Kumulative Dissertation
zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)
an der Fakultät für Psychologie und Sportwissenschaft der Universität Bielefeld

Eye movements, attention, and memory processes during sensorimotor learning and automatization

Bielefeld, November 2011
vorgelegt von **Dipl.-Psych. Rebecca M. Förster**

Acknowledgements

First, I would like to thank my supervisors *Werner X. Schneider* and *Elena Carbone*. I was very fortunate to have supervisors who encouraged me to explore my own ideas and at the same time provided me the guidance I needed to stay focused. They did not only bring forward this work, but also promoted me with respect to advanced career carrier management. Most important, they were always within reach and had practicable advices that helped me to solve emerging problems. This was very helpful, especially in problematic phases of my work. In addition, I would like to express my gratitude to *Hendrik Kösling*, *Thomas Hermann*, and *Bettina Bläsing* for their constant technical, professional, and personal support on this PhD project. It should also be mentioned that I was in the privileged situation to get my PhD research supported by grants of the *Center of Excellence Cognitive Interaction Technology (CITEC)* of Bielefeld University. Furthermore, I would like to thank *Slobodan Vukanovic* from Bielefeld University who agreed to read parts of this thesis and who provided valuable comments. Finally, I send my special thanks to *Heike Hartwig-Jakobs*, *Verena Donnerbauer*, *Okka Risius*, and *Frank Lehmke* as well as to all members of the *AE 1 – Neuro-cognitive Psychology* for their support that reduced my workload, and for lending an ear to me for all my needs and worries.

CONTENTS

ABSTRACT

Humans move their eyes via saccades to informative locations in the environment to accomplish a sensorimotor task. Eye movements and their characteristics have been investigated in both restricted laboratory tasks and less restricted more natural tasks. However, only a few studies addressed the development of attentional control during learning of natural tasks and its relationship to memory processes. In the present work, eye movements and memory processes were investigated during learning and automatization of a bimanual sensorimotor high-speed stacking task. This task requires grasping, lifting, moving, rotating, and placing of objects. By means of three studies, the present work investigated how attention control changes during learning and automatization in such a task, and how attention control, eye-movements, and memory processes interact after intensive training. The first study (manuscript 1) compared participants‟ eye movement patterns exhibited at the beginning of the learning process with their eye movement patterns exhibited after automatization. The second study (manuscript 2) compared eye movements of the same participants in light and dark illumination conditions while they had to execute the automatized stacking task. The third study (manuscript 3) tested working-memory performance while participants had to execute the LTM-based automatized stacking task. The first study (manuscript 1) revealed that the eye led the hand with shorter absolute delays after intense training, yet with longer delays when correcting them by the decreasing task duration. In addition, fewer fixations were performed with increasing practice. However, all participants fixated on the same task-relevant locations in a similar sequence across training days. These results indicate that sequences of saccades (scan paths) are transferred to the long-term memory (LTM) during learning and are retrieved from LTM after automatization. In favor of this assumption, highly systematic eye movements were observed in the dark illumination condition of the second study (manuscript 2). The comparison between eye movements in the dark and in the light revealed that number

of fixations, scan paths, and eye-hand dynamics were very similar between illumination conditions, while fixation rates were lower and fixation durations were longer in the dark. Moreover, the stacking task as well as an additional verbal automatized high-speed task recruited internal attention and domain-specific working memory (WM) even after automatization in the third study (manuscript 3). In summary, these results point to a gradual transition from a more sensory-based to a more LTM-based visual selection mode during learning and automatization of a high-speed sensorimotor task. In addition, attention, eye movements, WM, and LTM processes seem to work tightly together in the following way after automatization. Task-relevant LTM representations are retrieved by internal attention. The same internal attention process maintains the retrieved information which corresponds to retention within domain-specific WM. The information is maintained until it is used to determine where-to-attend and where-to-look next in the environment. By means of the subsequent fixation, task-relevant information is sampled from the environment which is used to determine the execution of the upcoming sensorimotor act.

DEUTSCHE ZUSAMMENFASSUNG

Um eine sensomotorische Aufgabe auszuführen, springt das menschliche Auge mittels Sakkaden zu informativen Positionen in der Umgebung. Diese Augenbewegungen wurden nicht nur in Laborsituationen, sondern auch während der Ausübung natürlicher Aufgaben hinreichend untersucht. Nur wenige Studien haben hingegen den Entwicklungsprozess der Aufmerksamkeitskontrolle und ihren Zusammenhang mit Gedächtnisprozessen während des Lernens natürlicher Aufgaben untersucht. In der vorliegenden Arbeit wurden Augenbewegungen und Gedächtnisprozesse während des Lernens und Automatisierens einer zweihändigen, sensomotorischen, ultraschnellen Stapelaufgabe untersucht. Diese Aufgabe besteht aus dem Greifen, Anheben, Bewegen, Rotieren und Platzieren von Objekten. Anhand dreier Studien untersuchte die vorliegende Arbeit einerseits die Veränderung der Aufmerksamkeitssteuerung während des Lernens und Automatisierens und andererseits das Zusammenspiel von Aufmerksamkeit, Augenbewegungen und Gedächtnisprozessen nach intensivem Training. In einer ersten Untersuchung (Manuskript 1) wurden die Augenbewegungsmuster von Versuchsteilnehmern am Anfang des Lernprozesses verglichen mit ihren Augenbewegungsmustern nach der Automatisierung. In einer zweiten Untersuchung (Manuskript 2) wurden die Augenbewegungen derselben Versuchsteilnehmer in einer beleuchteten und einer unbeleuchteten Versuchsbedingung verglichen, während sie die Stapelaufgabe ausführen mussten. In einer dritten Untersuchung (Manuskript 3) wurde schließlich parallel zur Ausführung der langzeitgedächtnisgesteuerten, automatisierten, sensomotorischen Aufgabe die Arbeitsgedächtnisleistung getestet. Die erste Untersuchung (Manuskript 1) konnte zeigen, dass das Auge die Hand nach intensivem Training mit einem kleineren absoluten Zeitabstand führte als vor dem Training. In Relation zur abnehmenden Aufgabendauer wurde dieses Intervall zwischen Auge und Hand jedoch größer. Weiterhin wurden mit zunehmender Übung weniger Fixationen gemacht. Außerdem fixierten alle Teilnehmer über die verschiedenen Trainingsta-

ge hinweg ähnliche aufgabenbezogene Orte in einer ähnlichen Reihenfolge. Die Ergebnisse deuten darauf hin, dass Sequenzen von Sakkaden (scan paths) durch den Lernprozess in das Langzeitgedächtnis (LZG) transferiert werden und nach der Automatisierung wieder aus dem LZG abgerufen werden. Im Einklang mit dieser Hypothese waren die beobachteten Augenbewegungen in der dunklen Beleuchtungsbedingung der zweiten Studie (Manuskript 2) im höchsten Maße systematisch. Der Vergleich zwischen normaler Beleuchtung und Dunkelheit zeigte, dass die Anzahl der Fixationen, die Trajektorien der Fixationen, sowie die Dynamik zwischen Auge und Hand in den beiden Beleuchtungsbedingungen sehr ähnlich waren, während im Dunkeln die Fixationsraten geringer und die Fixationsdauern länger waren. Darüber hinaus benötigten die Stapelaufgabe als auch eine zusätzliche verbale, ultraschnelle Aufgabe in der dritten Untersuchung (Manuskript 3) sogar nach der Automatisierung noch internale Aufmerksamkeitsprozesse und das domänenspezifische Arbeitsgedächtnis. Zusammengefasst deuten die Ergebnisse auf einen graduellen Übergang von einer eher sensorisch-gesteuerten zu einer eher LZG-gesteuerten visuellen Selektion während des Lernens und Automatisierens einer ultraschnellen, sensomotorischen Aufgabe. Außerdem scheinen Aufmerksamkeit, Augenbewegungen, Arbeitsgedächtnis- und LZG-Prozesse nach der Automatisierung in der nachfolgend beschriebenen Art und Weise eng zusammenzuarbeiten. Aufgabenbezogene LZG-Repräsentationen werden durch internale Aufmerksamkeitsverlagerung abgerufen. Derselbe internale Aufmerksamkeitsprozess sorgt dafür, dass die abgerufene Information aufrecht gehalten wird. Dies entspricht der Aufrechterhaltung innerhalb des domänenspezifischen Arbeitsgedächtnisses. Anschließend wird die Information genutzt, um die nächste Aufmerksamkeits- und Blickverlagerung in der Umgebung zu bestimmen. Mit Hilfe der nachfolgenden Fixation wird aufgabenrelevante Information aus der Umgebung gesammelt, welche schließlich die Ausführung der bevorstehenden sensomotorischen Handlung bestimmt.

1 Introduction

The present synopsis reports on a dissertation that consists of three manuscripts and aims at better understanding the interrelationship of eye movements, attention, and memory processes during sensorimotor learning and automatization. For this purpose eye movements, attention, and memory processes were investigated during and after learning and automatization of a sensorimotor task. Before studying how attention, eye movements, and memory processes interact during the execution of automatized tasks, it is important to understand each of these components separately. The following synopsis therefore begins with a review of some fundamental characteristics of eye movements, attention, and memory as well as theories of automatization. Subsequently, important prior research on the relation of two or more of these components is outlined. Afterwards, the purposes, questions, and hypotheses as well as the experiments of the underlying dissertation are briefly described. Finally, a general discussion summarizes the main results of the underlying work and discusses them with regard to automatization theories and the interplay of attention and memory during and after automatization.

1.1 Eye movements

Humans are highly dependent on visual perception. The eyes gather visual information that is important for interacting with the environment. While fixating a specific location in the surrounding, the retina of the eye can process the visual input (Dodge, 1900). Due to the inhomogeneity of the retina, only a limited part of the visual image can be processed at the highest quality (between .3° and 2° of visual angle). This part of the retina is called the fovea. Departing from the fovea, the resolution decreases progressively, reaching the tenth of the value of the highest quality in the far periphery of the retina. The fact that only the fovea can be processed at the highest resolution at a given time necessitates the execution of successive eye, head, and trunk movements to bring interesting visual stimuli on the fovea. The eyes move about three times a second in a staccato fashion. These fast, conjunctive, erratic eye movements are called saccades. The functional role of saccades is to bring informative parts of the visual environment to the retina, so that these parts can be processed during the interval between two successive saccades, by a fixation.

The resulting sequence of alternating saccades and fixations is called the saccadic scan path (Noton & Stark, 1971a, 1971b). Scan paths differ across conditions, both across visual inputs, and across task instructions. Yarbus (1967), for instance, observed not only different scan paths when presenting participants different paintings, but also when presenting one and the same painting with different instructions. When participants had to judge the age of persons in a painting of a living room, they almost exclusively looked at the persons" faces. In contrast, if the task was to indicate the wealth of the owner of the living room, participants looked more frequently at the furniture. Therefore, gaze control is highly goal driven and changes with task and context affordances. An overview of the physiological and psychological aspects of eye movements can be found, for instance, in Dodge, (1903), Findlay and Gilchrist (2003), Land and Tatler (2009), and Palmer (1999).

1.1.1 Eye movements in natural tasks

In the past decades, eye movements were investigated exclusively in laboratory settings and the research questions addressed visual processing of isolated, arbitrary stimuli. However, the inherent function of vision is selecting and processing important visual information for natural, everyday activities. It is questionable if the outcomes of laboratory studies also hold true in the real world. „t Hart, et al. (2009) showed for instance that eye movement characteristics differed between laboratory and natural settings, even if the same visual input was provided. The authors recorded eye movements as well as head-centered videos while participants were engaged in real-world activities. Other participants were asked to watch these head-centered videos, either continuously or as sequences of one-second frames, while their gaze was recorded. Results revealed that the scan path in the natural setting was more similar to the continuous than to the one-second frame observation condition. In addition, fixation locations were biased towards the centre in both laboratory settings. Thus, the dynamics of real world scenes cannot be simulated completely by real-world video presentation and even less by real-world picture viewing.

Hayhoe, Shrivastava, Mruczek, and Pelz (2003) argued that nothing can compensate for the exploration in the real world because of the infinite spatial scale and the three-dimensional depth information of real world exploration. In the course of this conclusion and with the emergence of mobile eye-tracking systems, researchers began to investigate eye movements during the execution of natural tasks such as tea making (Land, Mennie, & Rusted, 1999), sandwich making (Hayhoe et al., 2003), and car driving (Land, & Tatler, 2001). Thereby, important principles about the control of visual selection in natural tasks have been revealed (for a review, see Land & Tatler, 2009). As an example, the own hands or moving objects in the hands are hardly ever fixated on (Hayhoe et al., 2003; Johansson, Westling, Bäckström, & Flanagan, 2001; Land & Hayhoe, 2001; Land & Tatler, 2009). This finding illustrates how the investigation of gaze in natural tasks extends laboratory experiments. However, what can eye movement studies tell us about the attentional processes of the performer? The next section on attention deals with this question and will outline how eye movements and visual attention are interrelated.

1.2 Attention

In 1890, William James noticed that an object is more likely processed if it had been attended than if it had not been attended. Attention is necessary for multiple sensory and cognitive control mechanisms because of the limited capacity of human information processing. Humans are confronted with a huge amount of input from all their senses, but they cannot extract and process all available information in parallel within their processing limit (Cowan, 2000). Hence, efficient human information processing should ensure that the most relevant sensory input is processed, so that it will be available for upcoming tasks (Pashler, Johnston, & Ruthruff, 2001). This mechanism of selecting the most relevant sensory input from the environment among a variety of sensory inputs is called selective attention. Attention is selective in that only the limited amount of the available sensory input, namely the attended, is processed, while the unattended sensory information is not and gets lost. The decision of what will be attended is based on the one hand on bottom-up saliency of the material and on the other hand on top-down relevance for currently active goals of the individual person (Bundesen & Habekost, 2008).

Attention can also be divided between multiple sensory inputs as long as the processing limit is not exceeded. The processing depth of the attended material is dependent on the depth of attention allocation. In addition, attention cannot only be allocated to external sensory input from the environment, but also to internal representations, for example, memory structures or mental images (for a review on attention, see Chun, Golomb, & Turk-Browne, 2010). Attention is needed for every sensory input available to the five modalities of vision, hearing, touch, smell, and taste. However, the present work is mainly concerned with visual attention.

1.2.1 Visual attention

Visual attention is restricted to the visual modality. Visual attention can be directed to a location in space without the necessity to move the eyes to that location (Helmholtz, 1910). Shifts of attention without moving the eyes are called covert shifts of attention because the human focus of attention cannot be inferred from visible “overt” behavior. However, attention can be allocated to the current point of fixation and this symbiosis of covert and overt attention is usually the case in real-world interactions (Findlay & Gilchrist, 2003). This observation raises the question whether covert attention and eye movements are independent processes that accompany each other because they are usually initiated simultaneously or if there is an obligatory relationship between covert and overt attention. In 1978, a study by Posner revealed that participants could saccade faster to a location if a stimulus had cued this location previously than if no such cue was available. In this paradigm, Posner (1978) considered the cue as a prime for spatial attention and postulated a relationship between covert visual attention and eye movements. However, Posner‟s (1978) finding and its interpretation did not answer the question how obligatory the relationship between covert and overt attention might be.

The first empirical evidence for a tightly coupled relationship between attention and saccades originated from Deubel and Schneider (1996). In their study, participants had to discriminate a visual target while preparing a saccade. Discrimination performance was worse if the discrimination and the saccade target differed than if the same target was important for both tasks. The authors concluded that the allocation of covert attention for perception and discrimination and the selection of the next saccade target are determined by the same mechanism (Deubel & Schneider, 1996; Schneider, 1995). Therefore, it seems that the covert allocation of attention to a location in space is necessary to perform a saccade. Since then converging evidence demonstrated (e.g., Findlay, 2009) that saccadic control depends on covert attention. However, the shift of attention to a specific location does not necessarily result in a saccade to that location. A separate go-signal decides whether a saccade will be performed or not. Accordingly, the allocation of attention to a spatial location does not require saccades, while programming a saccade seems to require a prior covert shift of attention.

In conclusion, attention and eye movements are interrelated, i.e., a location of the environment has to be attended before it can be saccaded to. Therefore, the eyes mirror the attentional processes, revealing how attentional resources are applied to the environment during the execution of a task. Investigating eye movements while participants perform a specific sensorimotor task thus provides insights into the attentional processes during task execution. Not only eye movements and attention shifts should be important for sensorimotor control, but also memory processes. On the one hand, selected information has to be maintained until motor outcome. On the other hand, memory from prior task executions might be used to improve attention control during later repetitions, especially in everyday sensorimotor activities.

1.3 Memory

The human ability to memorize important information is an important necessity to perform sensorimotor tasks. Memory is defined as a collection of internal represented information built up by repetitive activation of neuronal circuits (Hebb, 1949) and stored as interrelated nodes in cortical regions of the brain. However, several distinct storage systems have been conceptualized in order to cope with the different memory characteristics and memory contents. Atkinson and Shiffrin (1968) differentiated between three memory systems according to the permanence of stored information: an iconic store, a short-term memory (STM) and a long-term memory (LTM). Internal representations can survive for milliseconds in the iconic store, for seconds to few minutes in STM, and for a lifespan in LTM. However, while the iconic store and LTM are not limited in capacity of stored information, STM can hold only about four to seven units of information.

1.3.1 Working memory (WM)

An additional short-term store, the working memory (WM), was postulated by Baddeley (1986; Baddeley & Hitch, 1974) as a gateway between STM and LTM. WM was conceptualized as an active memory system that processes and maintains stored information. The multi-component model of WM (Baddeley, 1986, 2000, 2012; Baddeley & Hitch, 1974) consists of passive stores and an active control system. One passive store, the phonological loop, deals with verbal information and another passive store, the visuospatial sketchpad, deals with visual and spatial information. The active control system is called the central executive and manipulates incoming and stored information. In 2000, Baddeley added the episodic buffer as an additional passive store that combines memory information from different sources to coherent episodes.

WM is closely connected to attention (for a detailed description, see Awh, Vogel, & Oh, 2006). First, the limited capacity of the short-term store necessitates attentional selection of external information. Only the representations of attended information will gain access to WM and can thus be maintained (Awh et al., 2006). Second, attention probably also underlies the rehearsal process that keeps representations active in WM (e.g., Awh, Jonides, & Reuter-Lorenz, 1998; Theeuwes, Kramer, & Irwin, 2011; Smyth, 1996; Smyth & Scholey, 1994).

1.3.2 Visual short-term memory (VSTM)

Concerning the internal representation of visual information, the concept of visual short-term memory (VSTM) – that equates at least partly to the visuospatial sketchpad - was used to explain how visual information survives the execution of an eye movement. VSTM was originally postulated by Phillips (1974) to explain maintenance of visual information above the sensory iconic storage but beneath the verbal short-term storage. Relatively abstract, schematic representations of visual information can be stored in VSTM up to its capacity limitation. About four objects can be maintained in VSTM during the execution of an eye movement (Irwin, 1992; Luck & Vogel, 1997; Pashler, 1988). Interestingly, far more than four features such as color, shape, or orientation can be kept in VSTM across saccades as long as these features belong to only about four objects (Irwin & Andrews, 1996; Luck & Vogel, 1997). Recently, it has been debated on whether establishing recognition of objects across fixations is a functional role of VSTM (Hollingworth, Richard, & Luck, 2008). An excellent review on visual memory systems is provided by Luck and Hollingworth (2008).

1.3.3 Long-term memory (LTM)

Long-term memory (LTM) retains information in the form of long-term synaptic weights in cortical regions of the brain (Olivers, Peters, Houtkamp, & Roelfsema, 2011). The LTM has been subdivided into separate stores according to its content. Most prominent is the taxonomy of a declarative (or explicit) and a non-declarative (or implicit) memory system (Schacter, 1987). Declarative memory can be further divided into semantic and episodic memory while perceptual memory, procedural memory, and priming are defined as non-declarative (Squire, 1982; Tulving, 1995). LTM has several important functional roles for sensorimotor tasks. With respect to the present work, it is important that LTM has a key role within most concepts of automatization as will be described in the following section.

1.4 Automatization

Automatization is becoming automatized in a specific operation, so that processing is automatic. Definitions of automatic processing are numerous and highly debated. A consensus might be that automatic processing can be characterized as efficient and effortless, i.e., maximal performance can be achieved with minimal effort. In contrast, more effort is usually needed to achieve a comparable high performance during non-automatic processing. Detailed definitions of automatic versus non-automatic processing as well as the mechanisms underlying automatization vary heavily. An excellent review and criticism was recently provided by Saling and Phillips (2007). Importantly, the authors noted that automaticity was often used to explain observations that did not fit to a specific theory. As a result, definitions of automaticity are often circular, i.e. "a*utomaticity is typically defined in terms of the very behavior it seeks to explain*" (Saling & Phillips, 2007, page 1). In the following, the most prominent theories of automatization will be outlined and compared with a main focus on their conceptualization of memory and attention during automatic versus non-automatic processing.

1.4.1 Schneider and Shiffrin"s (1977a, 1977b) two-process theory of human information processing

The two-process theory of human information processing, developed by Schneider and Shiffrin (1977a, 1977b) differentiates between automatic and controlled processes. According to the theory, an automatic process is initiated automatically based on externally or internally generated input. The input activates the automatic process through a set of associative connections in LTM without attention or active control, e.g., by the WM. Once established, an automatic process is difficult to modify or to suppress. In contrast, a controlled process is temporarily activated based on control and attention of the subject. Controlled processes are capacity-limited and can therefore only be activated serially or with reciprocal interference. The advantage of controlled processes is that they are easy to set up and to modify for other purposes. Interestingly, Schneider and Shiffrin (1977a, p. 51) allow automatic processes to attract attention, though this is normally not required. Concerning the development of automatic detection in visual search tasks, Schneider and Shiffrin (1977a, 1977b) suppose two prerequisites: consistent mapping of stimuli to be either targets or distractors and a large number of search repetitions. These prerequisites enable participants to learn to direct their attention to relevant features that are characteristic for the targets. Afterwards, attention is automatically directed to both the representation of the visual input in STM and the representation of the particular memory set target in LTM. Moreover, participants learn to detect automatically if a target is amongst the visual input, and, in some situations, an additional automatic overt motor response is also triggered (Schneider & Shiffrin, 1977b, p. 153). Although the two-process theory assumes that specific input automatically triggers specific LTM sequences, indirect control is allowed through threshold tuning for the activation of automatic processes (Schneider & Shiffrin, 1977b, p. 156). Thus, a wide range of input stimuli will trigger an automatic process if the activation threshold is low, while only specific input stimuli will trigger the same automatic process when the activation threshold is high. Unfortunately, little information was provided by Schneider and Shiffrin (1977a, 1977b) about the mechanisms that cause threshold tuning. Note also that Schneider and Shiffrin (1977a, 1977b) are inconsistent in their definition of automatization. In the theoretical framework of the two-process theory of human information processing, an automatic process is defined as an automatic activation of a LTM sequence by a specific input without active control and attention by the subject (Schneider & Shiffrin, 1977a, p. 2, 1977b, p. 155). Automatic detection in visual search tasks is referred to as an example for an automatic process (Schneider & Shiffrin, 1977a, p. 3). At the same time, automatic detection is described as a case when a stimulus gives rise to an *automatic-attention response* (Schneider & Shiffrin, 1977b, p. 143). However, the two-process theory of human information processing has found prominent advocates (e.g., Neves & Anderson, 1981; Treisman & Gelade, 1980) and is frequently applied to explain differences in human behavior (e.g., Cole & Schneider, 2007; Glöckner, & Betsch, 2008; Hofmann, Gschwendner, Frise, Wiers, & Schmitt, 2008). Moreover, some researcher still characterize automatic processing as uncontrollable and capacity-free (e.g., Küper, & Heil, 2010; Palmero, & Rhodes, 2006).

