving the classification task tested prior knowledge. No participant was familiar with the concept of "parallelism" of beta-sheets and no one mentioned a possible alternative strategy. Therefore, it is reasonable to assume that prior knowledge was comparable across all participants and conditions.

4.2 Classification Task

In the classification task, the proportion of correct answers and classification times were analyzed. Correctly classified proteins were coded with "1", wrong answers with "0". Chance

performance was accordingly 0.5. The time from the beginning of the presentation to pressing the key, which indicated whether the protein was classified as “parallel” or “anti-parallel”, was recorded. For each dependent measure, a separate ANOVA, including the factors “presentation mode” (split screen vs. vexing image vs. overlaid; between-subjects) and “molecule complexity” (low vs. high; within-subjects), was calculated. The results are displayed in Table 2.

(Table 2 about here)

4.2.1. Proportion Correct

Across all conditions and participants, a mean of $M = .70$ ($SD = .10$) was measured. Although low complex molecules led to higher classification rates than high complex molecules, $F(1, 81) = 5.14, p = .026, \hat{\omega}^2 = .05$, the presentation mode had no impact, $F < 1$ (see also Table 2). The interaction of these two factors also failed the level of significance, $F(2, 81) = 1.41, p = .25, \hat{\omega}^2 = .01$.

4.2.2. Classification Time

The mean classification time across all conditions and participants was $M = 28.65$ ($SD = 7.15$) seconds. Whereas low and high complex molecules did not lead to different classification times, $F < 1$, presentation mode did, $F(2, 81) = 3.53, p = .034, \hat{\omega}^2 = .21$. Classification time was longest in the “overlaid” condition; in both the “vexing image” and the “split-screen” conditions, classification times were comparable (see Table 2). Again, the “complexity” x “presentation mode” interaction did not reach the level of significance, $F(2, 81) = 1.56, p = .22, \hat{\omega}^2 = .01$, (see Table 2).

4.3. Learner Related Factors

In the further analysis, we focused on learner related factors and analyzed them with methods known from research on aptitude-treatment interaction effects (ATI; Snow, 1991; see also Cronbach & Snow, 1977). To test the relationship between participants' cognitive characteristics and the dependent measures "proportion correct" and "classification time" regression analysis was applied with the predictors "mental rotation ability" (measured with the mental rotation test, MRT, Vandenberg & Kuse, 1978) and "cognitive load" (measured with the self-report questionnaire NASA-TLX, Hart & Staveland, 1988). Both measures were not correlated, $r = .06$, $t(78) = 0.55$, $p = .58$. For both "proportion correct" and "classification time", a separate linear regression analysis, including the between-subjects factor "presentation mode", was calculated.

4.3.1. Mental Rotation Abilities

Participants' MRT scores were calculated by subtracting the incorrectly answered items from the correctly answered items. They were comparable across the three cells of the between-subjects factor "presentation mode", $F < 1$, (split screen: $M = 22.77$, $SD = 7.69$, vexing image: $M = 20.93$, $SD = 7.96$, and overlaid: $M = 21.03$, $SD = 9.52$).

Regarding the dependent variable "proportion correct", the regression analysis showed a significant effect, multiple $R^2 = .13$, $F(5, 78) = 2.38$, $p = .046$ (see Figure 3). Whereas there was no influence of MRT abilities on "proportion correct" in the "vexing image", $t(78) = 0.12$, $p = .90$, and "overlaid", $t(78) = 1.09$, $p = .28$, conditions, the "split screen" presentation mode was influenced by MRT abilities, $t(78) = 2.92$, $p = .004$. Learners with high MRT scores achieved higher classification rates than participants with low MRT scores. By comparing the slopes of the conditions with low complex presentations ("vexing image" and "split screen"), we

demonstrated that especially participants with a low MRT score profit from the vexing image presentation mode. Participants with a high MRT score, on the other hand, were better in the split screen mode, $t(78) = -2.23, p = .028$.