1.4.2 Neumann"s (1984, 1990) direct parameter specification theory

While supporting the qualitative distinction of automatic and non-automatic processes, Neumann (1984, 1990) argues contrastingly with Schneider and Shiffrin (1977) that automatic processes need intention and attention, and that they interfere with other processes, albeit to a smaller quantitative degree than non-automatic processes. Neumann thereby reconsiders an idea of Wilhelm Wundt (1903) that automatization is not independent of control, but the underlying control mechanisms are at levels below conscious awareness. Automatic processes have to be distinguished nevertheless from reflexive processes that are purely involuntary bottom-up driven actions. Even if attention might be directed automatically to relevant locations, as assumed by the *automatic-attention response* of Schneider and Shiffrin (1977), this response should not be free from interference. According to Neumann (1984, 1990), a sensorimotor skill and not just a separated process or processing stage is automatized. For Neumann"s (1984, 1990) parameter specification, action-relevant information has to be specified either in advance or during the execution of an action. Parameter specification is a necessary prerequisite to perform an action because an action comprises several processes that can be executed in different ways. The process of identifying an apple within a fruit bowl before grasping it, for instance, could be accomplished based on color, shape, or size information. Which information shall be used has to be specified before any classification process can start. Sources for parameter specification are threefold. First, the LTM contains procedures for skills. Second, the environment delivers the sensory input. Third, additional attention mechanisms are applied when links between sensory input and skill information do not suffice. Skill information in LTM can specify parameters directly and can contain the information of which stimuli in the environment are important for parameter specification. In car driving, for instance, the movement parameters necessary to release the gas pedal with the foot can be directly extracted from LTM of experienced drivers. However, sensory input from road signs and the speedometer is needed to specify whether pushing the gas pedal is the adequate action at a specific moment in time. LTM may again contain the information where the road signs and the speedometer are located in the environment. Nevertheless, attention has to be directed to the road signs and the speedometer before information can be extracted. Likewise, attention has to be directed to the gas pedal before pushing it. Additional attention is needed to provide concrete specifications if not all action parameters are specified (underspecification). Moreover, additional attention is necessary to select one of multiple available specifications (overspecification). In the car driving example, attention is needed to make a decision about which direction to take at an intersection (underspecification), unless the sensory input tells the driver that entering a particular street is prohibited, or LTM tells the driver which road leads to the intended goal. Attention in this example may be directed to a road sign or a road map containing information about the city. In addition, in case of mutually exclusive information on road signs (overspecification), selective attention will be utilized to bring about a decision of which road sign to comply with. Following Neumann (1984, 1990), a process is automatized if LTM and sensory input are sufficient for parameter specification, which is called the direct parameter specification. In contrast, non-automatic processes need additional, attentional mechanisms. Moreover, automatization is the acquisition of skill information or, in other words, the transformation of action-relevant information to LTM. Nevertheless, the necessity of sensory input survives for automated processes. Importantly, automatization is an emergent property that requires practice and task consistency.

1.4.3 Logan"s (1988, 1990, 1992) instance theory of automatization

Logan (1988), in agreement with Neumann (1984, 1990), rejects the idea that automatic and controlled processes can be differentiated through rigidly opposing characteristics (Schneider & Shiffrin, 1977) and advocates for a distinction between automatic and non-automatic processes based on the degree to which processing is LTM-controlled. According to Logan (1988, 1990, 1992), knowledge is represented as instances in LTM. An instance is defined as a representation of a single, prior task exposure. Automatic performance is characterized by direct retrieval of past solutions stored as instances in LTM. Conversely, general algorithms are applied for the execution of new tasks that cannot be related to any stored instances. Referring back to the driver example, when having to decide which way to take at an approaching intersection, an algorithm has to be used if the route is unknown. The driver may, for instance, apply an algorithm to transfer the sensory input of a road map or even a compass into the necessary information to decide which road to take. However, if the route is well-known, the driver will very likely retrieve the correct way from LTM. Logan (1988, 1990, 1992), in agreement with Neumann (1984, 1990) assumes that allocation of attention is needed throughout all levels of experience. Attention is directed to relevant stimuli in the environment during non-automatic processing to ensure encoding of the applied algorithm into LTM. Hence, each practice trial of a specific task induces a new memory instance. The depth of attention allocation determines the quality of the LTM representation. During automatic processing, attention is still directed to relevant stimuli to encode a further instance. Additionally, attention allocation to sensory input automatically triggers memory retrieval of previously encoded instances. Thus, attention allocation does not differentiate between automatic and non-automatic processing but the consequence of attention allocation, namely encoding versus encoding and retrieval, differentiates between non-automatic and automatic processing, respectively. Logan (1988, 1990) assumes that each exposure triggers a race between all memory instances and the algorithm resulting in a final decision and thus an action. The race takes the current task goal into account and finishes as soon as an adequate action is retrieved either from an instance or from the algorithm. As all instances are racing with one algorithm, the probability that processing will be memory-based increases along with the number of stored instances. In 1992, Logan modified his theory by assuming two races. A first race between all stored memory instances finishes as soon as the first adequate instance is retrieved from memory. A second race runs between the fastest instance and the algorithm. Altogether, according to Logan (1988, 1990, 1992) automaticity is LTM-controlled processing and automatization is the acquisition of LTM instances. The prerequisites for automatization are consistency and practice. Stored instances are only useful in case of high consistency of task constraints during acquisition, so that encoded instances will again be useful for successful task execution. Practice increases the amount of instances available. This in turn heightens the probability that a retrieved instance will be appropriate and accelerates the speed with which an appropriate instance can be retrieved. As a result, performance will become faster and more accurate. In short, Logan conceptualizes automatization as a gradual transition from algorithm-based to memory-based task execution because of the accumulation of experienced instances.

1.4.4 Theories of automatization in comparison

The outlined theories of automatization agree that practice and consistency are necessary prerequisites to automatize processing (Table 1). In addition, there is agreement that automatization is LTM acquisition and that LTM contents specify automatic responses (Table 1). However, theories disagree in the representation format in LTM. While Schneider and Shiffrin (1977a, 1977b) as well as Neumann (1984, 1990) assume that memory is represented as traces, Logan (1988, 1990, 1992) postulates that separate instances are stored (Table 1). Both concepts, the strengthening of memory traces as well as the accumulation of memory instances, can explain more accurate and faster performance through automatization.

Fundamental disagreement between automatization theories can be observed with respect to attention allocation to sensory input (Table 1). Schneider and Shiffrin (1977a, 1977b) make contradictory statements about the role of attention to sensory input and during the process of automatization. In the outline of their theory, automatic processes are postulated to completely lack attention and the sensory input is seen to trigger memory retrieval automatically. However, later in their work, an *automatic-attention response* to sensory input is introduced. According to this extended concept, automatic processes need no intentionally controlled, conscious attention allocation to sensory input, but attention is automatically drawn to specific sensory input that is related to automatic processes. Concerning the role of attention allocation to sensory input, Logan (1988, 1990, 1992) claims that all sensory input is attended independently of the degree of automatization. The quality of attention allocation determines the quality of instance encoding. However, the instance theory does not specify the conditions causing more or less intense attention allocation and the representation of worse compared to better encoded instances in LTM. In later versions (e.g., Logan, 2002; Logan, Taylor, & Etherton, 1999), it is assumed that attention allocation is also based on a race just like the choice between instances and between instances and algorithms. The theory of visual attention (Bundesen, 1990; Bundesen & Habekost, 2008) is referred to as best describing the mechanisms of attention allocation to the sensory input. Neumann (1984, 1990) assumes that a varying number of attentional mechanisms are necessary depending on the available sensorimotor skill information and thus depending on the degree of automatization. Sensory input has to be attended in any case. In low automatized processing, additional attention has to be allocated to the sensory input to link it to the necessary actions. In highly automatized processing, linking sensory input with necessary actions is based on LTM information, a phenomenon that is called direct parameter specification.

Further disagreement exists concerning the role of sensory input (Table 1), regardless of whether attention has to be (automatically) allocated to sensory input or not. In Schneider and Shiffrin‟s (1977a, 1977b) automatization theory, sensory input should automatically trigger memory retrieval. The triggered memory trace contains the automatic response of the specific sensory input. Similarly, attention to sensory input triggers the retrieval of memory instances in Logan‟s (1988, 1990, 1992) theory. Instances are weighted with an algorithm and the winner determines the response. In contrast, Neumann (1984, 1990) assumes that attention allocation to sensory input is used to extract action-relevant parameters from the sensory input. LTM skill information links the extracted parameters with the stored LTM parameters. While LTM directly contains the automatic response in Schneider and Shiffrin‟s (1977a, 1977b) as well as Logan‟s (1988, 1990, 1992) theory, the response is specified based on LTM content and sensory in Neumann‟s (1984, 1990) theory (Table 1).

The comparison of the three automatization theories reveals that they differ mainly with respect to the assumptions about attention allocation after automatization (Table 1). Schneider and Shiffrin (1977a, 1977b) assume that attention allocation is not needed or automatically drawn to specific sensory input in automatic processes. In contrast, attention allocation is still needed in Neumann‟s (1984, 1990) and Logan‟s (1988, 1990, 1992) theory. Logan (1988, 1990, 1992) assumes that the same mechanism of attention allocation is active in non-automatic and automatic processing. Attention allocation to sensory input initiates memory encoding in non-automatic processing, while it triggers additional retrieval of memory instances in automatic processing. Neumann (1984, 1990) finally proposes that more attentional mechanisms are needed for non-automatic than for automatic processing. As research on automatic processing was strongly neglected in the last decades, the interplay of attention, eye movements, and memory processes during learning and automatization is still unclear.

Table 1: Theories of automatization in comparison

	Schneider and Shiffrin	Neumann	Logan	present work
Prerequisites	practice and consistency			practice and consistency
Definition	automatization is LTM acquisition			automatization is LTM acquisition
LTM representation	traces	Traces	instances	traces
attention to sensory input (SI)	no or automatic attention to SI	attention to every SI	attention to every SI	attention to relevant SI
role of sensory input (SI)	SI automatically triggers LTM retrieval	SI provides action-relevant features	SI automatically triggers LTM retrieval and instance encoding	SI provides action-relevant features
role of LTM	LTM determines response	LTM and SI determines response	LTM determines response	LTM and SI determines response
attention after automatization	no or automatic	more attentional mechanisms	unchanged	changed: more LTM-based

1.5 Eye movements, attention, and memory during learning

Understanding the interrelationship between eye movements, attention, and memory during learning is an important step towards the understanding of automatization. Some studies provided insights into the relationship of at least some of these components. Sailer, Flanagan, and Johansson (2005) for instance analyzed eye movements during learning of an arbitrary tool-cursor mapping task. Results revealed a transition from exploratory reflexive eye movements towards anticipatory target-directed eye movements. After the mapping was learned, targets were fixated just in the moment before their location information was necessary to perform the cursor movement. This gaze pattern has been called the just-in-time strategy (e.g., Hayhoe, 2000; Hayhoe et al., 2003; Land & Hayhoe, 2001; Land et al., 1999; Land & Tatler, 2009). Is the transition from reflexive to anticipatory just-in-time eye movements a general pattern of everyday sensorimotor learning? The investigated tool-cursor task used an arbitrary mapping between hand and cursor movements. This deviation from natural tasks may have caused gaze patterns to differ from usually arising gaze patterns during learning of everyday sensorimotor tasks. It is very likely that it was the arbitrary mapping that caused the reflexive saccades in the beginning, because the arbitrary mapping forced humans to act against their well-learned hand-to-effect mappings. In everyday sensorimotor tasks, the consequences of effectors on the environment are usually well-known.

In contrast to reflexive eye movements, the just-in-time gaze strategy is a dominant finding in well-practiced natural tasks such as tea making or sandwich making (e.g., Hayhoe, 2000; Hayhoe et al., 2003; Land & Hayhoe, 2001; Land et al., 1999; Land & Tatler, 2009). This strategy implies that humans make usually little use of their memory during the execution of well-known sensorimotor tasks. Instead, they seem to use the world as external memory (O‟Regan, 1992). However, other results indicate that memory comes into play along with specific task constraints (Droll & Hayhoe, 2007; Droll, Hayhoe, Sullivan, & Triesch, 2008; Mennie, Hayhoe, & Sullivan, 2006). In a virtual brick sorting task (Droll & Hayhoe, 2007), participants had to pick up specific bricks and to put them down on specific layers according to the bricks‟ features. If participants knew in advance which feature indicated how to sort the bricks, the relevant features were more likely stored to WM. However, if different randomly chosen features specified pick-up and put-down actions, the just-in-time strategy was applied.

Droll & Hayhoe (2007) interpreted from their results that memory is used if it is advantageous for task performance.

Another finding that reflects memory use under specific constraints is the phenomenon of look-ahead fixations. A look-ahead fixation is defined as a fixation that gathers important information for an action taking place several seconds after the fixation had been performed. Humans seem to perform look-ahead fixations to save re-fixations as well as to perform more accurate or faster re-fixations. In a model-building task (Mennie et al., 2006), for instance, target-directed saccades were more accurate after look-ahead fixations to the targets. This observed advantage by look-ahead fixations indicates that visual information was retained across saccades.

Thus, humans can either gather visual information just-in-time or use their visual STM that has been acquired through look-ahead fixations to control their movements in non-automatized tasks. In automatized tasks, visual LTM information acquired through prior practice is very likely an additional source of information. The decision between the opposing strategies is probably based on cost-benefit tradeoffs (Gray, Sims, Fu, & Schoelles, 2006). The main advantage of sensory strategies for motor control is the fact that sensory information is up-to-date and thus accurate. The main advantage of memory use for motor control seems to be that it saves re-encoding and is thus faster.

1.6 Purposes, Questions, Hypotheses, and Experiments

The present work aimed at better understanding the cognitive processes underlying automatization. In order to gain more insights into attentional control, eye movement patterns, and memory processes during and after sensorimotor learning and automatization, three experiments were conducted along three research questions and their accompanying hypotheses.

As was outlined above, the most prominent automatization theories differ mainly in their assumptions about attention allocation. None of them would predict a change of eye movement and attention control during sensorimotor learning and automatization. However, the recent findings that humans have the opportunity to use either sensory information or memory information to control their actions (e.g., Droll & Hayhoe, 2007; Gray et al., 2006; Mennie et al., 2006) demonstrate that task constraints determine attentional control strategies. The level of automatization is also a constraint that might be able to determine attention and gaze control.

Therefore, the first question of the present work was whether and how eye movements might change during learning and automatization of a sensorimotor task. If attention and gaze control shifts, it should be possible to assign different gaze patterns to the different stages of automatization during the learning process. The study reported in manuscript 1 (Foerster, Carbone, Koesling, & Schneider, 2011a) compared eye movement patterns in early and late stages of the learning process. Participants were trained for fourteen days in the speed-stacking task (also known as sport stacking). This task consists of a fixed sequence of stacking up and down pyramids of plastic cups as fast as possible and requires grasping, lifting, moving, rotating, and placing actions. Number, order, and direction of the stacking movements are predetermined.

The second question of the present work was if eye movements can be controlled by LTM information after automatization. Through training of a highly structured sensorimotor task, task-related LTM representations are probably built, so that attention and eye movement control might be based on LTM after intense training. If this hypothesis is correct, participants‟ eye movements should be very similar across task repetitions, even if visual information cannot be used to control for eye movements. Therefore, the investigation reported in manuscript 2 (Foerster, Carbone, Koesling, & Schneider, 2011b) compared eye movements in the dark with eye movements in the light, while the speed-stacking experts that had been trained in the first study performed the automatized stacking task.

The third question of the present work was whether the execution of a LTM-based automatized task interferes with internal attention and WM processes. The traditional theories of automatization propose a minor part of attention and WM for automatic processing. According to the theories, attention is only allocated - if at all - to sensory input of the environment. This external attention allocation to the environment suffices to trigger LTM retrieval and also directly action execution. Temporary storage within WM is denied. Contrastingly, recent research points to a contribution of attentional and WM processes for LTM retrieval (e.g., Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Wagner, Shannon, Kahn, & Buckner, 2005). However, how attention and WM processes precisely interact for LTM retrieval, e.g., during the execution of automatized tasks, is unclear. If the execution of automatized tasks requires internal attention and WM processes, they should interfere with parallel STM retention tasks that also need internal attention and WM processes. Participants in the study described in manuscript 3 (Foerster, Carbone, & Schneider, 2011c) were asked to maintain either verbal or visuospatial material in WM, while they had either to wait, to perform the automatized speed-stacking task, or to perform an automatized high-speed poem-reciting task.

2 Empirical Part

2.1 Manuscript 1

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision*, *11*(7), *9*, 1-16.

2.2 Manuscript 2

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X (2011). Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task. *Manuscript submitted for publication in Journal of Vision*.

2.2 Manuscript 3

Foerster, R. M., Carbone, E., & Schneider, W. X (2011). The interplay of attention, working memory, and long-term memory: An interference study with automatized tasks. *Manuscript submitted for publication in Journal of Experimental Psychology: Human Perception and Performance*.

2.1 Manuscript 1

“Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization”

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision*, *11*(7), *9*, 1-16.

*Co-author statements are attached.

Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization

Rebecca M. Foerster
Department of Psychology, Bielefeld University,
Bielefeld, Germany, &
Cluster of Excellence "Cognitive Interaction Technology",
Bielefeld University, Bielefeld, Germany

Elena Carbone
Department of Psychology, Bielefeld University,
Bielefeld, Germany, &
Cluster of Excellence "Cognitive Interaction Technology",
Bielefeld University, Bielefeld, Germany

Hendrik Koesling
Department of Neuroinformatics, Bielefeld University,
Bielefeld, Germany, &
Cluster of Excellence "Cognitive Interaction Technology",
Bielefeld University, Bielefeld, Germany

Werner X. Schneider
Department of Psychology, Bielefeld University,
Bielefeld, Germany, &
Cluster of Excellence "Cognitive Interaction Technology",
Bielefeld University, Bielefeld, Germany

Principles of saccadic eye movement control in the real world have been derived by the study of self-paced well-known tasks such as sandwich or tea making. Little is known whether these principles generalize to high-speed sensorimotor tasks and how they are affected by learning and automatization. In the present study, right-handers practiced the speed-stacking task in 14 consecutive daily training sessions, while their eye movements were recorded. Speed stacking is a high-speed sensorimotor task that requires grasping, moving, rotating, and placing of objects. The following main results emerged. Throughout practice, the eyes led the hands, displayed by a positive eye–hand time span. Moreover, visual information was gathered for the subsequent manual sub-action, displayed by a positive eye–hand unit span. With automatization, the eye–hand time span became shorter, yet it increased when corrected by the decreasing trial duration. In addition, fixations were mainly allocated to the goal positions of the right hand or objects in the right hand. The number of fixations decreased while the fixation rate remained constant. Importantly, all participants fixated on the same task-relevant locations in a similar scan path across training days, revealing a long-term memory-based mode of attention control after automatization of a high-speed sensorimotor task.

Keywords: natural tasks, attention, saccades, eye–hand span, sensorimotor learning, automatization, high speed, bimanual

Citation: Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision, 11*(7):9, 1–16, http://www.journalofvision.org/content/11/7/9, doi:10.1167/11.7.9.

Introduction

Humans have to covertly attend to a location before the eyes can be directed to it (e.g., Deubel & Schneider, 1996). These saccadic eye movements are performed several times per second to informative locations in the environment. It is well known that the process of "where to look next?" is strongly shaped by the current task (e.g., Yarbus, 1967). This task dependence has recently been studied in natural everyday tasks in real-world environments. Studies have, for instance, investigated tea making (Land, Mennie, & Rusted, 1999), sandwich making (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003), or car driving (Land & Tatler, 2001). Important new principles about the control of visual selection in these "natural" tasks have been revealed (for a review, see Land & Tatler, 2009). First, locations that are fixated most frequently are similar between agents. Second, agents rarely look at task-irrelevant areas. Third, agents select visual information just when they need it (Hayhoe, 2000), a pattern that Hayhoe et al. (2003) called "just-in-time" strategy. The idea included in this strategy is that

doi: 10.1167/11.7.9 Received October 26, 2010; published June 10, 2011

the world is used as external memory (O'Regan, 1992) to save capacity load instead of relying on memorized environmental information. Fourth, agents hardly ever look at their own hands or at moving objects in their hands (Hayhoe et al., 2003; Johansson, Westling, Bäckström, & Flanagan, 2001; Land & Hayhoe, 2001; Land & Tatler, 2009), and, fifth, agents' eyes lead their hands by approximately 1 s or less (Land & Hayhoe, 2001; Land et al., 1999; Land & Tatler, 2009).

Some principles hold across different natural tasks, while others seem to be more task and context dependent (Droll, Hayhoe, Triesch, & Sullivan, 2005; Yarbus, 1967). Droll et al. (2005) demonstrated that participants were more likely to detect a feature change in virtual bricks when the changing feature was relevant for the task at hand than when it was irrelevant. Furthermore, in brick sorting, participants often refixated on relevant information if one feature was relevant for brick pick-up and another for brick placement (just-in-time strategy). In contrast, participants made use of their working memory for relevant information if the same feature indicated both the pick-up order and the placement location. Such results indicate that the allocation of gaze in space and time is highly goal driven and changes with task and context affordances.

Concerning object manipulation in natural tasks, a few studies investigated eye movement strategies during learning and automatization of a novel task (e.g., Epelboim et al., 1995; Sailer, Flanagan, & Johansson, 2005). To understand how humans learn to adjust their attentional control, changes of eye movement patterns during skill acquisition are of particular interest (Hayhoe, Droll, & Mennie, 2007; Land & Hayhoe, 2001). In addition, insights into visual selection processes during learning and automatization can help understand and improve the learning process itself. Sailer et al. (2005) examined eye–hand coordination during learning of an arbitrary mapping task, in which forces and torques on a rigid tool were mapped to cursor movements on a computer screen. Their results suggest three stages of learning: a first initial exploratory stage of poorly controlled movements, a second skill acquisition stage of rapid improvement, and a third skill refinement stage of gradual improvement. However, this tool–cursor mapping task deviates from most daily tasks as it takes place in 2D on a computer screen and not in the real 3D world. In addition, the arbitrary mapping makes participants oppose their well-learned, common mappings. Therefore, stage one and part of stage two may be specific to this arbitrary mapping, as these stages reflect the processes of learning a new and uncommon relationship between movements and their visual consequences, which is common to tool use tasks. In other real-world visuomotor tasks, the effects of sensorimotor acts on objects are known. In this class of tasks, there might be no exploratory phase, but refinement in speed and accuracy may only occur. Changes of eye movements found during learning of a sequential tapping task (Epelboim et al., 1995) support this hypothesis. While tapping the same sequence of targets ten times, participants became faster and performed less irrelevant fixations. In the tenth trial, all target locations were sequentially fixated and empty locations were not selected anymore. Epelboim et al. (1995) concluded from this finding that the fixations to empty locations during the first trials displayed the process of searching for the next target location. In the last trials, target locations were known and could, thus, be fixated in succession without any sensory-based search process. Becoming more effective at a "natural" task resembles the change from conscious to unconscious execution (Land & Hayhoe, 2001), which is, in most cases, an important feature of automatization (Schneider & Shiffrin, 1977). Because Epelboim et al. changed target locations after ten trials of training, it is still an open question how eye movement patterns might change with more practice and an increasing level of automatization.

An important prerequisite for automatization of a task is not only practice but also its consistency (Logan, 1988; Neumann, 1984, 1990; Schneider & Shiffrin, 1977). The consistency of a specific task is determined by the consistency of its elements. Task elements are the manipulated objects, the executed actions, and the sequence in which specific actions are performed on specific objects. Here, the term action refers to Cooper and Shallice's (2000, 2006) motor response schemas, which contain a class of similar subordinate actions. Subordinate actions of a single motor response schema share the relationship between initial conditions, response specifications, sensory consequences, and response outcomes. Examples of motor response schemas are grasping, placing, pressing, and pushing. Across different trials of the same task, these task elements may remain constant or may vary. A single or multiple objects and actions—either identical or different—can be elements of a task. Moreover, the action sequence can be completely fixed, partly fixed, or variable. An action sequence is partly fixed, if some of its actions have a fixed position, while others are interchangeable. Action sequences can be fixed by instruction, by practice, or even by physics. Thus, sensory as well as long-term memory (LTM) information can specify the action sequence. However, the more fixed the task elements are, the more likely long-term memory can be used. This may explain why consistency facilitates automatization. An open question, however, is how eye movements are integrated in this relationship between task consistency, memory, and automatization.

In addition to the lack of studies on gaze strategies during learning and automatization, most studies of gaze in natural tasks have investigated self-paced tasks. Their goal is performing as accurately as possible and avoiding action errors without trying to maximize execution speed. To our knowledge, only few studies investigated eye movements under time pressure and no previous study analyzed gaze patterns in a high-speed sensorimotor task. The tool–cursor mapping study by Sailer et al. (2005) described above is one of the few studies with a speed instruction element. Irrespective of the fact that the task concerns tool use learning, it consisted of only three objects, namely, the 2D target, the 2D cursor, and the 3D tool. Furthermore, only one sub-movement had to be performed, namely, rotating the tool in the 3D world to hit the target with the cursor in the 2D world. Additionally, participants performed the task for about 17 min in a single day. In another study (Flanagan & Johansson, 2003), trials of a block-stacking task were compared with different movement speeds. Flanagan and Johansson (2003) found shorter time intervals between eye and hand arrival at relevant locations with time pressure than without. However, the authors did not explicitly analyze and discuss the demands of different movement speeds on attention control, since the main research issue of their study was a comparison of eye movements in action production and in action observation. Although participants become faster in the sequential tapping task of Epelboim et al. (1995), accuracy was the primary goal. Speed was not emphasized by the instruction. The available task completion time allowed 9 s for tapping a six-target sequence, 6 s for four targets, and 4 s for two targets. In sum, no previous study investigated the demands of a high-speed sensorimotor action on gaze control.

Finally, few studies conducted so far were concerned with visual guidance of bimanual sensorimotor control. Some studies investigated eye–hand coordination solely in tasks with one acting hand, for instance, in obstacle avoidance (Johansson et al., 2001), block stacking (Flanagan & Johansson, 2003), and target contacting (Bowman, Johannson, & Flanagan, 2009). Although other tasks had to be performed with both hands such as making a cup of tea (Land et al., 1999) or a sandwich (Hayhoe et al., 2003), no study investigated the similarities and differences of specific gaze strategies guiding either the right or the left hand. Moreover, the two hands were almost always engaged with the same object in these tasks. As Land and Hayhoe (2001, p. 3561) stated: "In a few cases the two hands had separate roles. Rarely this involved actions on different objects…." In fact, in the tea-making study, sequences were excluded from analysis if the two hands were engaged in different tasks at the same time (Land & Tatler, 2009, p. 86). Therefore, it is still an open empirical question how different movements of the two hands performed simultaneously on different objects are visually guided by only one gaze point at a time.

Altogether, to our knowledge no previous study has investigated eye movement patterns in a high-speed bimanual, sensorimotor task with fixed task elements. In addition, research concerning possible changes of visual selection during learning and automatization of natural tasks is limited. The sensorimotor task used in the present experiment is speed stacking (also known as sport stacking). Speed stacking consists of a fixed sequence of stacking up and down pyramids of plastic cups as fast as possible. The number, the order, and the direction of the stacking movements are predetermined. For example, a six-cup pyramid is stacked up using six interleaved cups by arranging the cups with both hands in such a way that three cups form the base, two cups are stacked up on the base, and the last cup is placed on top of the two (for an illustrative example, see Movie 1 or visit http://www.speedstacks.com/about/history.php).

Movie 1. A participant performing the speed-stacking task on the first, second, and last training days.

Studying eye movements in speed stacking has several advantages with regard to the four previously mentioned neglected issues. The fact that speed stacking is a largely unknown activity allows for recruitment of naive participants. Moreover, it is fast and easy to learn and to automatize. Therefore, the whole learning process can be investigated, from the first contact with the task until a high degree of automatization has been achieved. Furthermore, the task elements (object, action, and the order) of speed stacking are fixed, i.e., the task has a high degree of consistency. Furthermore, it is a task that can be executed at an amazingly high velocity, i.e., its 44 sub-movements can be accomplished within approximately 19 s by participants who were trained for 45 min a day over a period of only 2 weeks. In comparison, it took participants in Land et al. (1999) approximately 4 min to accomplish the 40 to 50 sub-movements required for tea making. Finally, the task involves simultaneous movements of the two hands on different objects. Speed stacking enables us to analyze the role of gaze during the execution of a bimanual, high-speed sensorimotor task in which objects are grasped, moved, rotated, and placed.

The present study focuses on four topics: First, which similarities and dissimilarities can be observed between self-paced tasks and high-speed tasks? This question relates to the four principles of visual selection in natural tasks, to fixation functions, and to eye–hand dynamics found in self-paced tasks. Second, how and where do the eyes select visual information for the two hands that have to manipulate different objects simultaneously? Third, how do processes of visual selection change during learning and automatization of a new sequential, high-speed sensorimotor task with fixed task elements? How do people, for instance, adapt their eye movement strategy to speed up the sensorimotor task? To be pressed for time implies the need for parsimonious information gathering, which may force a decrease in number and rate of fixations. The proportion of fixations related to different functions may change such as guiding versus monitoring (e.g., described by Land et al., 1999, see Discussion section). For the current experiment, participants were asked to practice the speed-stacking task for 45 min a day over a period of 14 consecutive days. Participants were instructed to perform the task as fast as possible, while their stacking performance and eye movements were measured.

Methods

Participants

Nine right-handed students from Bielefeld University, Germany, participated in the experiment. Participants' age ranged from 22 to 26 years with a mean of 25. All participants had either normal or corrected-to-normal vision, were naive with respect to the aims of the study, and were paid for their participation. None had tried the speed-stacking task before.

Apparatus

A mobile head-mounted SMI eye tracker (iView X HED) and speed-stacking equipment (cups, timer, and mat) were used. Speed-stacking cups are 7.5 cm wide and 9.5 cm high. The SMI eye tracker features two video cameras (one for the scene and one for recording the participant's eye), an infrared light source, and a dichroic mirror attached to a cycle helmet. The eye tracker recorded gaze positions of the right eye at 200 Hz using an infrared video-based system. The direction of the eye relative to the head was detected by capturing the center of the pupil and the corneal reflection. A scene camera recorded the participant's field of view. Gaze position was indicated by a red circle superimposed on the scene camera image. The resulting gaze video of the task performance was recorded at 25 Hz. Gaze position accuracy was approximately 0.5 degree of visual angle with a tracking resolution below 0.1 degree of visual angle. Participants were seated in front of a 100-cm-high table with speed-stacking equipment placed on it in a distance of approximately 30 cm. The speed-stacking task was performed in an area that was approximately 60 cm wide, 40 cm high, and 30 cm deep. The distance between participants' eyes and the cups varied from 20 cm to 50 cm during task execution. Speed-stacking velocity was measured by a speed-stacking timer and transferred to and stored on a laptop computer. The speed-stacking errors were annotated manually during the experiment.

Task

We report data obtained from the bimanual, high-speed stacking task, which had to be performed as fast as possible. The speed-stacking "cycle" consists of three sequences. First, a three-cup, a six-cup, and another three-cup pyramid had to be stacked up and then stacked down. Second, 2 six-cup pyramids had to be stacked up and then stacked down. Third, a ten-cup pyramid had to be stacked up and then stacked down (see movies).

Gaze calibration procedure

Before the start of the actual gaze measurement, we used a five-point calibration procedure. Participants were asked to sequentially fixate five 10-mm-diameter colored points on a cardboard box with a width of 60 cm and a height of 40 cm. One of the points was located at the center and each of the remaining four points was located in one of the four corners of the box. The viewing distance between the participants and calibration plane was 40 cm. Calibration accuracy was checked after each trial and the calibration was repeated when necessary.

Procedure

The experiment consisted of 14 consecutive training days of 45-min speed-stacking practice each. The experiment began with an initial speed-stacking video instruction of 25-min duration on the first training day. Afterward, the trials started and participants were instructed to stack as fast as possible. On days 1, 2, and 14, participants had to practice in the laboratory. On the remaining days, participants practiced at home. Each laboratory session was divided into 30-min speed-stacking practice without eye movements being measured, the calibration procedure, and 15-min eye movement recording. Speed-stacking performance measures were recorded by the experimenter on the laboratory days and by the participants themselves on the remaining days. Thus, speed-stacking times and error rates were measured throughout the whole experiment.

Analysis

The gaze videos of two trials per participant were analyzed frame by frame, one trial of the first training day and one trial of the last training day. For maximum comparability, each participant's fastest speed-stacking trial without errors was analyzed. To standardize gaze positions despite their varying absolute *x*- and *y*-locations within the video frames, the frame-by-frame analysis was based on the topological structure of the cup arrangement. To allow for the investigation of gaze positions depending on the temporal sequence of the speed-stacking task, despite the varying trial durations, we standardized the gaze analysis by dividing the task into 44 "object-related actions" (ORAs). According to Land and Hayhoe (2001), an ORA is an act that is performed on a particular object without interruption. In our case, an ORA was defined as stacking up or down a single cup or stack of cups to other cups or stacks. In addition, the two occasions when cups were rotated were also defined as ORAs. In contrast to ORAs in other tasks, speed stacking entails ORAs in which the right hand manipulates objects as well as ORAs in which the left hand manipulates objects. These ORAs are performed simultaneously with slight temporal delays, i.e., an ORA of one hand begins while an ORA of the other hand is ending and vice versa. The cup's starting configuration of each of the 44 ORAs was drawn schematically in Power Point slides. Fixations that were performed during an ORA were superimposed manually based on the video information. Each fixation was plotted into the corresponding ORA box at the corresponding location with respect to the cup arrangement, i.e., each ORA box afterward contained the position of every fixation that started during this ORA. The frame-by-frame analysis of one participant stacking up a six-cup pyramid is presented as an example in Figure 1. The trapeziums resemble the speed-stacking cups with the broader horizontal line marking the open side. The red circles symbolize the gaze points. In ORA 5, the right hand has to stack up the two upper cups from the three-cup pile to the second row, so that they rest on the middle cup and the right cup that was formerly the lowest cup in the three-cup pile. In ORA 6, the left hand has to stack up the upper cup from the left two-cup pile to the second row, so that it rests on the middle cup and the left cup that was formerly the lowest cup in the two-cup pile. Finally, in ORA 7, the right hand has to stack up the cup from the two-cup pile of the second row on the top of the pyramid. The boxes contain the cup's starting configuration of the present ORA and, at the same time, the end configuration of the previous ORA. ORA 6, for instance, begins when the configuration displayed in its box is reached (also see video frame) and ends when the configuration of box ORA 7 is reached (also see video frame).

The number of fixations per ORA, the fixation-associated hand, and the eye–hand span (time and unit indices, see below for definition) were enumerated and listed. Fixations that continued in other ORAs were counted only once. We defined the fixation-associated hand as the hand that reached a fixated location immediately before or after the fixation was made. This variable is independent of the hand that is active in the present ORA. As an example, in ORA 6 (Figure 1), the upper cup in the leftmost two-cup pile has to be manipulated by the left hand, while the fixations made during this ORA are clearly associated with the right hand. On the first training day (Figure 1, left), the fixated area in ORA 6 is the location in which a stack was previously placed with the right hand. On the last training day (Figure 1, right), the fixated area in ORA 6 is the location where a stack will be placed afterward with the right hand.

The eye–hand span is defined by the movement onset asynchrony between eye and hand movements given that both movements are directed to the same location in space. The eye–hand span can be measured as a time index or as a unit index. In the present study, the time index is called eye–hand time span and the unit index is

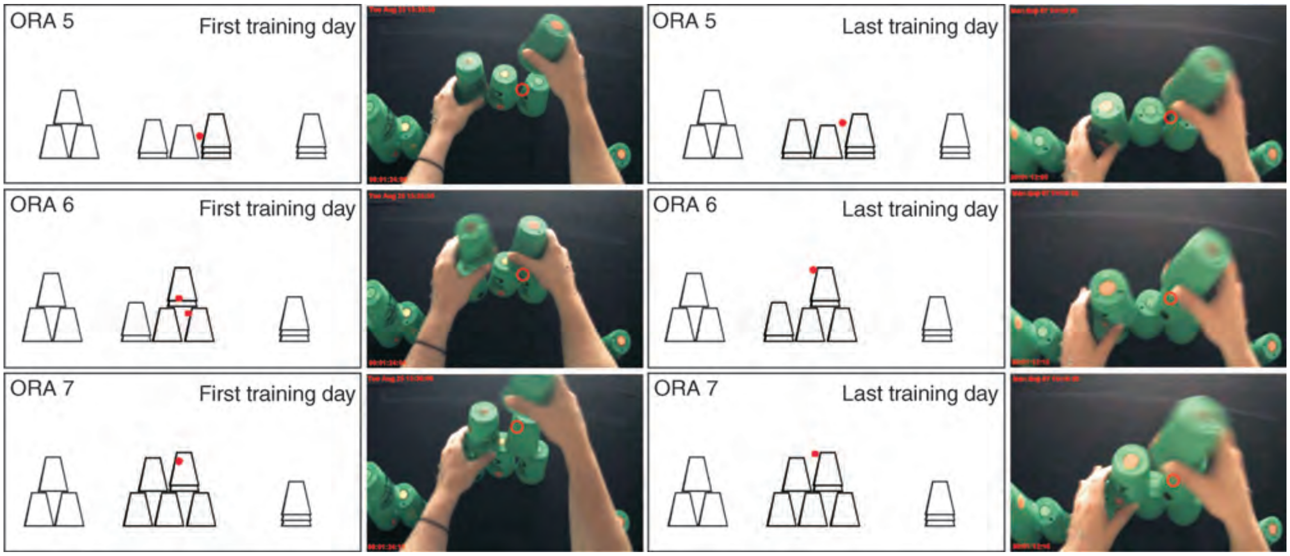

Figure 1. An example of ORA boxes for analyzing the gaze positions. The cup's starting configurations for ORAs 5 to 7 are represented in boxes and in video frames on the (left) first and (right) last training days. In ORAs 5 and 7, the right hand is manipulating a cup or stack. In ORA 6, the left hand is manipulating a cup. Each cup is illustrated as a trapezium with the long horizontal line as the open part of the cup. Additional horizontal lines near the open part of a cup illustrate a pile of cups. Each line corresponds to one cup. The boxes contain the cup's starting configuration of the present ORA and, at the same time, the end configuration of the previous ORA. The red dots represent the fixation locations of the participant in the interval between the start configuration of the present ORA and the start configuration of the successive ORA.

called eye–hand unit span. The eye–hand time span was defined as the time delay between gaze and cup in hand or the thumb landing at the same location. Locations were counted as the same if gaze and cup in hand/thumb lay within the half of a cup's height and width. Eye–hand time spans are positive if the eye reaches a location first and the hand follows. The eye guides the hand to a location. Eye–hand time spans are negative if the hand moves first and the fixation follows. The eye is driven by the location of the hand position. The eye–hand time span is commonly used in natural task approaches (e.g., Hayhoe et al., 2003; Land et al., 1999) and it is analog to the time index in music sight reading (Furneaux & Land, 1999). The eye–hand unit span is defined as the number of ORAs between the ORA in which gaze is directed at a specific location and an ORA in which a hand reaches this location. Eye–hand unit spans are positive if the fixation happens first and the hand follows. Eye–hand unit spans are negative if the hand moves first and the fixation follows. The eye–hand unit span is analog to the note index (the number of notes played after a specific note is fixated until the fixated note is played) in music sight reading (e.g., Furneaux & Land, 1999; Van Nuys & Weaver, 1943; Weaver, 1943) and the letter index (the number of letters typed after a specific letter is fixated until the fixated letter is typed) in typewriting (e.g., Butsch,1932; Hershman & Hillix, 1965; Shaffer & Hardwick, 1969). Finally, the *x*- and *y*-coordinates of each fixation with regard to the scene in the box were determined with millimeter accuracy within a graphics program (Microsoft Power Point). The left upper corner was the point of origin of the coordinate system. Coordinates were transformed into real-world coordinates and with the left lower corner as point of origin for further analysis. Interrater reliability on *x*- and *y*-coordinates of four trials (first and last days of the two fastest participants) analyzed by two independent data scorers revealed high consistency with Pearson's correlation coefficients ranging from 0.90 to 0.99.

In order to determine similarities of fixation sequences (the so-called scan paths) within and across participants, an action-sequenced linear distance method was used. This method is a combination of the minimum string-edit distance method (Brandt & Stark, 1977; Foulsham & Underwood, 2008; Levenshtein, 1966; Myers & Gray, 2010) and the mean linear distance method (Foulsham & Underwood, 2008; Henderson, Brockmole, & Castelhano, 2007; Mannan, Ruddock, & Woodman, 1995) that quantifies the scan path similarity. The action-sequenced linear distance method first assigns fixations to the ORAs in which they appear (Figure 1). Then, scan paths are compared according to the mean linear distances between its fixation locations within ORAs. In the present study, we computed between-training distances, between-subject distances, and random baseline distances. The random baseline distance is used to evaluate the size of the two experimental distances (the computation is analog to the method reported in 't Hart et al., 2009). In the first step, mean fixation locations were calculated for each participant's ORA for the first and the last training days, respectively (Figure 2a). The distance measures were calculated based on these averaged fixation locations. The between-training distance indicates scan path similarity across training days. It is the Euclidean distance between a participant's mean ORA fixation locations on the first training day and the same participant's mean ORA fixation locations on the last training day (Figure 2b). The between-subject distance indicates scan path similarity between participants. It is the Euclidean distance between mean ORA fixation locations of all participant pairs on the same training day (Figure 2c). The random baseline distance indicates random scan path similarity. It is the Euclidean distance between an observed and a randomly assigned mean ORA fixation location of a participant within the same training day (Figure 2d).

Using these action-sequenced linear distances to measure scan paths similarity has several advantages for the current study compared to the minimum string-edit distances or the mean linear distances alone. Mean linear distances (Foulsham & Underwood, 2008; Henderson et al., 2007; Mannan et al., 1995) are computed as precise Euclidean distances between nearest located fixations of to-be-compared paths. Unfortunately, no prior sequencing is performed in this method, i.e., identically located fixations performed in reverse order lead to maximal scan path similarity. Alternatively, the string-edit method (Brandt & Stark, 1977, Foulsham & Underwood, 2008; Levenshtein, 1966; Myers & Gray, 2010) categorizes fixations into labeled regions and calculates the minimum number of editing steps (insertions, deletions, and substitutions) needed to transform one fixation sequence into another. One disadvantage of this method is that it uses spatial regions instead of the precise *x*- and *y*-coordinates. The similarity index is, therefore, affected by the scale of single regions and by the placement of region borders. Thus, the comparison of fixations within a region leads to smaller similarity indices than the comparison of fixations across adjacent regions even if the absolute distance of the latter pair is smaller than that of the former pair. In addition, the similarity index reduces to the same extent if fixations are located in adjacent or distant regions instead of being located within the same region. A second and more important problem of the string-edit method for the present study is editing paths by deletions. When comparing sequences of different numbers of fixations, every deletion operation reduces the similarity index. During learning of the speed-stacking task, a performance speedup is expected, which will likely result in a decreased number of fixations on the last day. Nevertheless, similar locations might be looked at in a distinct order to perform the task. By assigning fixations to ORAs, it can be investigated whether similar locations are fixated within the same actions across expertise levels. As an example, we are not interested to know whether the tenth fixations of each day are similarly located but whether the fixations

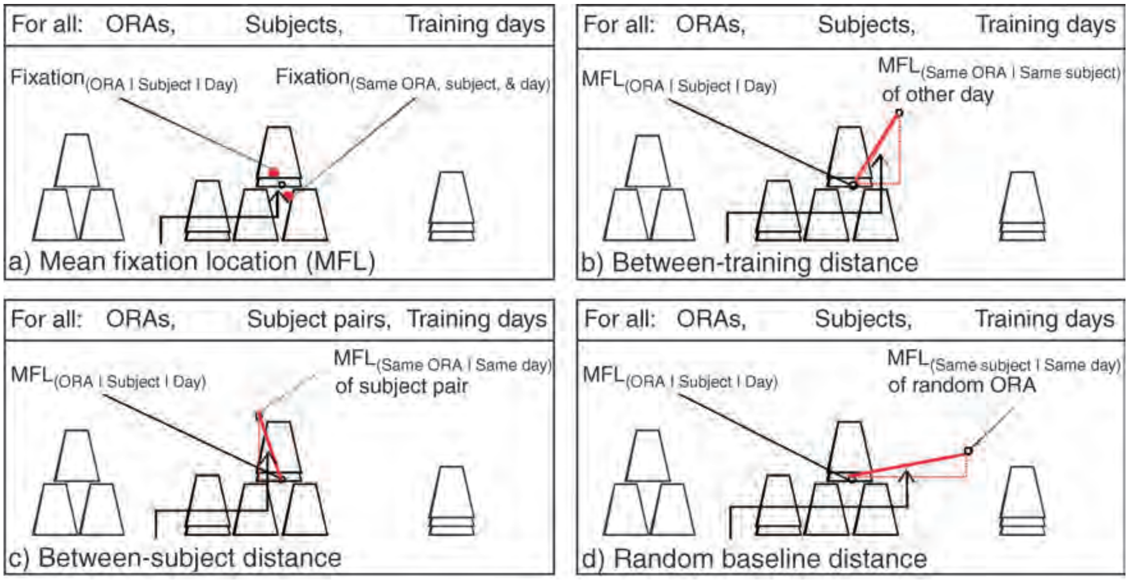

Figure 2. Schematic illustration of the calculations of (a) mean fixation location, (b) between-training distance, (c) between-subject distance, (d), and random baseline distance. (a) Mean fixation locations are the averaged fixation locations within the same ORA, subject, and training day. (b) Between-training distance is calculated between training days and within the same ORA and subject. (c) Between-subject distance is calculated between subject pairs and within the same ORA and training day. (d) Random baseline distance is calculated between random paired ORAs but within the same subject and training day. Cups and fixations are symbolized as in Figure 1. Averaged fixation locations of single ORAs are illustrated as black dots. Distances are illustrated as thick red lines. The figure contains no observed fixations as it serves only for illustrative purposes.

made during ORA 10 are similarly located across days. The present study is mainly interested in the similarity of action-sequenced fixation locations, indicating whether similar task-relevant points were fixated in the same sequence. This is conveniently measured by the action-sequenced linear distance method.

Design

The within-subject variables were the degree of speed-stacking experience (first day vs. last day) and the associated hand (left vs. right). The dependent variables were times and error rates of speed-stacking performance, as well as number, rate, location, and eye–hand dynamics of eye movements. The speed-stacking time was defined as the duration of a complete speed-stacking cycle. We defined a speed-stacking error as cups falling or sliding down (Movie 2). If an error occurred, participants had to correct it before continuing.

Movie 2. Exemplary errors of a falling and a sliding cup.

Results

Speed-stacking performance

Time

All participants learned the speed-stacking task as is reflected in the highly significant overall decrease of stacking time between the first (35.62 s) and last (18.56 s) training days [$t(8) = 10.01$, MSE = 1.70, $p < 0.001$]. Participants achieved a mean stacking time of 18.56 s with a mean best time of 14.05 s on the last training day. Because of the long-lasting practice of approximately 1300 trials per participant and the small increase in learning at later stages of training (Figure 3), the task can be executed with a high degree of automaticity on the last training day.

Error rate

Overall mean error rate was 43.20%. Unsurprisingly, error rates were high because participants were instructed

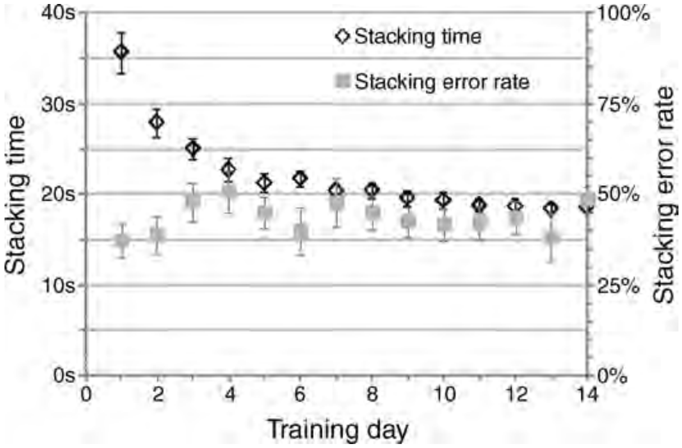

Figure 3. Mean speed-stacking time (dark gray diamonds and left *y*-axis) and error rates (light gray squares and right *y*-axis) with error bars indicating the standard error of the mean per training day.

to perform the task very quickly, regardless of accuracy. Error rates did not change significantly from the first to the last training day [$F(1,8) = 0.07$, MSE = 8.62, $p > 0.05$]. Mean stacking times and error rates per training day are depicted in Figure 3.