(Figure 3 about here)

A similar result pattern was found for the dependent variable “classification time”. The linear regression analysis showed a significant effect, multiple $R^2 = .20, F(5, 78) = 3.92, p = .003$, with the following result pattern (see also Figure 3): While the “vexing image”, $t(78) = -0.15, p = .880$, and “overlaid”, $t(78) = -0.41, p = .685$, conditions were not affected by participants’ mental rotation abilities, there was an effect in the “split screen” condition, $t(78) = -3.50, p < .001$; the MRT score was positively correlated with the dependent measure “classification time”. Learners with high MRT scores outperformed those with low MRT scores. Especially participants with higher MRT scores benefited from the split screen presentation. Single comparisons of the slopes of the “vexing image” and the “split screen” conditions revealed a significant difference, $t(78) = 2.39, p = .019$.

4.3.2. Cognitive Load

Cognitive load was assessed with a German adaptation of the NASA-TLX (Hart & Staveland, 1988). Cronbach’s alpha was .28 and correlations showed that the item “scale own performance” was negatively correlated with the other items. Therefore, we computed the cognitive load measure from the sum of all items with the exception of “scale own performance”, for which the Cronbach’s alpha was .58 (this finding is in line with results from Fischer, Lowe,

& Schwan, 2008). The cognitive load was comparable across all presentation modes, $F < 1$ (split screen: $M = 33.73$, $SD = 13.36$, vexing image: $M = 38.47$, $SD = 13.22$, and overlaid: $M = 41.71$, $SD = 13.59$).

The linear regression with the independent variable “presentation mode” and the cognitive load score revealed no significant effect regarding the dependent variable “proportion correct”, *multiple* $R^2 = .04$; $F < 1$. Regarding the “classification time”, the model was significant, *multiple* $R^2 = 0.16$, $F(5, 78) = 2.96$, $p = .017$ (see also Figure 4). Again, there was no influence of the cognitive load score on “classification time” in the “vexing image”, $t(78) = -1.17$, $p = .246$, and “overlaid”, $t(78) = 0.87$, $p = .385$, conditions. However, in the “split screen” condition, participants with higher cognitive load scores showed longer classification times than participants with lower cognitive load scores, $t(78) = 2.43$, $p = .017$. Pairwise comparisons of the “vexing image” and the “split screen” conditions underline this finding by showing a significant effect, $t(78) = -2.52$, $p < .014$.

(Figure 4 about here)

5. Discussion

In the present article, the results of an experiment that contributes to the ongoing discussion on multiple external representations in multimedia learning scenarios were presented. Two theoretical approaches, namely the *Cognitive Load Theory* (CLT; Sweller, van Merriënboer, & Paas, 1998) and the *Cognitive Theory of Multimedia Learning* (CTML; Mayer, 2001), were proposed to predict knowledge acquisition from (multiple) external representations. According to the spatial and temporal contiguity principle, related information should be presented in an

integrated manner (Ginns, 2006; Mayer, 2001). Prior research focused mainly on text-picture or audio-animation relationships and their impact on learning outcomes. Recent developments in technology (wide use of dynamic animations), empirical findings (Lee et al., 2006), as well as theory (DeFT approach by Ainsworth, 2006), make the further development of theoretical accounts of multimedia learning necessary. Lee et al. highlighted the importance of reducing visual complexity of learning material. In doing so, the authors extended multimedia learning to the important field of purely pictorial information and its influence on learning outcome.

Ainsworth (2006) describes different functions of multiple external representations. In the current study, we focused on the cognitive processing of complementary information. Based on these deliberations, we designed a study that implemented a classification task, in which participants were required to relate pictorial information from two external dynamic representations. In order to address generic cognitive processes, which are the basis for multimedia learning (like the integration of information from distributed information), we chose a task that primarily requires mental rotation abilities (Pillay, 1994). Participants had to relate information from two external dynamic representations (rotating protein molecules) and subsequently answer a question. The main finding of the present study is that, in the case of pairs of visualizations, both split attention effects and complexity effects must be taken into account and that this tradeoff is further qualified by the mental rotation abilities of the learners.