Gaze analysis

Gaze–hand coordination in object-related actions (ORAs)

We begin the description of the gaze results by presenting an exemplary ORA analysis with the help of three ORAs (5, 6, and 7). The description will reveal some of the general principles of natural task control, such as avoidance of effector-related fixations or hand guidance by the eye. We will show that these principles hold for the whole speed-stacking cycle and for all participants. Figure 1 shows the schematic fixation locations of one participant while performing the three consecutive ORAs 5 (upper part), 6 (middle part), and 7 (lower part) on the first (left) and last (right) training days. As mentioned before, these three ORAs belong to the upstacking of a six-cup pyramid. In ORA 5, the two upper cups of the three-cup pile have to be stacked up with the right hand to the second row of the six-cup pyramid so that they rest on the middle cup and the right cup that was formerly the lowest cup of the three-cup pile (see Movie 1). Achieving the configuration depicted in ORA 6 completes ORA 5. In ORA 6, the participant has to take the upper cup from the two-cup pile on the left side with the left hand and has to place it in the second row of the six-cup pyramid so that it rests on the middle cup and the left cup that was formerly the lowest cup of the two-cup pile. Finally, in ORA 7, the cup from the two-cup pile of the second row has to be stacked up on the top of the pyramid with the right hand. The fixations in Figure 1 illustrate that the participant did not track his own hand or moving cup during the task but looked at the goal position for the next action. This observation is quantified for all participants by the high percentage of positive eye–hand time spans (94.79%) and eye–hand unit spans (68.91%). When acting with the right hand, as, for instance, in ORAs 5 and 7, participants' fixations were associated with the right hand. This is quantified by the 64% right-hand-associated fixations. In contrast, only 36% were associated with the left hand, implying that participants were fixating less frequently on the location where the left hand had to place a cup, e.g., in ORA 6. Participants rather looked at that location where they were going to place the next cup with the right hand (62% positive right-hand spans). In summary, gaze led hand movements, the own hands were rarely fixated, and foveal information was extracted to guide the right hand but not the left hand. These results will be further quantified in the following sections and can also be observed in Movie 3.

Scan path similarity

In speed stacking, gaze was almost exclusively directed at task-relevant points—locations that contain important visuospatial information to perform the actions of the given task—such as the grasp area of cups and the target area where a cup had to be placed. Less than 0.01% of all fixations were directed at task-irrelevant points. Importantly, scan paths were highly similar between participants (Figure 4). For a statistical analysis of scan path similarities, we analyzed the calculated distances (see the Methods section). Mean between-training distance of 8.72 cm was significantly smaller than mean between-subject distance of 10.05 cm [$t(8) = 2.33$, MSE = 0.57, $p < 0.05$], indicating that scan paths were more similar across training days of the same participant than between participants within the same training day. In addition, both distance measures were significantly smaller than the random baseline distance of 23.39 cm [$t(8) = 13.76$, MSE = 1.07, $p < 0.001$ for between-training distance and $t(8) =$

Movie 3. Eye movements during the first, second, and last training days of a participant in slow motion.

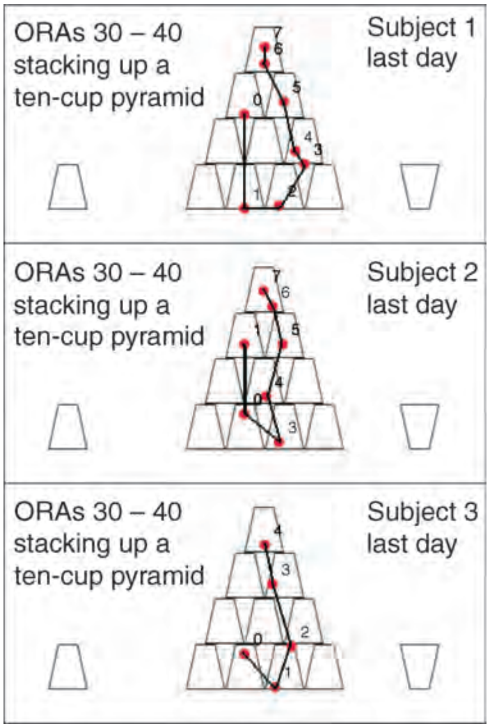

Figure 4. Scan paths of three different participants while stacking up the ten-cup pyramid out of a ten-cup stack on the last training day. Participants' fixations made during 10 successive ORAs (30 to 40) were superimposed on the schematic illustration of the upstacked ten-cup pyramid. Cups are illustrated as trapeziums and fixations are illustrated as red dots. Scan paths are indicated by numbers and black connection lines.

18.25, MSE = 0.73, $p < 0.001$ for between-subject distance], indicating that scan paths were similar across days as well as between participants.

Number of fixations

Participants made, on average, 78.5 fixations per speed-stacking trial, i.e., they performed 2 or less fixations per ORA. More fixations were made during the first (95) than during the last (62) training day [$t(8) = 4.70$, MSE = 7.03, $p < 0.01$], indicating that foveal information of fewer locations was used when performing the 44 ORAs with more experience. In contrast, the rate of fixations (the number of fixations during a trial divided by the speed-stacking time in this trial) did not change significantly between the first (3.43) and last (3.94) training days [$t(8) = 0.97$, MSE = 0.52, $p > 0.05$]. In addition, significantly more fixations were related to the right (23) than to the left (13.1) hand [$t(1,8) = 4.46$, MSE = 2.28, $p < 0.01$], suggesting that participants gathered more foveal information to guide the right hand than to guide the left hand.

Omitting fixations

The decrease in the number of fixations from the first to the last training day shows that some fixations were omitted on the last day. We examined these omitted fixations and categorized them according to their function. There were fewer fixations on the same task-relevant points on the last training day compared to the first training day and fixations to cups that had just been stacked were left out on the last training day (e.g., in ORA 6 of Figure 1, see also Movie 3). Moreover, fixations were more focused on specific task-relevant points on the last day. To illustrate this, Figure 5 shows all participants' fixations made for ORA 40 on the first (top) and last (bottom) training days. On the first training day, the two "outer" cups (Figure 5), which had to be grasped and rotated before they were used for downstacking the ten-cup pyramid, both presented a gaze target, at least for some participants. In contrast, participants less frequently fixated on these cups on the last training day. In addition, after having acquired a high degree of expertise, fixations on the pyramid were much more focused on the top cups, which had to be used for downstacking.

Eye–hand dynamics

The overall mean eye–hand time span was 423 ms ranging from −360 to 2600 ms with a standard deviation of 332 ms; 94.79% of all fixations had positive eye–hand time spans, indicating that gaze arrived at a location well before the cup in hand or the acting hand itself. Negative eye–hand time spans were observed in 0.92% of all fixations. These were performed by only two participants; 11% of one participant's fixations and 4% of the other participant's fixations had negative time spans. Both performed these fixations on the first training day. The

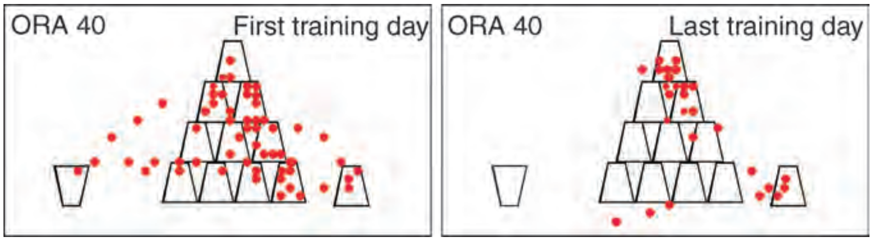

Figure 5. Fixations of all participants for ORA 40 on the (top) first and (bottom) last training days. Cups are illustrated as trapeziums and fixations are illustrated as red dots.

low percentage of fixations with negative eye–hand time spans indicates that hardly any checking fixations were used to assess hand movements. Furthermore, the few observable checking fixations occurred rather early in the learning process. The remaining 4.29% fixations were concurrent with the associated hand movement. We conducted a *t*-test for eye–hand time spans between the first and last training days. The analysis revealed a significantly longer eye–hand time span of 483 ms on the first day compared to 386 ms on the last day [$t(1,8) = 2.81$, MSE = 34.19, $p < 0.05$]. However, the size of the eye–hand time span depends on trial duration. Faster trials go along with shorter eye–hand time spans (Flanagan & Johansson, 2003; Furneaux & Land, 1999). In the present study, not only eye–hand time spans decreased from the first to the last day, but trial durations decreased as well. In order to determine whether the decrease of the eye–hand time span can be fully explained by the overall speedup in performance, we divided the mean eye–hand time span of each trial by its mean duration. This variable refers to relative eye–hand time span. If the decrease of the eye–hand time span can be fully explained by the speedup, the relative eye–hand time span should be constant across training days. A *t*-test was conducted for this relative eye–hand time span between the first and last training days. The analysis revealed a significantly higher relative eye–hand time span for the last (0.024) compared to the first (0.017) training day [$t(1,8) = 4.86$, MSE = 0.001, $p < 0.01$], i.e., the absolute eye–hand time span decreased to a lesser degree than trial durations. In contrast to the eye–hand time span, the eye–hand unit span does not depend on trial durations (Furneaux & Land, 1999). The eye–hand unit span specifies the number of ORAs performed after an eye movement until its associated hand movement is executed. The mean eye–hand unit span was 1.09 ORAs with a standard deviation of 0.93 ORAs, indicating that the visual information for the upcoming ORA was extracted and performance was dominated by a just-in-time strategy. A *t*-test was conducted for eye–hand unit spans between the first and last training days. The analysis revealed no significant difference of eye–hand unit spans between the first (0.99) and last (1.12) training days [$t(1,8) = 1.78$, MSE = 0.07, $p = 0.11$].

Discussion

A major aim of this study was to analyze eye movements during learning of a bimanual, high-speed sensorimotor task that required grasping, moving, rotating, and placing of objects and is performed with fixed task elements. Further, we were interested in how participants select visual information, provided that they had to manipulate different objects simultaneously with both hands and to perform the task as fast as possible. If automatization is characterized by a change of attention control, the relationship between attention and eye movements has to be specified. Visual selection can be performed overtly by an eye movement or covertly by a shift of attention without moving the eyes. Converging empirical evidence has demonstrated (e.g., Deubel & Schneider, 1996; Findlay, 2009) that saccadic control depends on covert attention. For instance, participants in Deubel and Schneider's (1996) study had to perform a perceptual discrimination task while they were preparing a saccade. Discrimination performance was heavily impaired if the discrimination task and the saccade had different target locations. It seems that the same mechanism that determines the allocation of covert attention for perception and discrimination also determines where to look next (e.g., Schneider, 1995; Wischnewski, Belardinelli, Schneider, & Steil, 2010). Therefore, the covert allocation of attention to a location in space should be necessary to perform a saccade. In addition, covert attention can be shifted without a subsequent eye movement (e.g., Posner, 1980). The present study is concerned with visual selection by saccades and examines whether and how this overt visual selection changes during learning and automatization. In the following parts of the Discussion section, visual selection processes in our task will be compared with visual selection processes in other self-paced natural tasks such as tea making and sandwich making. For this purpose, the results will be described according to the following issues. First, the five major principles derived from the investigation of gaze in natural tasks will be discussed with regard to speed stacking. Second, the present results will be linked to the four functions of gaze fixations in manipulation tasks proposed by Land et al. (1999). Third, we will analyze the asymmetries found in eye movements associated with left- and right-hand movements. Fourth, the dynamics relating the eye with the hand movements will be compared between different tasks. Fifth, we will contrast the just-in-time strategy with the working memory strategy of hand movement selection. Sixth, sensory-based and long-term memory-based eye movement selection will be discussed with regard to the role of fixed task elements in speed stacking (task consistency). Seventh, changes of visual selection during learning and automatization in the present task will be compared to a simple, single-step task (Sailer et al., 2005) and to a multi-step task with a short practice period (Epelboim et al., 1995). Eighth, we will derive task-independent conclusions concerning changes of overt and covert visual attention during skill learning and automatization. Finally, implications of our results will be outlined in relationship to theories of automaticity and attention.

Five major principle of eye movement control in natural tasks

The present results confirm the principles that have been derived from studying gaze in self-paced natural tasks without time pressure (Hayhoe et al., 2003; Johansson et al., 2001; Land et al., 1999). First, eye movements were highly similar between participants. It is important to note that not only fixation rate, fixation functions, and eye–hand dynamics were similar between participants but also the action-sequenced scan paths. The similarity of scan paths between participants as well as across training days was revealed by the small values of between-subject and between-training distances of fixation locations. They differed only about a cup's width and height. Thus, saccades tended, on average, to land on the same cup. In addition, both experimental distances were significantly smaller than a random baseline distance. Second, gaze was nearly exclusively directed at task-relevant areas. Third, selective vision followed the just-in-time strategy, indicated by the small positive eye–hand time spans and eye–hand unit spans. The eye–hand unit span reveals how many actions pass by after a fixation until that fixation is used to control a hand movement. This variable should be large if participants gather visual information far before they use it to control their hand movements, and it should be between zero and one if visual information is gathered just in time. In speed stacking, the eye–hand unit span was approximately one ORA on both training days, indicating that visual information was gathered just in time throughout practice. Fourth, acting hands or moving objects in hand were hardly fixated on, indicated by the small percentage of zero eye–hand time spans. Fifth, participants' eyes led their hands, reflected in the high percentage of positive eye–hand time spans. In summary, the five principles that were derived from studying self-paced natural tasks hold also for our high-speed bimanual sensorimotor task of speed stacking and they were not affected by learning and automatization.

Four functions of fixations in manipulation tasks

Land et al. (1999) proposed four functions of gaze fixations in manipulation tasks: locating, directing, guiding, and checking. The fixations on hand landing positions in speed stacking can be classified as directing fixations. The same pattern was found in Johansson et al.'s (2001) bar manipulation task but not in Hayhoe et al.'s (2003) sandwich-making task. Hayhoe et al. interpreted their results as evidence for foveal information being less critical for the control of placing actions. In contrast, the present results indicate that foveal information may be very critical to placing actions if the task demands fast and precise placing actions like in speed stacking. Alternatively, efferent gaze signals may be used in feedback loops to control hand movements without the need to extract foveal visual input. In our task, participants hardly performed any locating or guiding fixations or alternating fixations between approaching objects. In contrast, participants kept looking at hand landing positions. In self-paced tasks, fixations alternate between the approaching objects to perform the task as accurately as possible. This would be too time-consuming in high-speed tasks. Not surprisingly, hardly any checking fixations—with negative eye–hand time spans—were observed. Monitoring of successful movements is not functional for performance speed. Monitoring can be used to correct or prevent movement errors. However, in both cases, monitoring is an additional cognitive process that should decelerate performance. Since we analyzed accurate speed-stacking trials, the observed checking fixations must be concerned with error prevention. Only two participants performed checking fixations, all on the first day. It is possible that participants made more checking fixations directly after the video instructions during the very first trials. The participants might have needed monitoring to evaluate their performance in the beginning, before they learn that checking fixations do not help in realizing high-speed performance.

Hand asymmetry in eye movement control

We asked how a single gaze point is used to select visual information for the two hands in right-handers. The eyes do not select foveal visual information for both hands in an alternate fashion, as one might have expected. In contrast, foveal visual information is selected for the dominant right hand's landing positions but hardly for the non-dominant left hand's landing positions. Interestingly, the non-dominant hand could perform well, although it was not guided by foveal visual information of high resolution. The visual system may rely on peripheral vision to control the non-dominant hand and movements of the two hands may be planned and executed as a unit. This may be facilitated by the symmetrical task structure. However, it is an open question why participants decided to select foveal information of the dominant right hand's targets only.

Eye–hand dynamics

The eyes preceded the hand in speed stacking by approximately 400 ms, which is slightly shorter than the 560 ms in tea making but much longer than the 90 ms in sandwich making (Land & Hayhoe, 2001). Land and Hayhoe (2001) concluded from the dissimilarity between tea making and sandwich making that short eye–hand time spans only appear in the faster sit-down tasks. If the higher task speed was actually the only reason, then the eye–hand time span in the high-speed stacking task should have been even shorter. Eye–hand time spans were far longer in speed stacking than in sandwich making, although the latter task lasts for minutes while the former task lasts only for seconds. However, it is true that eye–hand time spans strongly depend on trial durations. This was verified by the longer eye–hand time span for longer trial durations in block stacking (Flanagan & Johansson, 2003) and sight reading (Furneaux & Land, 1999). It is difficult to decide what caused the 160-ms longer eye–hand time spans in tea making compared to speed stacking, as a speed-stacking trial is 12 times faster than a tea-making trial. A fair comparison between these two tasks could only be made based on the eye–hand unit span. As mentioned before, the eye–hand unit span is independent of trial durations as it counts the number of actions performed after a specific fixation until an action is performed on the fixated location. Comparing different tasks based on the eye–hand unit span could be an interesting topic for future research.

With more practice in the speed-stacking task, eye–hand time spans became shorter. However, when eye–hand time spans were normalized by division through their trial durations, the resulting relative eye–hand time spans became longer. Thus, the eye–hand time spans decreased less than what would have been predicted by the speedup in performance. In the following, three explanations for the increase of this relative eye–hand time span will be discussed. First, the absolute eye–hand time span might have reached a biological limit, in that the cognitive processing between visual input and motor output cannot be accomplished in less time. Then, the eye–hand time span might have stopped decreasing while the speedup in performance continued. The observation of far shorter absolute eye–hand time spans in sandwich making seems to contradict this explanation, yet the biological limit of eye–hand time spans might differ across tasks. Second, eye–hand time spans may decrease more slowly than trial durations. Third, eye–hand coordination might have become more dynamic. Relative eye–hand time spans would, for instance, increase if eye–hand cycles follow each other tighter after practice. This could be achieved either by shortening breaks between successive eye–hand cycles or by overlapping eye–hand cycles, where the next fixation is performed before the hand movement associated with the previous fixation is completed.

Sensory-based versus working memory-based hand movements

Humans can choose a capacity-saving just-in-time strategy or a more fixation-saving working memory strategy to guide hand movements (Droll et al., 2005; Hayhoe et al., 2003). When using the just-in-time strategy, participants extract sensory visual information just when they need it for hand movement execution. When using the working memory strategy, participants retrieve the relevant visual information from working memory. This is possible if the relevant visual information has been stored to working memory during prior fixations, the so-called look-ahead fixations. As each fixation needs time to be planned and executed, reducing the number of fixations by using working memory might speed up task performance. In high-speed tasks, it could, therefore, be advantageous to store relevant visual information in working memory for later hand movements instead of being forced to fixate a location again. On the other hand, working memory retrieval should be more error-prone than using the outside world as external memory (Gray & Fu, 2004; O'Regan, 1992). The just-in-time strategy has the advantage of gathering prompt, precise spatial information.

In speed stacking, the eye–hand unit span was close to one ORA, implying that visual information was extracted just when it was needed. This result indicates that participants used the just-in-time strategy not only in the beginning but also at the end of training. In speed stacking, the cup configuration changes rapidly and ORA relevant information cannot be extracted before a configuration provides this information. Thus, the necessity to update location information shortly before each ORA may have provoked the just-in-time strategy. In addition, refixations are less useful in speed stacking, as few locations specify more than one action.

Sensory-based versus long-term memory-based eye movements and the role of task consistency

Humans move their eyes to locations in the environment containing important information for the current task. However, both sensory and long-term memory (LTM) information may be used to select the saccade target. If an eye movement is directed to a location that has been extracted directly from the retinal input, the eye movement can be considered sensory-based. However, still the task determines which sensory information in the periphery is evaluated as important and, thus, will be fixated on. If an eye movement is directed to a location that has been stored in LTM, the eye movement is LTM-based. Therefore, in both cases, eye movements are controlled by the task in a top-down fashion. The decision between sensory-based versus LTM-based eye movement control may depend on the advantages and disadvantages of these control modes for the current task and context constraints. An advantage of sensory-based eye movements is the relatively high reliability of the outside world (Gray & Fu, 2004). In comparison, LTM information can only be encoded and stored in allocentric terms (object- or scene-relative) and may, therefore, be less accurate than egocentric retinal-based information. In addition, the environment can change, so that LTM information is no longer adequate. However, the resolution of spatial information in the periphery is probably worse than LTM information that had been encoded foveally. A further advantage of LTM-based saccades is that they should be less time-consuming than sensory visual selection as long as memory traces are strong. If memory traces are too weak and retrieval times are, therefore, relatively long (Gray & Fu, 2004), this advantage shall disappear or even be reversed.

A major prerequisite for a strong reliance on LTM information for eye movement control should be the consistency of task elements within and between trials as well as a high amount of practice. As defined in the Introduction section, a task consists of manipulated objects, executed actions, and, importantly, a sequence in which specific actions are performed on specific objects. In speed stacking, the same twelve cups have to be manipulated across trials and the cups have identical features. As a result, object features such as size, weight, or surface can be stored in LTM through practice. Moreover, the same set of motor response schemas such as grasp, rotate, and place has to be performed throughout the task. Most importantly, the action (ORA) sequence is fixed by task instruction and partly also by physics (top cups need a base to be placed on). Consequently, it is possible for participants to store the sequence of action-relevant locations in LTM through practice. After automatization, this action-relevant location sequence can be used to control eye movements, resulting in similar action-sequenced scan paths between participants. After speed-stacking automatization, it should be possible to initiate successive LTM-based eye movements, while hand movement control may depend, at least to a larger extent, on the time-consuming sensory just-in-time strategy for information extraction. Together, this may explain why the absolute eye–hand time spans were relatively large despite the high speed and short trial duration of speed stacking. In addition, this consideration would explain the increasing relative eye–hand time span on the last day, as it would lead to a tighter relation of consecutive eye–hand cycles after automatization.

However, if location sequences were stored in LTM during automatization, the following question arises: why did participants perform eye movements at all on the last training day rather than only direct the hands to the LTM-stored location sequence? For speed stacking, the answer is that participants needed to update the actual cup configuration just in time as the precise position of cups changes slightly from trial to trial. Fixations can reveal present deviations from LTM information, so that hand movement targets can be specified based on updated information. At the same time, the visual information can be used to update LTM information. If it would be possible to execute movements with marginal variation during trial repetitions, then LTM-stored locations could be used directly to specify hand movements with high precision. Interestingly, in tasks with even more fixed object locations, humans can perform an automatized task well without the necessity to move the eyes, e.g., playing a piece of music by heart.

Changes of attentional control during learning and implications for theories of automatization

We asked how visual selection changes during learning and automatization of a high-speed, sensorimotor task with fixed task elements. In contrast to the results of Sailer et al.'s (2005) study that revealed three stages of learning in an arbitrary cursor mapping task, we found only evidence for the last stage of skill refinement. A similar finding is reported by Epelboim et al.'s (1995) sequential tapping task. The number of fixations decreased with practice in all three tasks although the number of manual sub-movements could be reduced in the arbitrary cursor mapping task but not in speed stacking and tapping. In the bimanual stacking task, most fixations were associated with the right hand instead of the left hand, and this asymmetry did not change with expertise. In addition, the eyes led the hands already during the first training day of speed stacking and the first trial of tapping. In speed stacking, neither the absolute time index nor the unit index of the eye–hand span increased with practice. The change from negative to positive eye–cursor time spans in Sailer et al. (2005) may be a consequence of the arbitrary mapping. Participants seem to select visual information in advance even in new tasks if they know about their effectors' consequences, resulting in positive eye–hand time spans. Moreover, the same rate of approximately three fixations per second was maintained throughout the learning process. Perhaps, in natural tasks, the visual system of primates is limited to this maximal sampling rate that is determined by the minimal fixation duration needed to extract visual information.

We think that the reported results have task-independent implications of how covert visual attention and overt eye movements change during skill learning and automatization. Automaticity has traditionally been linked to attention. The two-process theory, most prominently advocated by Schneider and Shiffrin (1977), differentiates between automatic and controlled processes. Contrary to controlled processes, automatic processes are activated through long-term memory (LTM) and are performed without control, capacity, and attention. An alternative view (e.g., Logan, 1988; Neumann, 1984, 1990) characterizes the process of automatization by a change of attentional control. Following Neumann (1984, 1990), a sensorimotor skill is automatized if the conjunction of long-term memory skill information and sensory input is sufficient for parameter specification, while attentional selection is necessary for non-automatic processing. Extending and modifying Neumann's concept, we suggest that LTM information controls the attentional selection process for parameter specification (in the sense of Schneider, 1995) and determines which environmental information is relevant for movement parameter specification as well as where it can be extracted. A task-specific LTM representation should contain the sequences of task-relevant locations. Therefore, in our view, automatization in object-based sensorimotor tasks may imply a change of attentional control rather than its absence (Schneider & Shiffrin, 1977). After successful automatization, LTM structures may contribute substantially to the control of eye movements for actions. The selection of the next object for parameter specification—the information of where to look next for the eyes—should be guided to a larger degree by LTM information and to a lesser degree by sensory information. When reaching for a cup, its approximate location could be specified by LTM information, instead of being specified by peripheral sensory information. On the first training day of speed stacking, several fixations are used to guide the hands for a single ORA. There are not only fixations located on the target positions, but also further fixations located on positions in between the previous and next target positions. This result was also found in the first trials of the sequential tapping task (Epelboim et al., 1995). It is likely that participants have to shift their attention several times before the location that is important for the next ORA is found. Semantic LTM information built up during the task instruction probably determines an approximate region where the next relevant information has to be extracted from, e.g., on the left side. Therefore, a saccade is performed to a region outside the current visual field, increasing the possibility that the relevant location is available within the new visual field. However, a loop of more than one covert and overt shift of attention might be necessary until the relevant location is detected and fixated on. Then, the precise visual location information can be extracted and used to specify the parameters for the next hand movement. In contrast to this early stage of learning, participants may have built up a memory of location sequences on the last stacking training day and in the tenth tapping trial (Epelboim et al., 1995). This memory of location sequences is then used to guide the eyes directly to the next relevant location. This may explain the decreasing number of fixations during speed-stacking and tapping practice. It is important to note that this conception assumes that the change from sensory-based to memory-based selection contributes to the source specifying the relevant parameter for action control. The change does not contribute to the knowledge that parameter dimension has to be specified (e.g., location, shape, or color) for proper execution of the sensorimotor action. In addition, we think that the transition from a more sensory-based mode to a more LTM-based mode of attention control is gradual.