In terms of the experiment, the results show that the complexity of the presentation exerted a strong influence on behavior of the participants. In contrast to low complex molecules, high complex molecules led to a decrease in classification rates across all presentation modes. Of even greater importance, despite being an integrated format, the presentation mode with the highest complexity, namely the overlaid mode, led to a significant increase in classification time.

This replicated common findings in research on multimedia learning that showed that inherent complexity of the learning material is related with learning performance (Ginns, 2006; Sweller & Chandler, 1994).

In addition to complexity, integrated and split presentation formats also influenced classification behavior. Aptitude-treatment interaction analyses showed that the integrated “vexing image” condition was not influenced by participants’ mental rotation abilities. In contrast, in the “split screen” condition – in which the result pattern for the dependent variables “proportion correct” and “classification time” was comparable to the “vexing image” presentation mode – there was a significant influence of mental rotation abilities on dependent measures. Participants with a high mental rotation test score outperformed those with lower scores. In other words, the integrated vexing image format proved to be more robust against variations in mental rotation abilities than the split screen format.

This pattern of results also suggests some implications for the underlying perceptual and cognitive processes. Presenting information in a highly complex manner (as in the “overlaid” condition) forces participants to extract relevant information and to suppress task-irrelevant information. Although there is no need for split attention, classifying molecules in this “overlaid” condition is thus a time-consuming process, albeit not dependent on individual mental rotation abilities. In the “vexing image” condition, there is also no need for split attention and, again, solving the classification task does not depend on mental rotation abilities. However, in contrast to the overlaid mode, the complexity of the presentation is kept comparatively low, and the classification time is reduced accordingly. Finally, in the “split screen” condition, the complexity is low, but the information has to be integrated by shifting the visual focus between the two external dynamic representations (low spatial contiguity). Thus, in this condition, performance

(that is classification times and classification rates) depends on mental rotation abilities, with higher mental rotation abilities leading to better performance. This suggests that participants with high mental rotation abilities are able to cope with multimedia designs providing only low spatial contiguity. By implication, multimedia designs should consider spatial contiguity as one important factor for learning success. Especially learners with low mental rotation abilities suffer from designs providing only low spatial contiguity. Additionally, while keeping prior knowledge constant, we observed substantial differences in performance measures due to individual differences in mental rotation abilities.

In the “split screen” condition, we measured longer classification times for participants who reported higher cognitive load than those participants who reported lower cognitive load. This pattern is in line with research on cognitive load – increased extraneous cognitive load might be responsible for this finding. On the other side, we did not observe such a relationship in the “vexing image” condition. Here, higher cognitive load measures did not involve longer classification times. A tentative interpretation could be that higher cognitive load measures in the “vexing image” condition were not extraneous (as in the “split screen” condition) but reflect participants’ task related efforts. This cautious interpretation and the fact that the internal consistency of the NASA-TLX questionnaire (Hart & Staveland, 1988) is limited (Cronbach’s alpha was .58) require further research that could make use of more elaborate measures of cognitive load (Paas, 1992; Paas, Tuavinen, Tabbers, & Van Gerven, 2003).

Taken together, these findings are in line with those of Lee et al. (2006): (re)-integrating interrelated but separated information is a costly cognitive process, which depends on cognitive characteristics. Whereas Lee et al. (2006) demonstrated the importance of domain-specific prior knowledge, we underline the importance of generic cognitive functions. By keeping domain

expertise constant, we were able to address a generic cognitive process that influenced the classification process: Participants with high mental rotation abilities outperformed those with low mental rotation abilities.

In conclusion, the study has shown that established principles of multimedia design, such as the spatial contiguity principle and the split attention effect, apply to pairs of pictorial animations as well. Further, for the design of multimedia learning environments (which often contain classification tasks), it is important not only to consider the spatial relations between multimedia elements, but also the complexity of the resulting presentation and the cognitive characteristics of the learners that are necessary to solve the tasks successfully. In particular, in the case of two pictorial visualizations, providing an integrated format only proved beneficial if presentation complexity was kept low. Here, new forms of digital displays can play an important role, as was exemplified in the present study by the advantages of auto-stereoscopic displays.

References

- Ainsworth, S. E. (2006). DeFT: A conceptual framework for learning with multiple representations. *Learning and Instruction, 16*(3), 183-198. doi: 10.1016/j.learninstruc.2006.03.001
- Ayres, P. & Van Gog, T. (2009). State of the art research into cognitive load theory. *Computers in Human Behavior, 25* (2), 253-392. doi: 10.1016/j.chb.2008.12.007
- Ayres, P. & Paas, F. (2007). A cognitive load approach to the learning effectiveness of instructional animation. *Applied Cognitive Psychology, 21* (6), 695-820.
- Ayres, P. & Sweller, J. (2005). The split-attention principle in multimedia learning. In R.E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 135-146). New York: Cambridge University Press.
- Baddeley, A.D. (1997). *Human memory: Theory and Practice*. Hove: Psychology Press.
- Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology, 76*(3), 408-417. doi: 10.1037/0022-0663.76.3.408
- Brünken, R., Plass, J.L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 53-61. doi: 10.1207/S15326985EP3801_7
- Chandler, P. & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332. doi: 10.1207/s1532690xci0804_2
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York: Irvington.

- Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, *132*(1), 3-97. doi: 10.1037/0033-2909.132.1.73
- Fischer, S., Lowe, R.K., & Schwan, S. (2008). Effects of presentation speed of a dynamic visualization on the understanding of a mechanical system. *Applied Cognitive Psychology*, *22*(8), 1126-11. doi: 10.1002/acp.1426
- Forsythe, A., Sheehy, N. & Sawey, M. (2003). Measuring icon complexity: an automated analysis. *Behavior Research Methods, Instruments, & Computers*, *35*(2), 334-342.
- Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *15*, 313-331. doi: 10.1016/j.learninstruc.2005.07.001
- Ginns, P. (2006). Integrating information: Meta-analyses of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*, 511-525. doi: 10.1016/j.learninstruc.2006.10.001
- Hardiess, G., Gillner, S., & Mallot, H.A. (2008). Head and eye movements and the role of memory limitations in a visual search paradigm. *Journal of Vision*, *8*, 1-13. doi: 10.1167/8.1.7
- Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Hg.), *Human mental workload*. (pp. 139-183). North-Holland.
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32*, 717-742.
- Huff, M., & Schwan, S. (2008). Verbalizing events: Overshadowing or facilitation? *Memory & Cognition*, *36* (2), 392-402. doi:10.3758/MC.36.2.392

- Inamdar, S., & Pomplun, M. (2003). Comparative search reveals the tradeoff between eye movements and working memory use in visual tasks. *Proceedings of the Twenty-Fifth Annual Meeting of the Cognitive Science Society*, 599-604.
- Kester, L., Kirschner, P. A. & van Merriënboer, J. J. G. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer-simulated problem solving. *British Journal of Educational Psychology*, 75(1), 71-85.
doi:10.1348/000709904X19254
- Kozma, R., Chin, E., Russell, J. & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *The Journal of the Learning Sciences*, 9(2), 105-143. doi:10.1207/s15327809jls0902_1
- Lee, H., Plass, J. L. & Homer, B. D. (2006). Optimizing cognitive load for learning from computer-based science simulations. *Journal of Educational Psychology*, 98(4), 902-913.
doi:10.1037/0022-0663.98.4.902
- Linder, H. (2005). *Biologie*. Bayrhuber, H. & Kull, U. (Hg.) *Biologie*. Schroedel: Hannover.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32, 1-19. doi:10.1207/s15326985ep3201_1
- Mayer, R. E. (2001). *Multimedia Learning*. New York: Cambridge University Press.
- Mayer, R. E. (2006). *Cambridge Handbook of Multimedia Learning*. New York: Cambridge University Press.
- Messaris, P. (1996). *Visual literacy: Image, mind, reality*. Boulder, CO: Westview Press.
- Oestermeier, U. & Hesse, F. W. (2000). Verbal and visual causal arguments. *Cognition*, 75(1), 65-104. doi:10.1016/S0010-0277(00)00060-3

Paas, F. G W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429.