In summary, the present study addresses the question of how visual selection processes operate in bimanual, high-speed movements and how they change during learning and automatization. Results reveal similar scan paths between participants and across the learning process. In addition, the eyes lead the hands and are concerned with the upcoming action. Comparisons of eye–hand dynamics in high-speed tasks with those in self-paced tasks reveal similarities as well as dissimilarities. The eye–hand time span is longer in speed stacking than in sandwich making, although the latter task has longer trial duration. Eye–hand time spans are even longer in tea making than in sandwich making, but this may be caused by the fact that tea making is 12 times slower than speed stacking. It is difficult to infer what the eye–hand time span reveals about cognitive processes as this measure obviously not only depends on task speed. The eye–hand unit span reveals that the eyes gather visual information for the upcoming action both in the beginning and at the end of the learning process, a result consistent with the just-in-time strategy for movement control. As the eye–hand unit span is a valid measure to compare tasks with different trial durations, future research should investigate the eye–hand unit span supplementary to the eye–hand time span. Moreover, a right-side bias of foveal visual selection for bimanual movements has been found in our right-handed participants. Hence, sensorimotor control of the non-dominant hand may be based on peripheral vision. We would like to conclude that visual selection in high-speed sensorimotor tasks is parsimonious both in terms of number of fixations and working memory capacity and that automatization is characterized by a gradual transition from a more sensory-based to a more LTM-based mode of attention control.

Acknowledgments

This research was supported by a publication fund from Bielefeld University and by grants of the Cluster of Excellence Cognitive Interaction Technology (CITEC) at Bielefeld University. We would like to thank Thomas Hermann and Bettina Blaesing for their productive contributions to the speed-stacking project. Thanks are also extended to Okka Risius and Verena Donnerbauer who annotated the data for the interrater reliability. Finally, we would like to thank Wayne Gray and an anonymous reviewer for helpful comments on an earlier version of this article.

Commercial relationships: none.

Corresponding author: Rebecca Foerster.
Email: rebecca.foerster@uni-bielefeld.de.
Address: Department of Psychology, Bielefeld University, P.O. Box 100131, D-33501 Bielefeld, Germany.

References

Bowman, M. C., Johansson, R. S., & Flanagan, J. R. (2009). Eye–hand coordination in a sequential target contact task. *Experimental Brain Research, 195,* 273–283.

Brandt, S. A., & Stark, L. W. (1977). Eye movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 997–1010.

Butsch, R. L. C. (1932). Eye movements and the eye–hand span in typewriting. *Journal of Educational Psychology, 23,* 104–121.

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology, 17,* 297–338.

Cooper, R., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review, 113,* 887–916.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36,* 1827–1837.

Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1416–1438.

Epelboim, J., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C. J., et al. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research, 35,* 3401–3422.

Findlay, J. M. (2009). Saccadic eye movement programming: Sensory and attentional factors. *Psychological Research, 73,* 127–135.

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature, 424,* 769–771.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*(2):6, 1–17, http://www.journalofvision.org/content/8/2/6, doi:10.1167/8.2.6. [PubMed] [Article]

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye–hand span during music sight-reading. *Proceedings of the Royal Society of London B, 266,* 2435–2440.

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28,* 359–382.

Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition, 7,* 43–64.

Hayhoe, M. M., Droll, J., & Mennie, N. (2007). Learning where to look. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 641–659). Amsterdam: Elsevier.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*(1):6, 49–63, http://www.journalofvision.org/content/3/1/6, doi:10.1167/3.1.6. [PubMed] [Article]

Henderson, J. M., Brockmole, J. R., & Castelhano, M. S. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Compel, M. Fischer, W. Murray & R. W. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Amsterdam: Elsevier.

Hershman, R. L., & Hillix, W. A. (1965). Data processing in typing: Typing rate as a function of kind of material and amount exposed. *Human Factors, 7,* 483–492.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *Journal of Neuroscience, 21,* 6917–6932.

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41,* 3559–3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28,* 1311–1328.

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology, 11,* 1215–1220.

Land, M. F., & Tatler, B. W. (2009). *Looking and acting*. Oxford University Press.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics, Doklady, 10,* 707–710.

Logan, G. D. (1988). Towards an instance theory of automatization. *Psychological Review, 95,* 492–527.

Mannan, S., Ruddock, K. H., & Woodman, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9,* 363–386.

Myers, C. W., & Gray, W. D. (2010). Visual scan adaptation during repeated visual search. *Journal of Vision, 10*(8):4, 1–14, http://www.journalofvision.org/content/10/8/4, doi:10.1167/10.8.4. [PubMed] [Article]

Neumann, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 255–293). Berlin, Germany: Springer.

Neumann, O. (1990). Direct parameter specification and the concept of perception. *Psychological Research, 52,* 207–215.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46,* 461–288.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32,* 3–25.

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye–hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience, 25,* 8833–8842.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search and attention. *Psychological Review, 84,* 1–66.

Schneider, W. X. (1995). VAM: A neuro-cognitive model for visual attention control of segmentation, object recognition and space-based motor action. *Visual Cognition, 2,* 331–376.

Shaffer, L. H., & Hardwick, J. (1969). Reading and typing. *Journal of Experimental Psychology, 21,* 381–383.

't Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., et al. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition, 17,* 1132–1158.

Van Nuys, K., & Weaver, H. E. (1943). Studies of ocular behavior in music reading: II. Memory span and visual pauses in reading rhythms and melodies.

Weaver, H. E. (1943). Studies of ocular behavior in music reading: I. A survey of visual processes in reading differently constructed musical selections. *Psychological Monographs, 55,* 1–30.

Wischnewski, M., Belardinelli, A., Schneider, W. X., & Steil, J. J. (2010). Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. *Cognitive Computation, 2,* 326–343.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.

Universität Bielefeld

Erklärung zur Urheberschaft - Kumulative Dissertation von Rebecca M. Förster

Hiermit bestätigen wir, dass das Manuskript „*Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization*“ selbstständig von Frau Förster erstellt wurde. Frau Förster war hauptverantwortlich für die Planung, Durchführung und Auswertung dieser empirischen Arbeit. Außerdem verfasste und überarbeitete sie eigenständig alle Versionen des Manuskripts.

(Elena Carbone)

(Werner X. Schneider)

(Hendrik Kösling)

(Rebecca M. Förster)

2.2 Manuscript 2

“Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task”

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X (2011). Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task. *Manuscript submitted for publication in Journal of Vision*.

*Confirmation of submission, decision letter (minor revision), and co-author statements are attached.

Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task

Rebecca M. Foerster
Department of Psychology & Cluster of Excellence
‘Cognitive Interaction Technology’, Bielefeld University
Bielefeld, Germany

Elena Carbone
Department of Psychology & Cluster of Excellence
‘Cognitive Interaction Technology’, Bielefeld University
Bielefeld, Germany

Hendrik Koesling
Department of Neuroinformatics & Cluster of Excellence
‘Cognitive Interaction Technology’, Bielefeld University
Bielefeld, Germany

Werner X. Schneider
Department of Psychology & Cluster of Excellence
‘Cognitive Interaction Technology’, Bielefeld University
Bielefeld, Germany

Visual information for object-related natural tasks is usually selected by saccades. Nevertheless, humans can perform such a task in the dark provided it was automatized beforehand. It is largely unknown whether and how saccades are executed in this case. Recently, a long-term memory (LTM)-based mode of attention control during the execution of well-learned sensorimotor tasks was proposed [Foerster, Carbone, Koesling, & Schneider, 2011] that predicts task-relevant saccades in the dark. In the present study, participants performed an automatized speed-stacking task in the dark and in the light, while their eye movements were recorded. Speed stacking is a sequential high-speed sensorimotor task that requires grasping, lifting, moving, rotating, and placing of objects. Results demonstrated that participants indeed made systematic eye movements the dark. Saccadic scan paths and the number of fixations were highly similar across illumination conditions, while fixation rates were lower and fixation durations were longer in the dark. Importantly, the eye reached a location ahead of the hands even in the dark. Finally, neither eye-hand dynamics nor saccade accuracy correlated with hand movement durations in the dark. Results support the hypothesis of an LTM-based mode of attention selection during the execution of automatized sequential high-speed sensorimotor tasks.

Keywords: natural tasks, attention, visual selection, saccades, scan paths, eye-hand span, long-term memory, learning, automatization, high-speed

Introduction

The human eye has a rather small region on the retina where visual information can be processed with high resolution, the fovea. Therefore, humans move their eyes, head, and body to bring the image of informative stimuli on the fovea. Relevant information is then extracted from the foveal region and used for the current task. In object-related actions, relevant visual information is normally used to control eye movements as well as hand movements. For instance, when grasping an object, the eyes usually reach the location where the object has to be grasped sometime before the hand (e.g., Crawford, Medendorp, & Marotta, 2004; Desmurget, Pelisson, Rossetti, & Prablanc, 1998; Droll & Hayhoe, 2007; Droll, Hayhoe, Triesch, & Sullivan, 2005; Mennie, Hayhoe, & Sullivan, 2006). When placing an object, the eyes usually reach the placement location shortly before the object is set down (e.g., Ballard, Hayhoe, Li, & Whitehead, 1992; Droll & Hayhoe, 2007; Droll, et al., 2005; Epelboim, Steinman, Kowler, Edwards, Pizlo, Erkelens, & Collewijn, 1995). The same eye-hand dynamics have been found in well-learned everyday sensorimotor tasks such as sandwich or tea making (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land & Hayhoe, 2001; Land, Mennie, & Rusted, 1999; Land & Tatler, 2009) and throughout the learning process of a novel sequential sensorimotor tasks such as speed stacking (Foerster, Carbone, Koesling, & Schneider, 2011). Moreover, it has been shown that hand movements are executed more accurately if hand-target locations have been fixated beforehand (Prablanc, Desmurget, & Gréa, 2003; Prablanc, Echallier, Komilis, & Jeannerod, 1979; Prablanc, Pélisson, & Goodale, 1986). This “eye-guides-hand” benefit may explain the robustness of the finding that the eye leads the hands (positive eye-hand time spans) in sensorimotor tasks.

Apart from this finding of positive eye-hand time spans, Foerster et al. (2011) have found that there are also changes of gaze characteristics during learning and automatization. Participants in the study practiced the speed-stacking task for 45 minutes on 14 consecutive days. Speed stacking (also known as sport stacking) consists of a fixed sequence of stacking up and down pyramids of plastic cups as fast as possible. In addition, speed stacking has fixed task elements, i.e. it has a fixed set of objects to manipulate, a fixed set of actions to perform, and a fixed sequence in which specific actions have to be executed on specific objects. To compare gaze characteristics between low and high degrees of automatization, Foerster et al. (2011) recorded participant’s eye movements on the first and the last speed-stacking training day. On both training days, participants fixated on a location where a cup had to be placed shortly before the corresponding hand movement was initiated. However, on the last compared to the first training day, the eye preceded the hand by a shorter absolute time delay, but by a longer time delay relative to the overall stacking trial duration. Moreover, fewer fixations were needed to perform the sensorimotor task on the last training day and scan paths were highly similar between participants.

On the basis of these results, we suggested that the sequence of task-driven saccades might be learned and transferred to long-term memory (LTM) during automatization of sensorimotor tasks, provided that the sequence of object-related sub-actions is fixed. According to this idea, participants should rely on a more sensory-based mode of attention selection early in the learning process when no prior knowledge about objects, actions, and the action-to-object sequence is available. A sensory-based mode of attention control means that the next saccade target location is extracted from retina-based visual information. Later in the learning process, participants should rely on a more LTM-based mode of attention control. An LTM-based mode of attention control means that the next saccade target location is retrieved from LTM and that this retrieval process is guided by a stored scan path. This idea of an LTM-based mode of attention selection during the execution of an automatized sensorimotor task is tested in the present study. As a first step to examine this hypothesis, we will consider the functional role of eye movements when acting without sensory visual information, namely in complete darkness.

Object-related sensorimotor actions can be executed in the dark, especially if they are well-practiced. However, so far the “eye-guides-hand” benefits were always measured in the light (Prablanc et al., 1979; 1986; 2003). This study aims to clarify whether such a benefit also exists in the dark. This would imply that task-driven saccades to hand movement target locations are functional for object-related actions even in the dark. No “eye-guides-hand” benefits in the dark would be expected if the improvement of the hand accuracy exclusively depends on visual information that is gathered during fixation. In this case, saccades to hand-target locations in the dark would not be functional and task performance should be faster when no eye movements have to be planned and executed contemporaneously with the hand movements. Therefore, the eyes should not move during the execution of a sensorimotor task in the dark. In the light, humans fixate hand-target locations, although these locations often do not contain any visual object or salient feature. A reanalysis of the data reported in Foerster et al. (2011) revealed that more than 50 % of all fixations were directed to locations that did not contain any such visual object or salient feature. According to these results, one might suggest that the “eye-guides-hand” benefit does not depend on the availability of visual information of the hand-target locations.

If it is not visual information within a fixation, what else may cause the “eye-guides-hand” benefit? Flanagan, Terao, and Johansson (2008) proposed three possible mechanisms of how saccades to hand-target locations could be functional for hand movement execution when no visual input is available. First, sensorimotor transformations from fixated locations to hand movements may lead to better performance as they are well-practiced. Second, the computation of a hand motor command might be easier if a saccade to the same location has preceded it. Third, saccades to an action-relevant location may allow afferent and efferent signals of the current eye-ball position to be used as internal position marker for calculating a motor command. Thus, the improvement of hand movement accuracy after having fixated compared to having not fixated on hand-target locations might be inherent to the calculation of eye-hand transformations.

In favor of this assumption, Abrams, Meyer, and Kornblum (1990) observed that the accuracy of wrist rotations was better when participants could move their eyes than when eye movements were not allowed, although the moving wrist could not be seen in both conditions. On the basis of this finding, Rosenbaum concluded that the current eye-ball position could be used for hand movement planning (Rosenbaum, 2010). Computing target positions for hand movements may be more precise in the foveal range than in the periphery – independent of sensory visual information. If the “eye-guides-hand” benefit does not depend on visual information of the hand’s goal, performing task-relevant eye movements should be beneficial even in the dark.

How might saccades be controlled in the dark during object-related actions? Humans have to rely on sensory input other than visual information or on memory information to specify saccade target locations in the dark. If eye movements are controlled based on auditory feedback of objects in the dark, the eyes should be directed to a location where a sound has been recently produced, e.g., where two objects have recently contacted each other. Following the same logic, if haptic feedback is used to specify saccade target locations in the dark, the eyes should be directed to a location where a hand has recently contacted an object. For placing actions, this would lead to reversed eye-hand dynamics in the dark compared to those in the light, i.e., eye movements would follow hand movements instead of preceding them. Alternatively to sensory information, short-term memory (STM) information alone without the contribution of LTM motor chunks could be used to control eye movements in the dark. Before the light is switched off, visual information might have been extracted and stored in STM. This visual STM (VSTM) information might be used to select saccade target locations in the dark. Because of the capacity limitation of the VSTM, eye movement control based on VSTM in the dark should be restricted to the first three to four saccades. In complex sensorimotor tasks such as speed stacking, object configurations are changing in conjunction with task execution, so that the visual information extracted from the first action's start configuration in the light cannot specify saccade target locations for successive actions in the dark. Finally, if Foerster et al.'s (2011) idea of an LTM-based mode of attention selection in automatized sequential sensorimotor tasks is valid, it should be possible in the dark to select saccade target locations from LTM during the execution of such tasks with fixed sub-action sequences. As a result, scan paths as well as eye-hand dynamics should be very similar when performing an automatized sensorimotor task in the light and in the dark. However, this assumption does not imply a complete LTM-driven mode of selection in the dark, but allows for sensory-based corrections, e.g., by haptic feedback. Without sensory feedback, location errors of hand movements could increase along the scan path.

To our knowledge, only Flanagan et al. (2008) have investigated where people direct their gaze when performing a sensorimotor task in the dark. In their second experiment, participants had to reach for a bar, contact a support surface, and place back the bar in the dark at their preferred speed (for a detailed task description, see also Johansson, Westling, Bäckström, & Flanagan, 2001). The experiment consisted of eight object manipulation trials in the dark, four without an obstacle and four with an obstacle. Participants viewed the scene of the task and its objects for 3 seconds at the beginning of each trial. After the 3 seconds, an electric shutter was closed and a tone via earphones signaled that the object manipulation task had to be performed in the dark. Eye movements were recorded in the dark and compared with results of a previous study (Johansson et al., 2001) with the same object manipulation task in the light. On the basis of the findings of Prablanc et al. (1979, 1986, 2003), Flanagan et al. (2008) assumed that looking to hand-target locations in the dark might be useful because of well-learned eye-to-hand motor calculations, even though no visual information can be extracted. However, Flanagan et al. (2008) found mainly unsystematic eye movements during their object manipulation task in the dark, i.e., eye movements were dissimilar in several respects in the light and dark condition. Participants performed twice as many fixations in the light and the size of reaching errors in the dark did not increase with the distance between saccade landing position and target location. There was only a weak link between task phases (sub-actions) and eye movements. For instance, more fixations were located close to the bar during grasping than during lifting and targeting. The dissimilarity of eye movement patterns between light and dark condition and the absence of a correlation between saccade and hand movement accuracy in the dark condition of Flanagan et al. (2008) might support the interpretation that saccades to hand-target locations are no longer functional when no visual information is available. However, the object manipulation task analyzed in Flanagan et al. (2008) was not automatized beforehand. It was only performed eight times in the light (Johansson et al., 2001) before it had to be executed in the dark (Flanagan et al., 2008). Hence, these results cannot shed light on the hypothesis of an LTM-based mode of attention control during the execution of automatized sequential sensorimotor tasks.

The present study investigates eye movements and their relationship to hand movements while participants executed an automatized sequential object-related sensorimotor task in the light and in the dark. As the hypothesis of an LTM-based mode of attention control during automatized tasks was derived from the results with the speed-stacking task (Foerster et al., 2011), we chose the same task in the present investigation. The key objective was to clarify whether systematic eye movements are made in the dark. More specifically, we will address the following three research questions. First, are scan paths similar in light and dark condition? Second, does the eye lead the hand also in the dark? Third, are eye-hand dynamics and fixation locations in the dark related to the task performance? On the basis of our LTM-based mode of attention control, we expect similar scan paths and eye-hand dynamics in light and dark condition during the execution of the automatized sequential high-speed stacking task. Only well-practiced participants performed the speed-stacking task as fast as possible both in the light and in the dark, while their stacking performance and their eye movements were measured.

Method

Participants

A total of 7 right-handed students from Bielefeld University, Germany, with a mean age of 26 (*SD* = 1.38), participated in the experiment. All were highly trained in speed stacking (training is reported in Foerster et al., 2011). All participants had either normal or corrected-to-normal vision, were naive with respect to the aims of the study, and were paid for their participation.

Apparatus

Speed-stacking equipment (cups, timer, and mat), infrared light sources, and a monocular mobile head-mounted SMI eye tracker (iView X^{TM} HED) were used. The eye tracker features two head-mounted video cameras (one for recording the participants' right eye and one for recording the scene), infrared light source, and a dichroic mirror attached to a cycle helmet. To allow for gaze recording in the dark, the built-in infrared light source was replaced by an infrared light source with a wavelength range beyond the range visible for the human eye. This infrared light source was used to illuminate the participants' right eye. Additional infrared light sources were added to the scene, so that the SMI scene camera that records the participants' field of view could record the scene in the dark. Moreover, the SMI scene camera was modified by removing a built-in infrared filter. In the light, this infrared filter improves the image quality of the scene video, e.g., the brightness of colors. However, in the dark, this infrared filter would have interfered with the scene camera recording the IR-lit scene. The eye camera recorded gaze positions of the right eye at 200 Hz using an infrared video-based system. The direction of the eye relative to the head was detected by capturing the center of the pupil and the corneal reflection. Gaze position of the eye tracker was superimposed on the scene camera image and indicated by a red circle. The resulting gaze video was recorded at 25 Hz. Gaze position accuracy was approximately 0.5 degrees of visual angle with a tracking precision below 0.1 degrees of visual angle. Participants were seated in front of a table of 70 cm height, speed-stacking equipment placed on it at a distance of approximately 30 cm. The speed-stacking task was performed in an area of approximately 60 cm of width, 40 cm of height, and 30 cm of depth. The distance between the participant's eyes and the cups varied from approximately 20 cm to 50 cm during task execution.

Gaze calibration procedure

Before the start of the actual gaze measurement, the eye tracker was calibrated with a five point procedure in the dark. Participants were asked to sequentially fixate five 10-mm-diameter luminescent white stars on a 60 cm wide and 40 cm high cardboard box. One of the stars was located at the center, and each of the remaining four stars was located in one of the four corners of the box. The viewing distance of the calibration plane was 40 cm. Calibration accuracy was checked after each trial and the calibration was repeated if necessary.

Task

A speed stacking trial consisted of three sequences. First, a three-cup, a six-cup, and another three-cup pyramid had to be stacked up and then stacked down. Second, two six-cup pyramids had to be stacked up and then stacked down. Third, a ten-cup pyramid had to be stacked up and then stacked down (see Movie 1). Participants had to perform the task as fast as possible.

Procedure

The experiment started with a 30 minutes speed-stacking warm-up phase in the light. Afterwards, the light was switched off and the eye tracker was calibrated using the luminescent calibration stars. Calibration was checked, all remaining light sources were covered (control lights, computer screens, and the luminescent calibration stars), and participants had to perform the stacking task as fast as possible in the dark, while their eye movements were recorded. Stacking in the dark was repeated until participants had achieved at least five trials without errors (see design section for error definition). Finally, the light was switched on again, calibration was checked and repeated if necessary, and participants stacked in the light until they achieved at least one accurate trial, while their eye movements were recorded. An experimental session lasted for approximately one hour. Speed-stacking velocity was measured by a speed-stacking timer and stored on a laptop computer. Speed-stacking errors were annotated manually after each trial. In the dark, errors were reported by the participants. Error reports were checked based on the recorded gaze videos after the experiment.

Analysis

The gaze videos of one dark and one light trial per participant were analyzed frame-by-frame. For maximum comparability, eye movements during each participant's fastest speed-stacking trial without errors within a condition were analyzed. To standardize gaze positions despite their varying absolute x and y locations within the video frames, the frame-by-frame analysis was based on the topological structure of the cup arrangement. To allow for the investigation of gaze positions depending on the temporal sequence of the speed-stacking task, the action-sequenced analysis procedure reported in Foerster et al. (2011) was used. Following this analysis procedure, the task was first divided into 44 'object related actions' (ORAs). An ORA is an act which is performed on a particular object without interruption (Land & Hayhoe, 2001). Second, the schematic cup arrangement of each of the 44 ORAs was used for annotating the gaze positions manually. An annotation tool written in JAVA facilitated the frame-by-frame annotation process. The frame-by-frame analysis of one participant stacking up a six-cup pyramid is presented as an example in Figure 1. In ORA 39, the right hand has to stack up the upper cup from the two-cup pile to the top of the ten-cup pyramid. In ORA 40, the 'outer' cups have to be grasped (the right cup with the right hand and the left cup with the left hand), rotated by 180°, and tapped on the table. In ORA 41, the 'outer' cups have to be used to stack down the ten-cup pyramid from top to bottom by letting the cups fall into one another.