Paas, F. G W. C., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1), 63-71. doi: 10.1207/S15326985EP3801_8.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

Pillay, H. K. (1994). Cognitive load and mental rotation: structuring orthographic projection for learning and problem solving. *Instructional Science*, 22, 91-113.
doi:10.1007/BF00892159

Pomplun, M., Sichelschmidt, L., Wagner, K., Clermont, T., Rickheit, G. & Ritter, H. (2001). Comparative visual search: A difference that makes a difference. *Cognitive Science*, 25, 3-36. doi:10.1016/S0364-0213(00)00037-9

Rayner, K., Rotello, C. M., Steward, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, 7, 219-226.

Renkl, A., Atkinson, R. K. & Grosse, C. S. (2004). How fading worked-out solution steps works - a cognitive load perspective. *Instructional Science*, 32, 59-82

Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, 14, 101-120.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction*, 13, 141-156.

- Snow, R. E. (1991). Aptitude-Treatment Interaction as a Framework for Research on Individual Differences in Psychotherapy. *Journal of Consulting and Clinical Psychology, 59*(2), 205-216. doi:10.1037/0022-006X.59.2.205
- Stone, D. E., & Glock, M. E. (1981). How do young adults read directions with and without pictures? *Journal of Educational Psychology, 73*, 419-426.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185-233.
- Sweller, J., van Merriënboer, J. & Paas, F. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review, 10*(3), 251-296.
- Tarmizi, R. & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*, 424-436. doi:10.1037/0022-0663.80.4.424
- Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology-A, 57*, 165-182.
- Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotations, a group test of three dimensional spatial visualizations. *Perceptual and Motor Skills, 47*, 599-604.
- Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*, 345-376. doi:10.1207/s15326985ep2404_2

Tables

Table 1. Prognoses for proportion correct and classification times. (+) higher proportion correct and shorter classification times; (-) lower proportion correct and longer classification times.

| | | Presentation Mode | | |
|---------------------|------|-------------------|--------------|----------|
| | | Vexing image | Split screen | Overlaid |
| Stimulus complexity | Low | ++ | + | - |
| | High | + | o | -- |

Table 2. Mean proportions of correct responses and classification time (SD in parentheses).

| | Presentation Mode | | | | | |
|-------------------------------|-------------------|-----------------|-----------------|-----------------|-----------------|------------------|
| | Vexing image | | Split screen | | Overlaid | |
| | Low complex | High complex | Low complex | High complex | Low complex | High complex |
| Proportion correct | 0.71 (0.13) | 0.68 (0.17) | 0.72 (0.13) | 0.70 (0.17) | 0.72 (0.13) | 0.62 (0.15) |
| Classification time (sec.) | 26.89 (6.81) | 28.74 (6.76) | 26.03 (7.76) | 26.91 (8.09) | 32.08 (7.55) | 30.56 (10.30) |

Figure Captions

Figure 1. Examples of the stimuli. A: original material with trace-model (T) and sticks-and-dots model (S). B: variations of the factor “presentation mode”.

Figure 2. Schematic illustration of the classification task. Participants first had to locate the beta-sheet on the trace model, then switch to the sticks-and-dots model, where they had to locate the two H-bonds bounding the beta-sheet. Finally, they were asked to count the number of carbon atoms on both strings.

Figure 3. Learner related factor “mental rotation ability”: classification time in the “split screen” mode was influenced by participants’ mental rotation abilities. Participants with high mental rotation abilities outperformed those with low measures. In contrast, the “integrated” (vexing image and overlaid) modes were not influenced by the mental rotation abilities. Classification performance was highest for participants with high mental rotation abilities in the “split screen” mode. There was no influence of mental rotation abilities in the “vexing image” and “overlaid” modes.