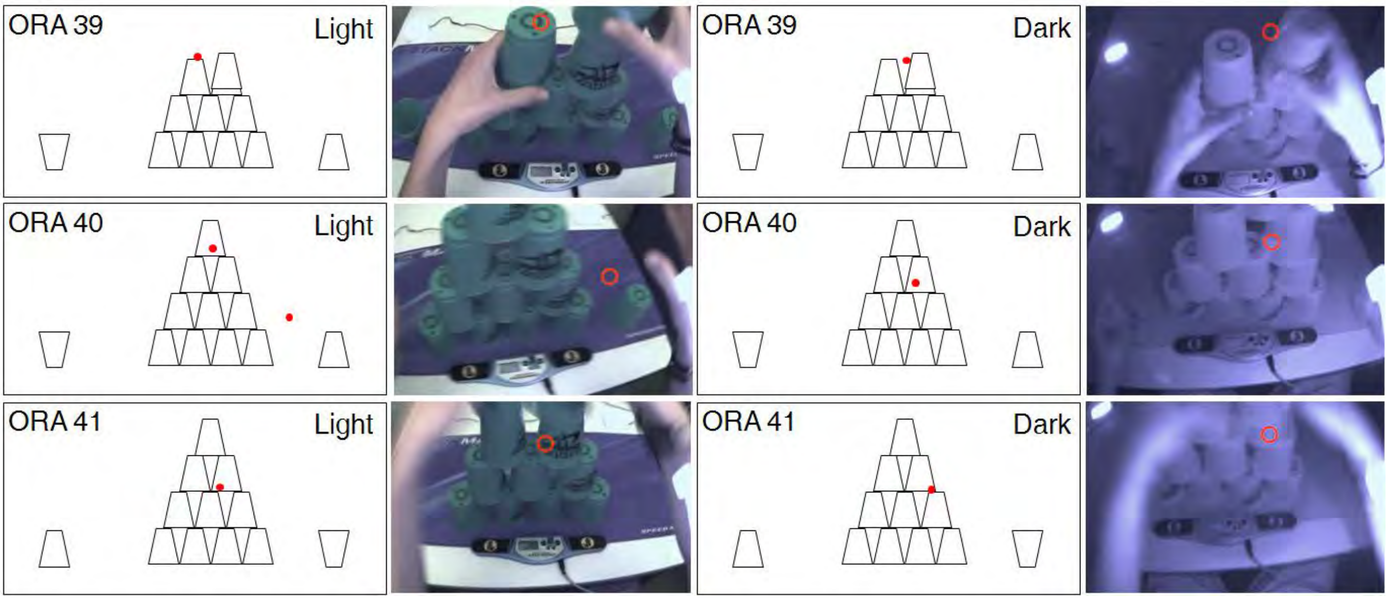

Figure 1. An example of ORA boxes for analyzing the gaze positions. The cups' starting configurations for ORAs 39 to 41 are represented in boxes and in video frames for the light (left side) and the dark (right side) condition. Each cup is illustrated as a trapezium with the long horizontal line as the open part of the cup. An additional horizontal line near the open part of a cup illustrates a pile of two cups. The boxes contain the cups' starting configuration of the present ORA and at the same time the end configuration of the previous ORA. The red dots represent the fixation locations of the participant in the interval between the start configuration of the present ORA and the start configuration of the successive ORA.

Based on the annotated video frame data, the number and duration of fixations, eye-hand spans (time and unit index, see below for definition), and saccade amplitudes were computed. Fixations were counted only once, when they continued in subsequent ORAs. The eye-hand span is defined by the movement onset asynchrony between eye and hand movement given that both movements are directed to the same location in space. The eye-hand span can be measured as a time index or as a unit index (Foerster et al., 2011; Furneaux & Land, 1999). As in Foerster et al. (2011), the time index was called eye-hand time span and the unit index was called eye-hand unit span. The eye-hand time span was defined as the time delay between gaze and cup-in-hand, or the thumb, landing at the same location. Locations were counted as the same if gaze and cup-in-hand/thumb lay within half of a cup's height and width. Eye-hand time spans are positive if the eye reaches a location first and the hand follows. They are negative if the hand moves first and the fixation follows. The eye-hand unit span is defined as the number of ORAs between the ORA in which gaze is directed at a specific location, and an ORA in which a hand reaches this location. Eye-hand unit spans are positive if the fixation is first and the hand follows. They are negative if the hand moves first and the fixation follows.

Finally, the x and y coordinates of each fixation with regard to the scene in the box were annotated with pixel accuracy. Pixel coordinates were transformed into centimeter coordinates for further analysis. Interrater reliability of two independent data scorers on x and y coordinates of four trials (light and dark trial of the two fastest participants) revealed moderate to high Pearson's correlation coefficients ranging from .71 to .99. In order to determine similarities of fixation sequences (the so-called scan paths) between the two conditions (light and dark), the action-sequenced linear distance method conducted by Foerster et al. (2011) was used. In the first step, this method calculates mean fixation locations for each participant's ORA separately for both conditions (Figure 2a). In the second step, Euclidean distance measures are calculated based on this mean fixation locations. In the present study, distances were calculated between light and dark condition (Figure 2b) as well as between observed and randomly shuffled fixation locations of the light condition (Figure 2c). The former between-condition distance (Figure 2b) indicates scan path similarity between illumination conditions. The latter random baseline distance (Figure 2c) is used to evaluate the size of the scan path similarity between illumination conditions (analog to the method reported in ‘t Hart, et al., 2009).

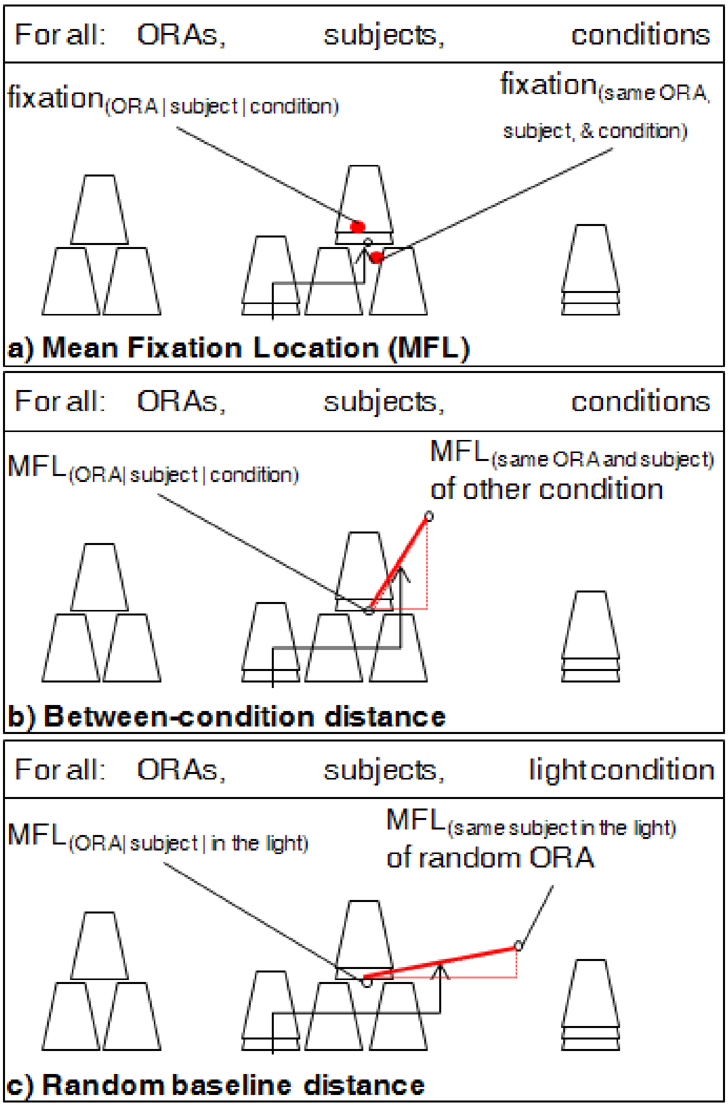

Figure 2. Schematic illustration of the calculations of (a) mean fixation location, (b) between-condition distance, and (c) random baseline distance. (a) The mean fixation location is the averaged fixation location within the same ORA, subject, and condition. (b) Between-condition distance is calculated between illumination conditions and within the same ORA and subject. (c) Random baseline distance is calculated between randomly paired ORAs within the same subject in the light. Cups and fixations are symbolized as in Figure 1. Each additional horizontal line near the open part of a cup corresponds to one further cup in the pile. Averaged fixation locations of single ORAs are illustrated as black dots. Distances are illustrated as thick red lines. The Figure does not depict observed fixations as it serves only for illustrative purposes.

The advantage of using action-sequenced linear distances to measure scan paths similarity across illumination conditions is that fixations belonging to the same sub-action are compared independently of their index number. In addition, different numbers of fixations within two comparable scan paths do not reduce the similarity index as long as fixations belonging to the same sub-action are similarly located. Moreover, absolute distances are computed instead of region compliance as the string-edit method does (Brandt & Stark, 1997, Foulsham & Underwood, 2008; Levenshtein, 1966; Myers & Gray, 2010). For a detailed comparison of the action-sequenced linear distance method and the string-edit method, see Foerster et al. (2011).

Design

The within-subject variable was illumination condition (light versus dark). The dependent variables were times and error rates of speed-stacking performance, as well as number, rate, and duration of fixations, eye-hand spans, and scan paths. The speed-stacking time was defined as the duration of a complete speed-stacking trial. We defined a speed-stacking error as cups falling or sliding down. If an error occurred, participants had to correct it before continuing.

Results

Task performance

All trials (accurate and erroneous trials) were used for the comparison of task performance between light and dark. Stacking was significantly faster in the light (19.50 s) than in the dark (46.89 s) [$t(6) = 6.25$, $MSE = 4.39$, $p < .01$]. Mean error rates in the light (35.54 %) were numerically, but not significantly smaller than in the dark (50.83 %) [$t(6) = 1.60$, $MSE = .10$, $p > .05$]. Mean stacking times and error rates for stacking in the light and in the dark are depicted in Figure 3.

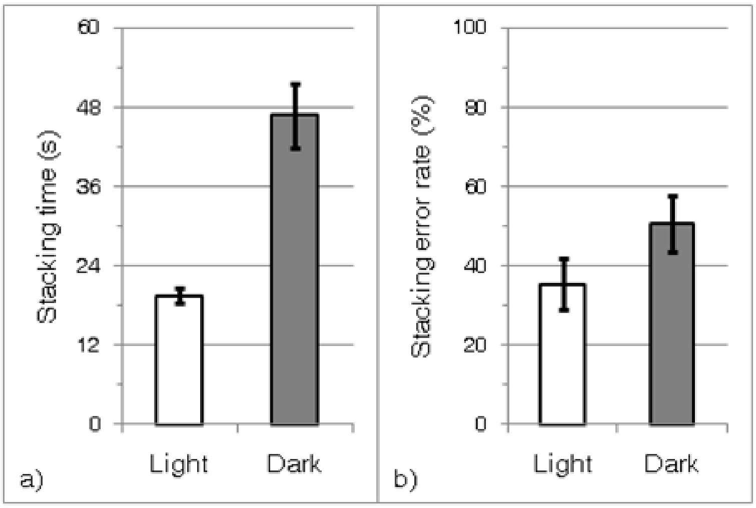

Figure 3. Speed stacking performance measures: (a) Mean speed-stacking time with standard error of the means in light and dark conditions. (b) Mean error rate with standard error of the means in light and dark conditions.

Eye movements

As mentioned in the analysis section, eye movements during each participant's fastest speed-stacking trial without errors per condition were analyzed in order to optimize comparability.

Fixations and saccades

Overall, the mean number of fixations for a trial was 31.86. There was no difference between the number of fixations in the light (31.86) and in the dark (31.86) condition [$t(6) = .00$, $MSE = 3.96$, $p > .05$]. On average, participants made 1.35 fixations per second. Fixation rate was significantly higher in the light (1.77) than in the dark (.93) condition

[$t(6) = 8.56$, $MSE = .10$, $p < .001$]. The mean fixation duration was 648 ms. Substantially shorter fixation durations occurred in the light (451 ms) than in the dark (844 ms) condition [$t(6) = 2.58$, $MSE = 152.66$, $p < .05$]. To analyze whether a similar number of fixations was made within the same ORAs, we analyzed Pearson's correlation of number of fixation per ORA in the light and in the dark. The correlation reached significance [$r = .82$, $p < .001$]. Saccades had a mean length of 19.53 cm and saccade lengths did not differ significantly between light (21.87 cm) and dark (17.13 cm) condition [$t(6) = 2.13$, $MSE = 2.23$, $p > .05$].

Eye-hand dynamics

Absolute eye-hand time spans did not significantly differ in light (399.32 ms) and dark (379.70 ms) conditions [$t(6) = .24$, $MSE = 82.50$, $p > .05$]. However, eye-hand time spans depend on trial durations (see Foerster et al., 2011; Furneaux & Land, 1999) and participants performed faster in the light than in the dark. Therefore, we conducted paired *t*-tests with relative eye-hand time spans (corrected by mean trial duration) and eye-hand unit spans as dependent variables. The relative eye-hand time span was significantly larger in the light (23.11) than in the dark (13.26) condition [$t(6) = 6.08$, $MSE = 1.62$, $p < .01$]. Accordingly, eye-hand unit spans were larger in the light (.85) than in the dark (.46) condition [$t(6) = 3.17$, $MSE = .12$, $p < .05$]. Thus, fixations were more often associated with the subsequent step (next ORA) in the light than in the dark, while fixations were more often associated with the current step (same ORA) in the dark than in the light. Importantly, eye-hand time and unit spans were positive across illumination conditions, i.e., the eyes reached a location prior to the hands even in the dark. In the light, 93.73 % of the eye-hand time spans and 47.19 % of the eye-hand unit spans were positive. In the dark, 82.27 % of the eye-hand time spans and 41.84 % of the eye-hand unit spans were positive. The lower percentage of positive eye-hand unit spans compared to relative eye-hand time spans in both illumination conditions is due to fixations with leading hand movements (positive eye-hand time spans) within the same ORA (zero eye-hand unit spans).

Scan path similarity

To quantify scan path similarity between illumination conditions, the between-condition distance was tested against the random baseline distance (see Method section). The between-condition distance (light versus dark) of 15.13 cm was significantly smaller than the random baseline distance (light observed versus light random) of 24.86 cm [$t(6) = 4.14$, $MSE = 2.35$, $p < .01$]. Thus, scan paths were similar across illumination conditions (see also Movie 1).

--------- Movie 1 about here ---------

Movie 1. A participant performs the speed-stacking task in the light (left) and in the dark (right) condition. To facilitate the comparison of eye movements across illumination conditions, the velocity of stacking in the light was dynamically adapted to the velocity of stacking in the dark.

Relations between Eye Movements and Task Performance

First, we asked whether the functionality of eye movements in the speed-stacking task is determined by the "eye-guides-hand" benefit. In self-paced reaching movements, spatial movement accuracy can be easily measured and is a central aspect of task performance. In speed stacking, hand movement accuracy is difficult to operationalize. Cups have to be stacked on top of two adjacent cups. Therefore, there is a relatively broad spatial area in which cups can be placed to accomplish the task. In addition, in speed stacking, speed is much more important than spatial movement accuracy. A relevant indicator of speed is hand movement duration. Because of these considerations, we decided to investigate the relationships of eye movements to hand movement durations instead of hand movement accuracy.

We analyzed whether eye-hand spans in the dark were related to hand movement durations in the dark. Therefore, eye-hand time spans as well as eye-hand unit spans were correlated with the duration of their corresponding hand movements in the dark. If large eye-hand spans were accompanied by short hand movement durations in the dark, this might be a hint for a non-spatial “eye-guides-hand” benefit in the dark. Pearson’s correlation of eye-hand time spans with hand movement durations in the dark was not significant [r = -.11, p = .33], nor was Pearson’s correlation of eye-hand unit spans with hand movement durations in the dark despite a tendency [r = -.34, p = .07]. Thus, a large delay between eye and hand was not associated with higher hand movement speed in the dark.

However, not only the temporal relationship between eye and hand might be important for the hand movement durations, but also the spatial relationship between eye and hand. The deviation of fixation locations in the dark from where participants normally fixate in the light might be related to hand movement durations in the dark. Therefore, mean linear distances of fixation locations between illumination conditions were correlated with mean hand movement durations in the dark. If small distances in fixation locations across illumination conditions were correlated with short hand movement durations in the dark, this might be a hint that spatial accuracy of fixations is advantageous for performing fast in the dark. However, Pearson’s correlation of between-condition linear distances with hand movement durations in the dark was not significant [r = .00, p = .50].

Discussion

In the present study, we asked whether systematic eye movements were made in the dark during the execution of an automatized sequential high-speed sensorimotor task - speed stacking. Our comparison condition refers to the same task in the light that already revealed systematic patterns (Foerster et al., 2011). More specifically, we compared scan path similarity between light and dark condition. Moreover, we wanted to know if the eye leads the hand even in the dark and if eye-hand guidance is beneficial for task performance in the dark reflected by a temporal “eye-guides-hand” benefit. Finally, we asked whether fixation location similarity across illumination conditions is beneficial for task performance in the dark. Based on the suggestion of an LTM-based mode of attention selection (Foerster et al., 2011), we expected similar scan paths in light and dark conditions during the execution of the automatized sequential high-speed stacking task.

The following major results emerged. Most importantly, systematic eye movements during automatized speed stacking were not only made in the light but also in the dark. Scan paths were highly similar in light and dark conditions. The relative eye-hand time span and the eye-hand unit span were smaller in the dark than in the light. However, the absolute eye-hand time span was similar across the two illumination conditions, i.e., the eye led the hands with a similar value of approximately 400 ms. More than 80 % of all eye-hand time spans were positive in both illumination conditions. Thus, the eye mainly preceded the hand movements not only in the light, but also in the dark. Speed-stacking trial durations were significantly faster in the light than in the dark, while error rates were only numerically smaller in the light. Nevertheless, participants performed on average a comparable number of fixations per trial in the light and in the dark. In compensation, they made more fixations per second in the light than in the dark. Consequently, fixation durations increased in the dark. There was a high correlation between illumination conditions concerning the number of fixations that were performed per object-related action (ORA). Hence, a sub-action unit guided by a large number of fixations in the light was associated with a comparably large number of fixations in the dark. Finally, neither large eye-hand spans, nor highly similar fixation locations across illumination conditions were correlated with fast hand movements in the dark.

The following discussion is divided into the following sections. First, we will discuss the possible mechanisms of saccade target selection in the dark. Second, we will consider the present results with respect to the theories of automatization. Third, we report further evidence for an LTM influence on attention allocation. Fourth, a possible coupling of eye and hand movements will be discussed. Fifth, we will propose mechanisms of the way that systematic eye movements in the dark might be beneficial for task execution, although not enhancing hand movement accuracy or speed.

Saccade target selection in the dark

The observation of positive eye-hand spans in sensorimotor tasks is a robust finding in the light (e.g., Foerster, et al., 2011; Hayhoe et al., 2003; Land & Hayhoe, 2001; Land, et al., 1999; Land & Tatler, 2009; Sailer, Flanagan, & Johansson, 2005). The eyes usually guide the hand movements in the light, especially in grasping and placing (e.g., Ballard et al., 1992; Crawford, et al., 2004; Desmurget et al., 1998; Droll & Hayhoe, 2007; Droll et al., 2005; Epelboim et al., 1995; Land & Hayhoe, 2001; Land et al., 1999; Land & Tatler, 2009; Mennie, et al., 2006). Directing the eyes to a hand-target location before initiating the corresponding hand movement is advantageous for the accuracy of the hand movements (Prablanc et al., 1979, 1986, 2003). However, it was not clear whether the eyes guide the hands only if visual information is available. We found positive eye-hand spans when speed stacking had to be performed in the dark. Participants saccaded to hand-target locations in the dark before moving the corresponding hand, although no visual input could have been extracted to specify the hand-target location. By implication, participants must have used other sensory input or memory information to specify where to look next in the dark.

In the speed-stacking task, sensory auditory and haptic input cannot specify hand-target locations until the corresponding hand movement is completed. At the moment a cup is placed on the ground or on other cups, an auditory as well as a haptic feedback signal is generated. If sensory auditory and haptic input had been used to specify saccade targets in the dark, the eyes would have followed the hands, so that negative eye-hand spans would have been observed. The fact that the eye movement preceded the hand movement excludes the possibility that eye movements were controlled based on auditory or haptic feedback signals. Therefore, memory information must have been used to control eye movements while performing the speed-stacking task in the dark. Two memory sources are available to specify where to look next in an automatized sensorimotor task: short-term memory (STM) and long-term memory (LTM). In the present study, participants had executed the speed-stacking task for 30 minutes in the light prior to the first trial in the dark. In addition, participants viewed the overall start configuration of the speed-stacking task before they performed the task in the dark. Did participants use STM information alone without contribution of LTM motor chunks to specify saccade-target locations in the dark condition? STM items can be stored for several seconds up to a few minutes (Atkinson & Shiffrin, 1968). The 30 minutes warm-up stacking phase in the light was followed by the eye tracker calibration procedure. It takes several minutes to calibrate the mobile head-mounted SMI eye tracker in the light and even more in the dark. Thus, it is very unlikely that STM information from the warm-up stacking phase was still available to control eye movements during speed stacking in the dark. In addition, the visual STM capacity spans approximately four items (Cowan, 2011). Speed stacking is a quite complex sequential sensorimotor task with 44 object-related actions (ORAs). The cup configurations in speed stacking are changing permanently, so that the visual information relevant for the next ORA is not available until its start configuration has been built up by previous ORAs. As only the start configuration of the very first ORA was viewed before the light was switched off, it is very unlikely that participants could use STM information from this very first start configuration alone without LTM contribution to control their eye movements during speed stacking in the dark.

Did participants use LTM information to specify the saccade-target locations in the present study? Participants had automatized the sequential high-speed sensorimotor stacking task (Foerster et al., 2011) before they were asked to perform the task in the dark. A high degree of automatization can be inferred from a long-lasting prior practice of approximately 1300 trials per participant and a small increase in performance at later stages of training (see, Foerster et al., 2011). As neither sensory nor STM information alone could be used to direct the eyes - prior to the hand - to hand-target locations in the dark, participants must have used LTM information to control their eye movements. Importantly, not only eye-hand dynamics, but also scan paths were very similar between illumination conditions. When participants performed the speed-stacking task in the dark, they seemed to retrieve successive saccade target locations - scan paths - from LTM.

How is saccade control related to attention allocation? Previous experimental research (e.g., Deubel & Schneider, 1996; Findlay, 2009, for a summary) revealed that overt eye movements ("where-to-look-next?") in space depend on covert attentional processes, i.e., a saccade location in space has to be selected covertly by attention before the eye will go to that location. Therefore, not only eye movements but also covert spatial attention must have been controlled based on LTM information. In other words, task-relevant information of 'where-to-attend-next' for specifying 'where-to-look-next' (see, e.g., Schneider, 1995) must have been retrieved from LTM. In conclusion, the results of the present study support the idea of Foerster et al. (2011) that attention selection in automatized sensorimotor tasks depends on LTM.

Implications of the present results for theories of automatization

The conception of automatic processing as a more LTM-based mode of attention control is opposed to traditional theories of automaticity (e.g., Schneider & Shiffrin, 1977a, 1977b) that deny the need for attention control during automatic processing. Alternative concepts of automatization allow for attention control during the execution of automatized tasks (e.g., Logan, 1988, 1990; Neumann, 1984, 1990). Logan's (1988, 1990) instance theory of automatization assumes that attention is directed to every sensory input. On the one hand, attention modulates the encoding of sensory input. On the other hand, attention to sensory input causes the activation of all corresponding LTM instances. According to the instance theory, acquiring LTM instances is substantial for automatization. However, attention control is quite similar before and after automatization because attention is used for selecting task-relevant sensory input among irrelevant input. Allocation of attention to sensory input does not differentiate between automatic and non-automatic processing. Instead, the consequence of attention allocation to sensory input, namely encoding versus encoding plus retrieval, differentiates between non-automatic and automatic processing, respectively. Neumann (1984, 1990) assumes that attention has to be allocated to relevant sensory input as well as to relevant LTM-stored content not only in automatic but also in non-automatic processing. However, processing is defined as automatic if sensory input and relevant skill procedures can be linked directly without "additional attention mechanisms" (Neumann, 1984, p. 281) and processing is defined as non-automatic if further attentional mechanisms are needed. Therefore, Neumann's conception is based on the assumption that several different attentional mechanisms exist, some of which are needed for automatic processing and some of which are additionally applied during non-automatic processing.