Figure 4. Learner related factor “cognitive load”: there was an effect of cognitive load (measured with the NASA-TLX) on classification time in the split screen condition. Participants with higher scores needed more time for the classification task.

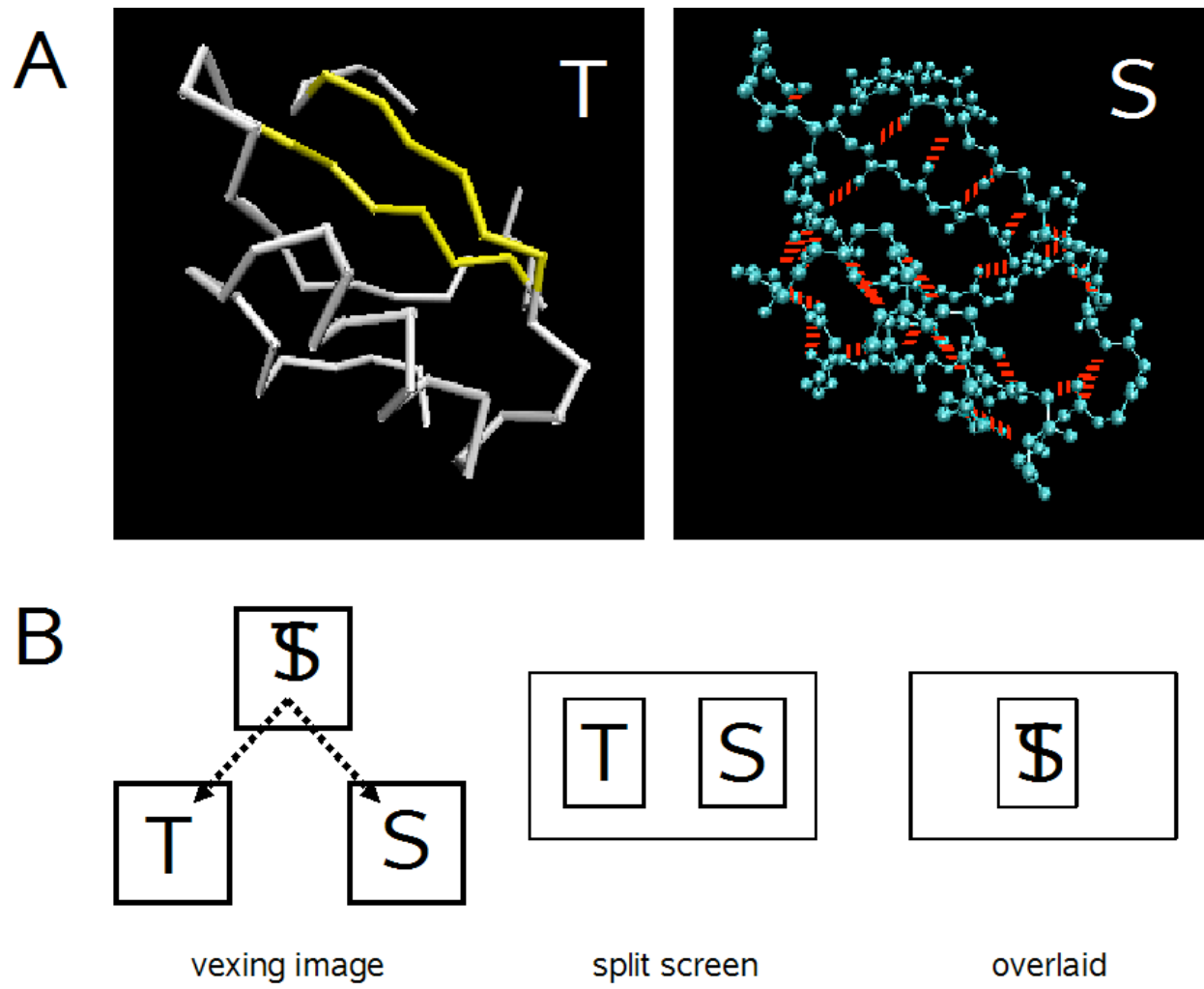


Figure 1