Modifying Neumann's (1984, 1990) conception of automatization, we propose that automatic processing does not involve "additional attention mechanisms". Instead, we suggest that automatization is associated with a change in attention control from a more sensory-based to a more LTM-based mode. In agreement with Logan (1988, 1990) and Neumann (1984, 1990), we think that attentional selection is needed for task execution irrespective of the degree of automatization. Attention has to be allocated to task-relevant target locations prior to action directed towards these target locations (Schneider, 1995). However, we think that in automatic processing, LTM is not only used to specify what is task relevant (e.g., the task instructions), but it also contains information about where relevant information can be found in the environment based on prior experience. Attentional control structures, the so-called "priority maps" (Bisley & Goldberg, 2010; Bundesen, Habekost, & Kyllingsbaek, 2011; Fecteau, & Munoz, 2006; Wischnewski, Belardinelli, Schneider, & Steil, 2010), should rely on this spatial LTM in automatic processing. A priority map is often conceptualized as a retinotopically organized map that contains representations of locations of objects. In addition, each location codes priority, that is, the importance of attending to this location (Wolfe, 1994). Priorities are computed based on bottom-up salience and top-down relevance. The location with the highest attentional weight (priority) determines the next saccade target location. For non-automatic processing, priorities are computed based on the external sensory input and top-down factors such as the current task. For automatic processing, we assume that priority computation should be based on a strong spatial signal from LTM that codes where the next relevant information of the environment might be found. The location of the priority map that receives this LTM signal should be selected as the next saccade target. However, priority computations should also be influenced by external sensory input, but to a much lesser degree than in non-automatic processing. In a sequential sensorimotor task, LTM information contains the sequence of task-relevant locations. Attention should be sequentially directed to the highly prioritized LTM locations generating the task-characteristic scan paths. In the dark condition of the present study, internal signals from LTM seemed to have won the competition against the low weighted auditory and haptic external signals that were available. Covert and overt attention was thus directed to hand-target locations before the hands have reached these locations.

Further evidence for an LTM influence on attention allocation

The phenomenon of contextual cueing (Chun, 2000; Chun & Jiang, 1998; Olson & Chun, 2001) constitutes further evidence that LTM contributes to the allocation of attention. Participants in the studies of Chun and Jiang (1998) as well as in those of Chun (2000) had to perform a visual search task with the following design. Without participants' knowledge, a fixed set of spatially invariant target-distractor configurations was presented. After sufficient repetitions, the invariant target-distractor configurations implicitly cued the target locations within the configuration, leading to faster target detection compared to a control condition with variable spatial relations between targets and distractors. Thus, the invariant target-distractor configurations must have been learned and stored to LTM, enabling faster attention allocation to target locations during repeated trials. However, processing of visual input was still necessary to specify the target locations in this spatial contextual cuing paradigm (Chun, 2000; Chun & Jiang, 1998). As the configurations followed each other in a random order, the last configuration was no indicator for the next configuration and the target location. In a later study, Olson and Chun (2001) showed that not only an invariant spatial configuration, but also an invariant temporal sequence can have an influence on attention, demonstrating that temporal contextual cueing is also possible. After training in Olson and Chun's (2001) experiment 1A, participants could identify a target faster when it followed an invariant sequence of stimulus durations than a sequence of random stimulus durations. Thus, the order of the target stimulus within a temporal sequence was used to predict when the target would appear. After training in experiment 3A of Olson and Chun (2001), participants could report a target faster when it followed an invariant spatio-temporal sequence of distractors that not only cued the time when the target would appear, but also where it would be located among distractors. This spatio-temporal contextual cueing can be seen as evidence for attention being directed to a specific location at a specific point in time within a spatio-temporal sequence based on information acquired during learning. Further evidence for an LTM influence on eye movements was reported by Noton and Stark (1971a, 1971b). Their participants viewed line drawings several times under conditions that prevented peripheral vision, so that only fixated parts of the line drawings could be seen clearly. Scan paths were highly similar between initial and repeated presentations. The authors concluded that these scan paths are connected to the internal pattern representation of the line drawings in memory and that stimulus-related scan paths are learned during the initial presentation. The results from the contextual cueing and the scan path paradigm as well as the key findings of the present study suggest that it is possible to store target locations as whole scan paths for covert and overt attention allocation in an automatized sensorimotor task with an invariant spatio-temporal sequence of sub-actions to objects.

Possible coupling of eye and hand movements

How are eye and hand movements related during the execution of an automatized sequential sensorimotor task? Eye-hand dynamics in the light and in the dark deliver an answer to this question. The absolute eye-hand time span was very similar between light and dark condition. The relative eye-hand time span and the eye-hand unit span were significantly smaller in the dark than in the light. The eye-hand unit span revealed that fixations in the light were mainly associated with the subsequent sub-actions (next ORA), while fixations in the dark were more often associated with the current sub-action (same ORA). While the hand followed the eye with the same absolute time delay across illumination conditions, the successive eye movement must have followed the last eye movement with a longer delay in the dark. This indicates that the interval between successive eye movements was flexibly adapted to the actual task requirements of acting in the dark. In contrast, the time between an eye movement and its corresponding hand movement has been found to be rather constant across sensorimotor tasks (e.g., Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land & Hayhoe, 2001; Land, et al., 1999; Land & Tatler, 2001, 2009) pointing to a tight coupling between eye and hand movements. Eye and hand movements might be coupled by a common selection of target locations for both movements (see Schneider, 1995, for the idea of a common spatial attentional signal for eye and hand movements during sensory processing). Alternatively, eye and hand movements could be selected separately first and coupled during motor initiation. As many actions such as pointing, grasping, and placing require both eye and hand to be directed to the same location in space, a common motor target selection might be efficient. After automatization, the same spatial LTM signal could be used for eye and hand movements. A common selection mechanism would not only be useful because of the common target locations in space, but also because a specific amount of time might be needed to up-date the hand movement parameters based on the preceding eye movement. There are psychophysical studies (Neggers & Bekkering, 2000; Song & McPeek, 2009) as well as functional imaging studies (Beurze, de Lange, Toni, & Medendorp, 2009; Levy, Schluppeck, Heeger, & Glimcher, 2007) that can be interpreted as support for a common selection mechanism of eye and hand-target locations. However, findings from psychophysical studies, functional imaging studies, as well as single-cell recordings (Calton, Dickinson, & Snyder, 2002; Prablanc, et al., 1979; Sailer, Eggert, Ditterich, & Sraube, 2000; Thompson & Westwood, 2007; Tosoni, Galati, Romani, & Corbetta, 2008; Van Der Werf, Jensen, Fries, & Medendorp, 2010) seem to support the idea of separate and largely independent selection of eye and hand-target locations. Recently, Jonikaitis and Deubel (2011) have demonstrated that participants could discriminate stimuli at a saccade and a reach-target location in parallel when they had to reach to one location and saccade to another location simultaneously. In addition, discrimination performance on both reach-target location and saccade-target location was equally good, when participants had to perform reach and saccade either separately or simultaneously. Initiation latency costs emerged in the separate location conditions. In conclusion, the majority of findings support the assumption of separate attention allocation mechanisms for eye and hand movements.

“Eye-guides-hand” benefit in the dark?

The question remains if and how eye movements to hand-target locations might be functional in the dark. Although no visual information can be extracted in the dark, saccading to a hand-target location may have specific functions as stated by Flanagan et al. (2008). Sensorimotor transformations from fixated locations to hand movements are well-practiced and computing target positions for hand movements may be facilitated by efferent and afferent signals from the eyeball. Thus, saccading to a hand-target location may facilitate the computation of the motor command even if no visual input is available. However, an “eye-guides-hand” benefit could neither be observed during the target-contacting task in Flanagan et al. (2008), nor during the automatized speed-stacking task in the present study. In Flanagan et al. (2008), the correlation between reach errors and the distance of saccade landing positions and target locations in the dark was not significant. In the present study, hand movement durations were neither correlated with temporal nor with spatial eye movement characteristics.

Why did participants saccade to hand-target locations in the dark anyhow? Systematic eye movements while performing a sensorimotor task in the dark might have other benefits than enhancing hand movement accuracy or speed. Four conceivable reasons why it might be beneficial to execute systematic eye movements in the dark will be outlined in the following.

First, eye movements might be used for timing the task steps. Evidence that eye movements can be used as an internal timer was reported in Huber and Krist (2004). In their study, participants who frequently tracked a falling target with their eyes - even when the target was not visible - performed better in judging the impact time of the occluded target than participants who less frequently performed such tracking eye movements. In the present study, the longer fixation durations and smaller eye-hand unit spans in the dark argue against the idea that eye movements had been used for timing the task steps. Instead, timing of successive spatial LTM signals seemed to be adapted according to the duration of task steps. The duration of task steps in turn increased with the advanced task difficulty in the dark.

Second, participants may use eye movements to facilitate the imagination of non-visible objects. This mechanism of eye movements was suggested by the imagery literature due to the highly similar eye movements during observation and imagination of the same visual stimuli (Brandt & Stark, 1997; Deckert, 1964; Heremans, Helsen, & Feys, 2008; Spivey & Geng, 2001). Visual imagination might be easier at locations that are currently fixated than at peripheral locations. Accurate visual imagination might in turn facilitate the execution of hand movements.

Third, eye movements might be used to retrieve the task steps. An indication that visuospatial rehearsal and eye movements are related was reported by Tremblay, Saint-Aubin, and Jalbert (2006). Participants in Tremblay et al. (2006) were better in remembering order and location of seven dots that appeared on a computer screen if they had fixated the successive locations in the correct order more frequently during the maintenance interval. In agreement with this idea, humans make use of their body to structure tasks cognitively. Gestures, for instance, support a speaker's thought process (e.g., Iverson & Goldin-Meadow, 1998; Krauss, 1998). Blind children's gestures are comparable to sighted children's gestures, and a participant gestures even if the dialogue partner is blind (Iverson & Goldin-Meadow, 1998). Similarly, eye movements might automatically retrieve the steps of a sensorimotor task independently from the illumination conditions. The plausibility of this idea depends on the still debated question whether eye movements are necessary for visuospatial rehearsal or whether covert shifts of attention suffice (Baddeley, 1986; Hale, Myerson, Rhee, Weiss, & Abrams, 1996; Lawrence, Myerson, Oonk, & Abrams, 2001; Smyth & Scholey, 1994).

Forth, participants may have executed their previously automatized scan paths in the dark because the prevention of overt eye movements may have attentional and other performance costs in the speed-stacking task. In order to test this idea, further investigations of speed stacking in the dark while the eyes must remain fixating are necessary.

In summary, the present study supplies several insights into eye movement control during the execution of automatized sensorimotor tasks. When performing an automatized sequential high-speed sensorimotor task in the dark, participants performed systematic eye movements that resemble saccades in the light in several respects. A similar number of fixations was executed in the dark as in the light, not only per trial, but also per sub-action. In addition, scan paths were very similar between light and dark conditions and the eyes reached an action-relevant location ahead of the hands even in complete darkness. However, performance was slower in the dark, resulting in longer fixation durations and smaller fixation rates in the dark than in the light. Finally, no relationship between temporal and spatial eye movement characteristics and task performance could be found in the dark. Results are interpreted as evidence for a more LTM-based mode of spatial attention control during the execution of automatized sequential sensorimotor tasks. In conclusion, automatized scan paths are maintained and used in the dark, although no visual input can be extracted by the eyes.

Acknowledgments

This research was supported by grants from the Cluster of Excellence Cognitive Interaction Technology (CITEC). We would like to thank Thomas Hermann and Bettina Blaesing for their support. Thanks are also extended to Michal Ljubljanac and Okka Risius who annotated the data for the interrater reliability as well as Frank Lehmke who wrote the JAVA annotation tool.

Commercial relationships: none.

Corresponding authors: Rebecca M. Foerster and Werner X. Schneider.
Email: rebecca.foerster@uni-bielefeld.de and wxs@uni-bielefeld.de.
Address: Department of Psychology, Bielefeld University, P.O. Box 100131, D-33501 Bielefeld, Germany.

References

Abrams, R. A., Meyer, D. E., & Kornblum, S. (1990). Eye-hand coordination: Oculomotor control in rapid aimed limb movements. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 248-267.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Baddeley, A. D. (1986). Working memory. Oxford, England: Oxford University Press.

Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *337*, 331-339.

Beurze, S. M., de Lange, F. P., Toni, I., & Medendorp, W. P. (2009). Spatial and effector processing in the human parietofrontal network for reaches and saccades. *Journal of Neurophysiology*, *101*, 3053-3062.

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*, 1-21.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, *9*, 27-38.

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2011). A neural theory of visual attention and short-term memory (NTVA). *Neuropsychologia*, *49*, 1446-1457.

Calton, J. L., Dickinson, A. R., & Snyder, L. H. (2002). Non-spatial, motor-specific activation in posterior parietal cortex. *Nature Neuroscience*, *5*, 580-588.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4, 170-178.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28-71.

Cowan, N. (2011). The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia*, *49*, 1401-1406.

Crawford, J. D., Medendorp, W. P., & Marotta, J. J. (2004). Spatial transformations for eye-hand coordination. *Journal of Neurophysiology*, *92*, 10-19.

Deckert, G. H. (1964). Pursuit eye movements in the absence of moving visual stimulus. *Science*, *143*, 1192-1193.

Desmurget, M., Pelisson, D., Rossetti, Y., & Prablanc, C. (1998). From eye to hand: Planning goal-directed movements. *Neuroscience and Biobehavioral Reviews*, *22*, 761-788.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*, 1827-1837.

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology, Human Perception and Performance*, *6*, 1352-1365.

Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1416-1438.

Epelboim, J., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research*, *35*, 3401-3422.

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, *10*, 382-390.

Findlay, J. M. (2009). Saccadic eye movement programming: Sensory and attentional factors. *Psychological Research*, *73*, 127-135.

Flanagan, J. R., Terao, Y., & Johansson, R. S. (2008). Gaze behavior when reaching to remembered targets. *Journal of Neurophysiology*, *100*, 1533-1543.

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision*, *11(7):9*, 1-16.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2):6, 1-17.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during music sight-reading. *Proceedings of the Royal Society of London B*, *266*, 2435-2440.

Hale, S., Myerson, J., Rhee, S. H., Weiss, C. S., & Abrams, R. A. (1996). Selective interference with the maintenance of location information in working memory. *Neuropsychology*, *10*, 228-240.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*, 49-63.

Heremans, E., Helsen, W. F., & Feys, P. (2008). The eyes as a mirror of our thoughts: Quantification of motor imagery of goal-directed movements through eye movement registration. *Behavioral Brain Research*, *187*, 351-360.

Huber, S., & Krist, H. (2004). When is the ball going to hit the ground? Duration estimates, eye movements, and mental imagery of object motion. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 431-444.

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, *396*, 228.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *The Journal of Neuroscience*, *21*, 6917-6932.

Jonikaitis, D., & Deubel, H. (2011). Independent allocation of attention to eye and hand targets in coordinated eye-hand movements. *Psychological Science*, *22*, 339-347.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, *7*, 54-60.

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*, 3559-3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311-1328.

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology*, *11*, 1215-1220.

Land, M. F., & Tatler, B. W. (2009*). Looking and acting*. Oxford: Oxford University Press.

Lawrence, B. M., Myerson, J., Oonk, H. M., & Abrams, R. A. (2001). The effects of eye and limb movements on working memory. *Memory*, *9*, 433-444.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics, Doklady*, *10*, 707-710.

Levy, I., Schluppeck, D., Heeger, D. J., & Glimcher, P. W. (2007). Specificity of human cortical areas for reaches and saccades. *The Journal of Neuroscience*, *27*, 4687-4696.

Logan, G. D. (1988). Towards an instance theory of automatization. *Psychological Review*, *95*, 492-527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*, 1-35.

Mennie, N., Hayhoe, M., & Sullivan, B. (2006). Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research*, *179*, 427-442.

Myers, C. W., & Gray, W. D. (2010). Visual scan adaptation during repeated visual search. *Journal of Vision*, *10*(8):4, 1-14.

Neggers, S. W. F., & Bekkering, H. (2000). Ocular gaze is anchored to the target of an ongoing pointing movement. *Journal of Neurophysiology*, *83*, 639-651.

Neumann, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 255-293). Berlin: Springer.

Neumann, O. (1990). Direct parameter specification and the concept of perception. *Psychological Research*, *52*, 207-215.

Noton, D., & Stark, L. W. (1971a). Scan paths in eye movements during pattern perception. *Science*, *171*, 308-311.

Noton, D., & Stark, L. (1971b). Scan paths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*, 929-942.

Olson, I. R., & Chun, M. M. (2001). Temporal context cueing of visual attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 1299-1313.

Prablanc, C., Desmurget, M., & Gréa, H. (2003). Neural control of on-line guidance of hand reaching movements. *Progress in Brain Research*, *142*, 155-170.

Prablanc, C., Echallier, J. F., Komilis, E., & Jeannerod, M. (1979). Optimal response of eye and hand motor systems in pointing at a visual target. I. Spatio-temporal characteristics of eye and hand movements and their relationships when varying the amount of visual information. *Biological Cybernetics*, *35*, 113-124.

Prablanc, C., Pélisson, D., & Goodale, M. A. (1986). Visual control of reaching movements without vision of the limb. I. Role of retinal feedback of target position in guiding the hand. *Experimental Brain Research*, *62*, 293-302.

Rosenbaum, D. A. (2010). *Human motor control*. London: Academic Press.

Sailer, U., Eggert, T., Ditterich, J., & Straube, A. (2000). Spatial and temporal aspects of eye-hand coordination across different tasks. *Experimental Brain Research*, *134*, 163-173.

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye-hand coordination during learning of a novel visuomotor task. *Journal of Neuroscience*, *25*, 8833-8842.

Schneider, W. X. (1995). VAM: A neuro-cognitive model for visual attention control of segmentation, object recognition and space-based motor action. *Visual Cognition*, *2*, 331-376.

Schneider, W., & Shiffrin, R. M. (1977a). Controlled and automatic human information processing: 1. Detection, search and attention. *Psychological Review*, *84*, 1-66.

Schneider, W., & Shiffrin, R. M. (1977b). Controlled and automatic human information processing: 2. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127-189.

Smyth, M. M., & Scholey, K. A. (1994). Interference in immediate spatial memory. *Memory and Cognition*, *22*, 1-13.

Song, J.-H., & McPeek, R. M. (2009). Eye-hand coordination during target selection in a pop-out visual search. *Journal of Neurophysiology*, *102*, 2681-2692.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychological Research*, *65*, 235-241.

‘t Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., & Einhäuser, W. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, *17*, 1132-1158.

Thompson, A. A., & Westwood, D. A. (2007). The hand knows something that the eye does not: Reaching movements resist the Müller-Lyer illusion whether or not the target is foveated. *Neuroscience Letters*, *426*, 111-116.

Tosoni, A., Galati, G., Romani, G. L., & Corbetta, M. (2008). Sensory-motor mechanisms in human partietal cortex underlie arbitrary visual decisions. *Nature Neuroscience*, *11*, 1446-1453.

Tremblay, S., Saint-Aubin, J., & Jalbert, A. (2006). Rehearsal in serial memory for visual-spatial information: Evidence from eye movements. *Psychonomic Bulletin and Review*, *13*, 452-457.

Van Der Werf, J., Jensen, J., Fries, P., & Medendorp, W. P. (2010). Neuronal synchronization in human posterior parietal cortex during reach planning. *The Journal of Neuroscience*, *30*, 1401-1412.

Wischnewski, M., Belardinelli, A., Schneider, W. X., & Steil, J. J. (2010). Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. *Cognitive Computation*, *2*, 326-343.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*, 202-238.

Betreff: JOV-[redacted]-2011 New Paper Received for JOV
Von: droddy@arvo.org
Datum: 19.09.2011 20:24
An: rebecca.foerster@uni-bielefeld.de

19th Sep 2011

Dear Miss Foerster,

Your manuscript titled "Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task" by Rebecca Foerster, Elena Carbone, Hendrik Koesling, and Werner Schneider has been successfully submitted online and will be forwarded to the Editor-in-Chief, Dr. Andrew Watson, for peer review assignment in the Journal of Vision.

Your manuscript has been assigned the Manuscript ID: JOV-[redacted]-2011.

Please mention the above manuscript ID in all future correspondence. If there are any changes in your institution or e-mail address, please log in to the Journal of Vision manuscript tracking site at http://jov.msubmit.net/ and update or edit your user information as appropriate. You may check on the status of this manuscript by selecting the "Check Manuscript Status" link under the following URL:

<http://jov.msubmit.net/cgi-bin
/main.plex?el=A6Gg3GC5A7DUb6F3A9MBCwVYWpiOSAV1CKXjzcsQZ>

Thank you for submitting your work to the Journal of Vision.

Sincerely,
Journal of Vision
Editorial Office

1801 Rockville Pike, Suite 400
Rockville, MD 20852-5622
Tel: +1.240.221.2930
Fax: +1.240.221.0355

Betreff: Journal of Vision - Decision on Manuscript ID JOV-[redacted]-2011
Von: mary@mail.cps.utexas.edu
Datum: 27.10.2011 15:27
An: rebecca.foerster@uni-bielefeld.de
Kopie (CC): mary@mail.cps.utexas.edu, elena.carbone@uni-bielefeld.de, hendrik.koesling@uni-bielefeld.de, wxs@uni-bielefeld.de

Manuscript ID JOV-[redacted]-2011 titled "Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task"

Dear Rebecca,

Your manuscript has been carefully reviewed and is potentially acceptable for publication, but there are a few minor issues that still need to be addressed before the paper can be accepted in its final form. The reviewer comments are appended below. Please prepare a revised manuscript and a point-by-point response to each issue identified by the reviewers. Only after satisfactory revision will your paper officially be accepted for publication. Please try and return your revised manuscript within 30 DAYS.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript using a word processing program and save it on your computer. Please also highlight the changes to your manuscript within the document by using the track changes mode in MS Word or by using bold or colored text.

To submit your revised manuscript, click on the "Revise #JOV-[redacted]-2011" link in the Author Tasks section at <http://jov.msubmit.net/cgi-bin/main.plex?el=A5Gg1GC1A5DUb5F3A9MBCwVYWpiOSAV1CKXjzcsQZ>

Please prepare a point-by-point response to the suggestions of the reviewers. This can be a Word or PDF file to be uploaded with the rest of the manuscript files. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewer(s).

IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any redundant files before completing the submission.

If it is not possible for you to submit your revision within the time requested please contact the journal as soon as possible.

Once again, thank you for submitting your manuscript to the Journal of Vision and I look forward to receiving your revision.

Sincerely,
Mary Hayhoe
Editor, Journal of Vision

Erklärung zur Urheberschaft - Kumulative Dissertation von Rebecca M. Förster

Hiermit bestätigen wir, dass das Manuskript „*Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task*" selbstständig von Frau Förster erstellt wurde. Frau Förster war hauptverantwortlich für die Planung, Durchführung und Auswertung dieser empirischen Arbeit. Außerdem verfasste und überarbeitete sie eigenständig alle Versionen des Manuskripts.

(Elena Carbone)

(Werner X. Schneider)

(Hendrik Kösling)

(Rebecca M. Förster)

2.3 Manuscript 3

“The interplay of attention, working memory, and long-term memory: An interference study with automatized tasks”

Foerster, R. M., Carbone, E., & Schneider, W. X (2011). The interplay of attention, working memory, and long-term memory: An interference study with automatized tasks. *Manuscript submitted for publication in Journal of Experimental Psychology: Human Perception and Performance*.

* Confirmation of submission and co-author statements are attached.

The interplay of attention, working memory, and long-term memory: An interference study with automatized tasks

Rebecca M. Foerster, Elena Carbone, & Werner X. Schneider

Department of Psychology and Cluster of Excellence „Cognitive Interaction Technology",
Bielefeld University, Germany

Correspondence to
Rebecca M. Foerster
Department of Psychology, Bielefeld University
Cognitive Interaction Technology - Center of Excellence (CITEC), Bielefeld University
P. O. Box 100131, D-33501 Bielefeld, Germany
Phone: 011-49-106-4347
Fax: 011-49-106-156934
E-mail rebecca.foerster@uni-bielefeld.de

Abstract

Previous studies demonstrated domain-specific interference between working memory (WM) maintenance and secondary tasks. For instance, visuospatial memory is impaired by spatial sensorimotor tasks. This interference seems to be due to competition for visuospatial attention. However, little is known of how attention and WM are engaged during long-term memory (LTM) retrieval. This is of particular interest for understanding cognitive processes involved in everyday activities such as driving or writing. These activities are usually highly automatized and LTM-driven. We asked whether such LTM-driven tasks interfere with WM retention and whether this interference is domain-specific. Participants performed either a highly practiced visuospatial sensorimotor (speed stacking) or verbal task (high-speed poem-reciting), while maintaining visuospatial or verbal information in WM. Results revealed unidirectional and domain-specific interference. Neither speed stacking nor high-speed poem reciting was influenced by WM retention. Stacking disrupted the retention of visuospatial locations, but did not modify memory performance of verbal material (letters). Reciting reduced the retention of verbal material substantially while it affected the memory performance of visuospatial locations to a smaller degree. We suggest that shifts of attention are needed to select information from LTM and to maintain it within domain-specific WM stores for the control of automatized actions.