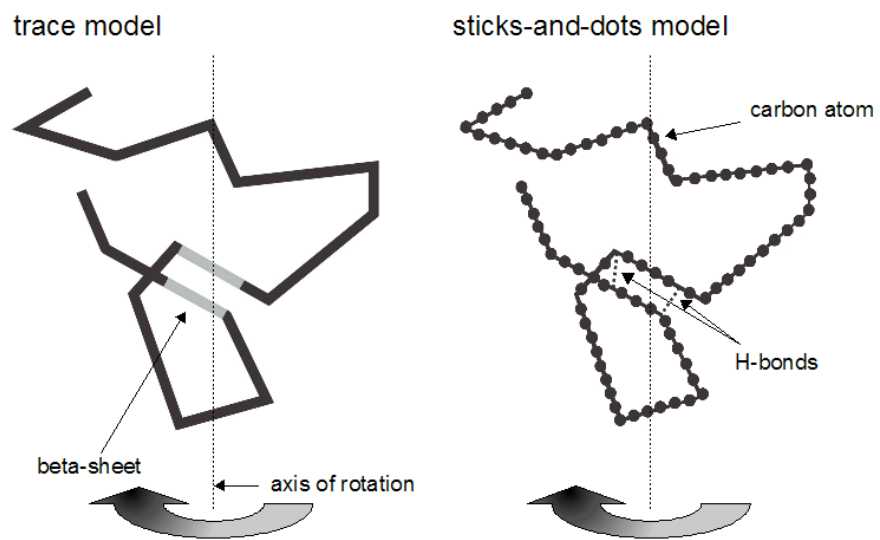


Figure 2

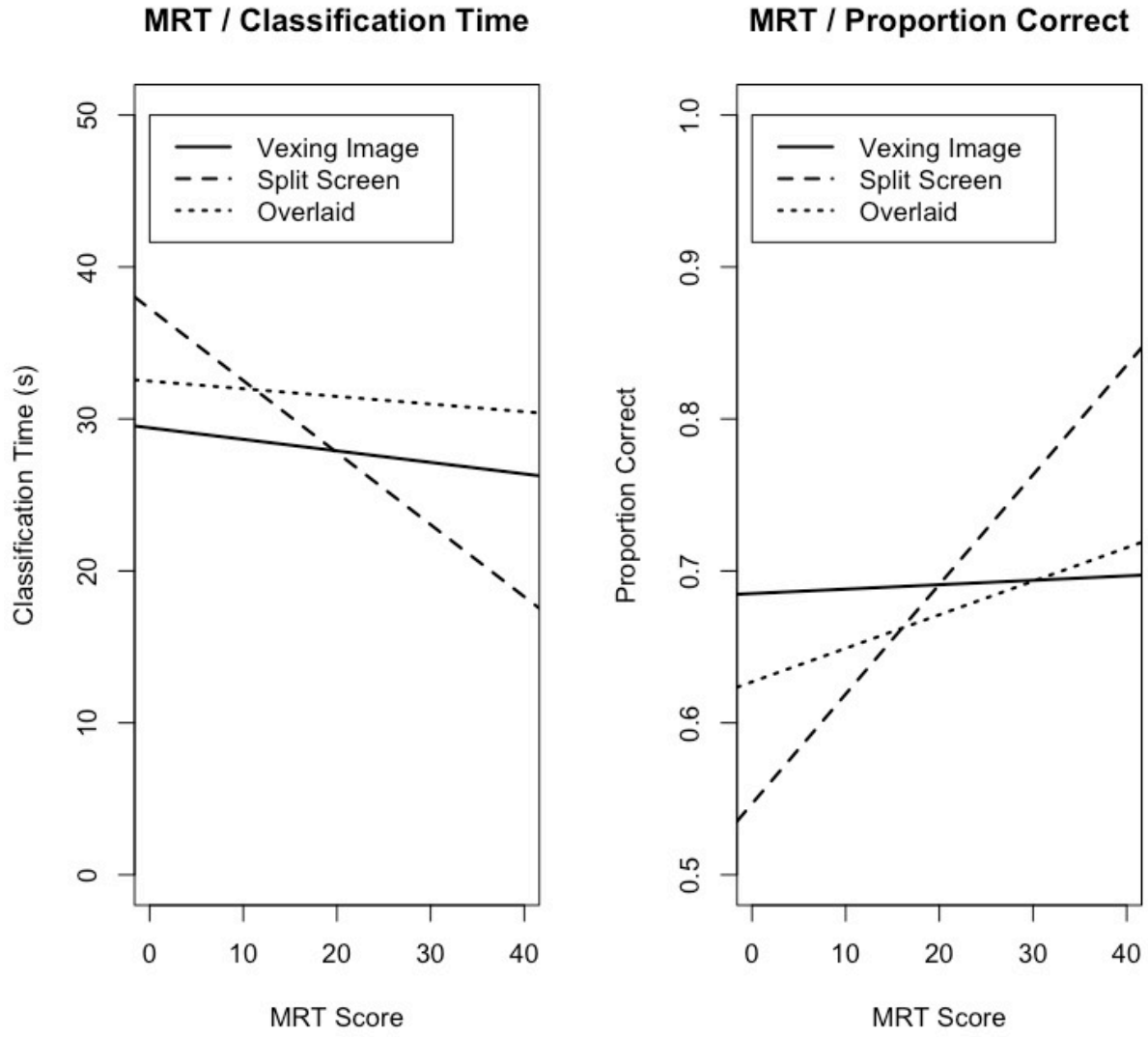


Figure 3

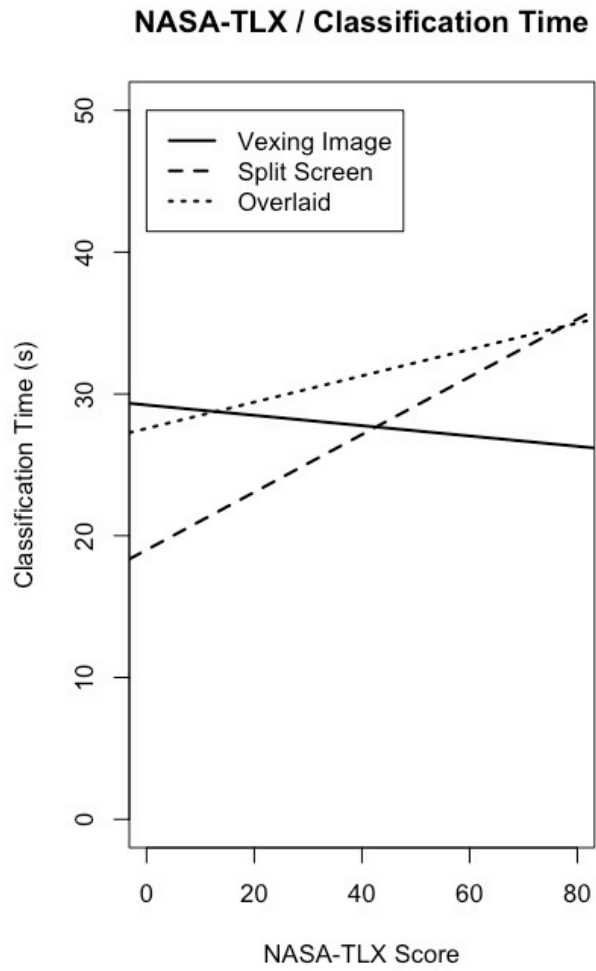


Figure 4