Keywords attention, working memory, long-term memory, automaticity, interference

Introduction

During the last decades, our knowledge on attention, working memory (WM), and long-term memory (LTM) has dramatically increased (for reviews see, e.g., Baddeley, 2012; Bundesen & Habekost, 2008; Squire & Wixted, 2011). Experimental studies and theoretical concepts on the interplay of these functional domains are still sparse. However, for understanding the mental processes underlying everyday human behavior – such as driving, writing, or tea making – attention, WM, and LTM should not be considered in isolation (e.g., Land & Tatler, 2009). Such everyday tasks are characterized by a high degree of automatization which implies a substantial LTM contribution (e.g., Logan, 1988; Foerster, Carbone, Koesling, & Schneider, 2011a, 2011b). Not much is known about whether and how attentional and WM processes are also engaged in these tasks. The present study aims at closing this knowledge gap by focusing on the role of attention and WM during LTM-based execution of automatized tasks.

A useful taxonomy for understanding the relationship between attention, WM, and LTM has been provided by Chun, Golomb, and Turk-Browne (2010). The taxonomy distinguishes between external and internal attention. Attention is hence categorized according to its targets which can either be located in the external or internal world of an attentive human. Research has mainly focused on external attention, for instance during visual search (e.g., Wolfe, 2010). Internal attention such as the selection of LTM representations is less well understood.

If WM connects internal LTM representations with action control within the external world (Baddeley, 2012), then it will be important to specify the structure and processes of WM first. According to Baddeley (1986; Baddeley & Hitch, 1974) WM consists of multiple components for temporary storage and manipulation of limited information. One passive store, the articulatory loop, is concerned with verbal information. Another passive store, the visuospatial sketchpad, is concerned with visuospatial information. An active control system, the central executive, manipulates incoming and stored information. A fourth component – the episodic buffer – was added later (Baddeley, 2000). That is a passive multidimensional store receiving input from both the verbal and the visuospatial store. It is connected to long-term memory (LTM), and controlled by the central executive (Baddeley, Allen, & Hitch, 2011). Support for separated WM stores (verbal vs. visuospatial) was provided by dual task studies that demonstrated domain-specific interference (e.g., Baddeley, 2003; Baddeley & Hitch, 1994; Logie, Zucco, & Baddeley, 1990). Further evidence has been found in neuroimaging studies (e.g., D‟Esposito, 2007) and patient studies (e.g., Hamamé et al., 2011; Vallar & Papagno, 2002). Support for a strict separation of the central executive and the episodic buffer is lacking. Moreover, it is not clear how the WM components interact with LTM and attention. For that reason, Baddeley (2012) stated in a recent review that the integration of perception, LTM, and action into the WM model is an important upcoming step.

Perhaps best understood is the relation of attention and WM during visuospatial rehearsal. A fruitful discussion focused on whether covert shifts of attention, eye movements, implicit, or explicit motor activity or spatial imagery underlie the visuospatial rehearsal process (Awh, Jonides, & Reuter-Lorenz, 1998; Awh, Smith, & Jonides, 1995; Baddeley, 1986; Hale, Myerson, Rhee, Weiss, & Abrams, 1996; Smyth, & Scholey, 1994). Meanwhile, the assumption of covert attention as a visuospatial rehearsal process has been supported not only by behavioral (e.g., Awh, Jonides, & Reuter-Lorenz, 1998; Smyth, 1996; Smyth & Scholey, 1994; Theeuwes, Kramer, & Irwin, 2011), but also by neuroimaging evidence (e.g., Awh & Jonides, 1998; Awh et al., 1995; Awh et al., 1999; Awh, Anllo-Vento, & Hillyard, 2000). Hence, recent reviews suggest that attention is covertly shifted to locations that have to be maintained in the visuospatial sketchpad of WM (Awh & Jonides, 2001; Awh, Vogel, & Oh, 2006).

It has been shown that both attentional and WM processes are involved in performing sensorimotor tasks. Selective visuospatial attention usually determines which information of the environment will access WM (Awh, et al., 2006; Bundesen & Habekost, 2008; Bundesen, Habekost, & Kyllingsbaek, 2005, 2011). WM access of task-relevant information in turn is likely a necessary precondition for performing a sensorimotor task. In line with these assumptions, several studies revealed that sensorimotor tasks such as touching or pointing interfere with the maintainance of visuospatial, but not verbal material in WM (e.g., Hale et al., 1996; Lawrence, Myerson, Oonk, & Abrams, 2001; Smyth, & Scholey, 1994). The observed interference seems to be due to a competition for visuospatial attention by sensorimotor execution and WM retention.

However, the stimulus-driven sensorimotor tasks investigated so far (e.g., Hale et al., 1996; Lawrence et al., 2001; Smyth, & Scholey, 1994) are very dissimilar to the sensorimotor tasks humans perform in their everyday life. Natural tasks such as tea making (Land, Mennie, & Rusted, 1999), sandwich making (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003), and car driving (Land & Tatler, 2001) are characterized by a high contribution of LTM (e.g., Foerster et al., 2011a, 2011b). Up to now, little is known about the role attentional and WM processes play for LTM retrieval in such automatized tasks.

Due to their high dependence on LTM retrieval, tasks with automatized processing are a good starting point for investigating the interplay of attention, WM, and LTM. Traditional conceptions of automatization (Logan, 1988, 1990, 1992; Neumann, 1984, 1990; Schneider & Shiffrin, 1977a, 1977b) do not ascribe WM and attention a central role to automatic processing. According to the two-process theory of information processing (Schneider & Shiffrin, 1977a, 1977b), automatic processes, contrary to controlled processes, do not need attention and can be performed interference-free in parallel with other processes. The direct parameter specification concept (Neumann, 1984, 1990) and the instance theory of automatization (Logan, 1988, 1990, 1992) postulated a need of – at least external – attention. In line with this assumption, we showed recently that external attention (indicated by saccadic eye movements) was allocated to task-relevant information in the environment when performing an automatized sensorimotor task (Foerster et al., 2011a, 2011b). Moreover, where-to-attend-next was determined by LTM information. However, the contribution of internal attention and WM for the execution of LTM-based automatized tasks is still unclear.

The present study aims at better understanding the interplay of attention, WM, and LTM by investigating automatized tasks in a dual task scenario. More specifically, we asked whether domain-specific WM retention – possibly by internal attention - interferes with the execution of LTM-based automatized tasks. Participants had to perform either an automatized verbal task (high-speed poem reciting) or an automatized sensorimotor task (speed stacking), while maintaining either verbal (letters) or visuospatial (locations) material in WM. We chose high-speed poem reciting (reciting a poem by heart as fast as possible) and speed stacking (stacking up and down distinct configurations of cups as fast as possible, for an illustrative video visit http://www.speedstacks.com/about/history.php) because both tasks can be learnt easily and automatized rapidly. In addition, one trial of such a task can be accomplished within a relatively short retention interval of a WM task. Interference effects between such automatized LTM-driven tasks and the retention of domain-specific material would point to a contribution of domain-specific internal attention and WM processes.

Method

Participants.

Ten students from Bielefeld University, Germany, participated in the experiment. Nine of them took part in a speed-stacking automatization study (Foerster et al., 2011a) and the tenth participant ran through the same speed-stacking training before participating in the present experiment. Participants" age ranged from 21 to 32 years with a mean of 26. All participants had either normal or corrected-to-normal vision, were naive with respect to the aims of the study, and were paid for their participation.

Apparatus and stimuli.

A notebook with a 15.4 inch screen, with a resolution of 1,024 × 768 pixels and speed-stacking equipment (cups, timer, and mat) were used for the experiment. For the speed stacking condition, participants were seated in front of the screen and the speed-stacking equipment was placed in-between them and the screen. The distance to the screen was approximately 60 cm. Stimulus presentation was controlled by the Experiment Builder software (SR Research, Ontario, Canada). For the WM task, stimuli were displayed on a black background. The verbal memory stimuli were yellow consonants (B, F, J, L, N, Q, R, V, and X), appearing successively inside of a white frame (subtending approximately 2.86° of visual angle) centered on the screen. For the visuospatial WM-span task, grey filled white squares (again subtending approximately 2.86° of visual angle) were distributed in a fixed layout across the screen, and individual frames successively changed their inner color to yellow and back to grey. The visuospatial task was similar to the Corsi Block task of De Renzi and Nichelli (1975). Neither a letter nor a location was repeated within a sequence. The poem consisted of four quatrains with rhyming couplets and iamb as measure (see Appendix).

Design.

The within-subject variables were WM-span task (none, verbal, and visuospatial) and automatized task (none, reciting, and stacking). WM-span condition was blocked starting without WM-span task as a first block and verbal and visuospatial WM-span task as second and third block (the order was counterbalanced across participants). The automatized task conditions were intermixed within the two latter WM-span blocks. The first block of the experiment (no WM-span task) consisted of 6 stacking and 6 reciting trials. Each of the other two WM-span blocks (verbal and visuospatial) consisted of 18 experimental trials, with 6 trials each for the three automatized task conditions (none, reciting, and stacking), adding up to a total of 48 trials. Two practice trials (one verbal WM-span trial and one visuospatial WM-span trial, both without automatized task) at the beginning of the second block were added to ensure that the participants followed the instruction.

The dependent variables were percentage correct for the WM-span task as well as completion time and error rate for the speed-stacking task and the poem-reciting task. Trials were considered correct when all memory items were reported in the correct order. The performance measure of the automatized tasks was the duration of a complete stacking or reciting sequence. We defined a stacking error as one or more cups falling or sliding down and skipping. Substituting or adding of one or more words was defined as a reciting error.

Procedure.

Each experiment was preceded by a speed-stacking and a poem-reciting automatization period as well as a refreshment day directly before the experimental day. Speed stacking consists of a fixed sequence of stacking up and down pyramids of plastic cups as fast as possible. Number, order, and direction of the stacking movements are predetermined (for an illustrative video visit http://www.speedstacks.com/about/history.php). The speed-stacking automatization phase consisted of 14 days with 45 minutes practice each day (details are reported in Foerster et al., 2011a). The poem-reciting automatization lasted 50 minutes on a single day consisting of 10 minutes silent memorization and 40 minutes reciting at maximum speed. This poem-reciting automatization was preceded and followed by reading aloud the poem three times. On the refreshment day, both stacking and reciting had to be performed as fast as possible for 30 minutes. The third day was the experimental day and started with the first block of high-speed stacking and high-speed poem reciting without parallel WM-span task. This initial calculation of the participants‟ performance in stacking and reciting served as a baseline for the automatized tasks. The trial speed of both automatized tasks was measured by the speed-stacking timer and then transferred and stored on the notebook. The accuracy was marked by the experimenter. Afterwards, the dual task trials started with a written instruction appearing on the screen. Each trial started with a left mouse button press followed by the sequence of memory items, either 4 consonants or 3 locations. This difference in number of retained items was necessary to ensure equal task difficulty (see Results section). Each item was shown for 400 ms with an inter-stimulus interval (ISI) of 400 ms. Following the stimulus sequence, a written message was shown on the screen for 20 s informing the participants about the automatized task they had to accomplish within this delay (none, reciting, or stacking). A tone signaled the start and the end of the delay. For the verbal WM-span test, a central frame was shown on the screen and participants had to type in the letters in the correct order via the keyboard. Spatially distributed frames were shown on the screen for the visuospatial WM-span test, and participants had to select the locations via the mouse cursor in the correct order and confirm each selection with a left mouse click. The recording of the WM span stopped as soon as the participants made an error or had reproduced the complete sequence correctly. The reproduction was followed by a feedback („correct‟ or „incorrect‟). Trial sequences for all six combinations of conditions are shown in

Figure 1. The participants were supposed to memorize the items as accurately as possible and to stack and recite as fast as possible.

---Insert Figure 1 about here---

Results

Learning curves of automatized tasks.

According to Logan (1988, 1990, 1992), a task or process is automatized if an improvement in performance occurs during practice and if this improvement decreases with training. The data of our participants confirm this assumption.

Stacking time decreased significantly from the first (38.83 s) to the last (18.49 s) training day [$t(9) = 8.55$, $MSE = 2.38$, $p < .001$] and participants achieved a mean stacking time of 18.49 s with a mean best time of 12.63 s on the last training day. The high level of speed-stacking performance, the long-lasting practice of approximately 1084 trials per participant, and the small increase in performance at later stages of training (Figure 2) indicate that participants achieved a high degree of automatization in the high-speed stacking task.

---Insert Figure 2 about here---

Because the whole poem-reciting automatization took place on a single day, we split up the training trials of each participant into ten percent bins and calculated means of reciting times for each bin. All participants learned the poem as reflected by the significant overall decrease of mean reciting time between the first ten percent (42 s) and the last ten percent (20 s) of all trials [$t(9) = 4.79$, $MSE = 4.52$, $p < .01$]. Participants achieved a mean best reciting time of 14.47 s which equals to a reciting rate of 7 syllables per second. This best reciting time did not even differ significantly from the mean best reading time of 12.43 s (8 syllables per second) after training [$t(1,9) = 1.84$, $MSE = 1.11$, $p > .05$]. The high reciting rate, its comparability to reading rate, and the small increase in learning at later stages of training (Figure 3) indicate that participants achieved a high degree of automatization in the high-speed reciting task.

---Insert Figure 3 about here---

Performance of automatized tasks with concurrent WM-span tasks.

The speed-stacking time was 19.15 s without dual task, 18.48 s in the verbal WM-span condition, and 18.92 s in the visuospatial WM-span condition. The poem-reciting baseline time was 14.65 s without dual task, 13.93 s in the verbal WM-span condition, and 13.46 s in the visuospatial WM-span condition. The speed-stacking error rate baseline was 36.70 % without dual task, 33.30 % in the verbal WM-span condition, and 31.60 % in the visuospatial WM-span condition. The reciting error rate baseline was 6.70 % without dual task, 6.70 % in the verbal WM-span condition, and 8.40 % in the visuospatial WM-span condition. Stacking and reciting performance measures are depicted in Figure 4.

To test whether the WM-span tasks affected stacking or reciting speed or accuracy, we conducted two 2 x 3 analyses of variance for task completion time and error rate as dependent variables with automatized task (reciting and stacking) and WM-span task (none, verbal, and visuospatial) as within-subject variables. The analysis of task completion time revealed a significant main effect of automatized task [$F(1,10) = 14.07$, $MSE = 351.19$, $p < .01$], indicating that participants could recite the poem faster (14.01 s) than they could stack the cups (18.85 s). Neither the main effect of WM-span task [$F(2,20) = 2.36$, $MSE = 3.28$, $p >$

.05] nor the interaction of automatized task and WM-span task [*F*(2,20) = 2.50, *MSE* = 1.47, *p* > .05] were significant. The analysis of error rate revealed a significant main effect of automatized task [*F*(1,10) = 17.28, *MSE* = 1.06, *p* < .01], indicating that participants made less errors when reciting the poem (7.27 %) than when stacking the cups (33.87 %). Neither the main effect of WM-span task [*F*(2,20) = .11, *MSE* = .002, *p* > .05] nor the interaction of automatized task and WM-span task [*F*(2,20) = .14, *MSE* = .01, *p* > .05] were significant.

---Insert Figure 4 about here---

Performance of WM-span tasks with concurrent automatized tasks.

The verbal WM-span accuracy was 88.33 % without automatized task, 90.00 % during speed stacking, and 18.33 % during poem reciting. The visuospatial WM-span accuracy was 83.33 % without automatized task, 45.00 % during speed stacking, and 50.00 % during poem reciting. Performance measures for the WM-span tasks are depicted in Figure 5.

To test whether the automatized tasks affected the verbal or visuospatial memory, we conducted a 2 × 3 analysis of variance for the memory performance with WM-span task (verbal and visuospatial) and automatized task (none, stacking, and reciting) as within-subject variables. The analysis revealed no significant effect of WM-span task [*F*(1, 9) = .80, *MSE* = .06, *p* > .05], indicating that task difficulty was comparable. The main effect of automatized task was significant [*F*(2, 18) = 51.69, *MSE* = 1.37, *p* < .001] with the highest memory accuracy without automatized task (85.83 %), intermediate memory accuracy during stacking (67.50 %), and worst memory accuracy during reciting (34.17 %). The analysis also revealed a significant interaction between WM-span task and automatized task [*F*(2, 18) = 24.14, *MSE* = .73, *p* < .001]. Paired *t*-tests revealed that the verbal WM-span accuracy did not differ significantly between the no dual task (88.33 %) and the stacking (90.00 %) condition [*t*(1,9)

= .36, *MSE* = .05, *p* > .05], while it decreased significantly from 88.33 % without dual task to 18.33 % with parallel reciting [*t*(1,9) = 7.87, *MSE* = .28, *p* < .001]. The visuospatial WM-span accuracy was reduced significantly from 83.33 % without dual task to 45.00 % in the stacking condition [*t*(1,9) = 4.64, *MSE* = .08, *p* < .01] and also decreased significantly from 83.33 % without dual task to 50.00 % in the reciting condition [*t*(1,9) = 4.05, *MSE* = .08, *p* < .01]. However, this cross-domain interference between the visuospatial WM span and reciting was significantly smaller than the domain-specific interference between the verbal WM span and reciting [*t*(1,9) = 2.80, *MSE* = .13, *p* < .05].

---Insert Figure 5 about here---

Discussion

The present study aimed at better understanding the interplay of attention, WM, and LTM. More precisely, we investigated the role of internal attention and WM processes during the execution of LTM-based automatized tasks. Participants were asked to maintain either verbal (letters) or visuospatial (locations) material in WM while they had to perform an LTM-based automatized task that was either a verbal (high-speed poem reciting) or a visuospatial (speed stacking) task. If domain-specific internal attention and WM processes contribute to LTM retrieval during the execution of automatized tasks, domain-specific interference effects should arise between short-term retention and the execution of the automatized LTM-driven tasks. In this case, speed stacking should mainly interfere with the visuospatial WM span, while high-speed poem reciting should mainly interfere with the verbal WM span.

The results of the present study revealed indeed mainly domain-specific interference effects. Speed stacking disturbed the visuospatial, but not the verbal memory performance, while poem reciting disturbed the verbal WM performance significantly stronger than the visuospatial WM performance. Moreover, high-speed poem reciting was in general faster and more accurate than speed stacking. Neither reciting nor stacking was affected by the WM-span tasks. The fact that even the highly automatized LTM-driven tasks distorted WM performance in a mainly domain-specific manner indicates that LTM-based automatized tasks need internal attention and domain-specific WM processes.

In the subsequent sections, we will discuss the implications of our results for a more detailed understanding of the interplay between attention, WM, and LTM. First, we will attempt to explain the main data pattern of domain-specific interference between WM retention and the execution of LTM-based automatized tasks. Second, we will evaluate whether this main result pattern is compatible with already existing assumptions about the interplay of attention, WM, and LTM, and we will present our own view. Third, we will propose explanations for two supplementary experimental results, namely that poem reciting also distorted the visuospatial WM span and that the automatized tasks were not influenced by the WM tasks. We will close the discussion with a summary of what we can learn from the data about the interplay of attention, WM, and LTM.

Domain-specific interference of WM and automatized tasks.

There are two alternative ways to explain the observed domain-specific interference of the WM tasks by the automatized tasks. First, domain-specific LTM retrieval might have been involved in WM-span retention (Unsworth & Engle, 2007), so that the LTM retrieval of task-relevant information for the automatized tasks selectively disturbed the WM spans. Unsworth and Engle (2006) and more recently also Rose, Myerson, Roediger, and Hale (2010) suggested that LTM is not involved in simple WM-span tasks with less than five memory items. Since the WM-span tasks of the present experiment consisted of three visuospatial and four verbal memory items, LTM involvement during the WM-span tasks does not seem to be a plausible explanation.

Alternatively, the WM-span tasks and the automatized tasks might compete for attentional and domain-specific WM processes. More specifically, shifts of visuospatial attention (either internal or external or both) during speed stacking may have disturbed the attention-based rehearsal process within visuospatial WM. This assumption is supported by results from the following studies. First, the maintainance of visuospatial material is very likely based on visuospatial attention (e.g., Awh et al., 1998, 2000, 2006; Smyth, 1996; Theeuwes et al., 2011). Second, external attention has to be directed to important locations in the environment while performing an automatized sensorimotor task such as speed stacking (Foerster et al., 2011a, 2011b). Third, internal attention seems to be necessary for LTM retrieval (e.g., Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Wagner, Shannon, Kahn, & Buckner, 2005).

However, apart from the interference between the two visuospatial tasks, the results also revealed interference between the two verbal tasks. External attention is not needed during poem reciting because verbal information is exclusively provided by LTM. Even if external shift of visuospatial attention were necessary for poem reciting, this could not explain interference with verbal rehearsal. Instead of external attention, internal attention to verbal LTM content may be required for poem reciting. Could these internal attention shifts during reciting disturb verbal rehearsal? The standard assumption is that verbal rehearsal is realized by a subvocal articulatory process (e.g., Awh, Jonides, Smith, Schumacher, Koeppe, & Katz, 1996; Baddeley, Lewis, & Valler, 1984; Salame & Baddeley, 1982). However, it is not specified what exactly constitutes this subvocal articulatory process and whether it is based on attention. More recently, behavioral and neuroimaging studies (e.g., Majerus et al., 2011; Zhijian & Cowan, 2009) showed that attention is involved in verbal short-term retention. The difference between verbal and visuospatial rehearsal is probably just the target of attention (verbal vs. visuospatial). Therefore, internal attention for LTM retrieval during the execution of the automatized verbal task may have competed with attention-based rehearsal for the WM-span tasks. Moreover, it is likely that internal attention is not only necessary to retrieve LTM content, but also to maintain it in WM until response execution. Therefore, LTM-retrieved information may also have competed with the WM-span items for attention-based rehearsal within domain-specific verbal as well as visual-spatial WM stores.

In line with these ideas, Wagner et al. (2005) reviewed neuroimaging studies showing that the posterior parietal cortex (PPC) – an important structure for WM (e.g., Fiehler et al., 2011; Funashi, Bruce, & Goldman-Rakic, 1989) – is activated during episodic memory retrieval. The authors proposed that the PPC is activated because memory representations have to be attended for retrieval. Cabeza et al. (2008) further differentiated the role of the PPC and attentional processes during LTM retrieval. As the ventral region of the PPC (ventral parietal cortex or VPC) is frequently active during recollection of memory items, the authors assumed that the VPC is primarily involved in allocating attention to relevant cues in the environment. The dorsal region of the PPC (dorsal parietal cortex or DPC) in contrast is frequently active during familiarity judgments of memory items, so that the DPC might be primarily involved in allocating attention to the memory content. In the taxonomy of Chun et al. (2010), the VPC seems to be associated with external attention allocation while the DPC seems to be associated with internal attention allocation. Further studies are needed to reveal whether external or internal attention is involved in WM rehearsal.

Implications of our result for the interplay of attention, WM, and LTM.

Next, we will investigate whether our result of domain-specific interference of WM tasks and LTM-based automatized tasks is compatible with already existing theories about the interplay of attention, WM, and LTM. The observed interference clearly contradicts the two-process theory of information processing (Schneider & Shiffrin, 1977a, 1977b). Its key assumption is that automatic processes do not need attention and can neither disturb nor be disturbed by concurrent tasks. In contrast, the direct parameter specification account (Neumann, 1984, 1990) and the instance theory (Logan, 1988, 1990, 1992) suggest that automatic processing still needs attention to sensory input and may interfere with concurrent tasks. According to Neumann (1984, 1990), an action is automatized if relevant parameters are directly selected from LTM information (acquired skill) and sensory input. Similarly, Logan (1988, 1990, 1992) defines an action as automatized if it can be directly selected from LTM instances. Both authors do not make explicit assumptions about the role of WM during automatic processing. However, the postulated direct access to LTM for motor planning implies that LTM selected information does not have to be maintained in WM until task execution. This conception of automatic processing without the need for internal attention or WM processes may be compatible with the observed visuospatial interference if it was based on competition for external attention. However, it is not compatible with the observed interference between poem reciting and verbal memory retention.

The recent conceptualization of WM by Baddeley (2012) addresses the question of how attention and LTM might interact with WM for action control. Within the multi-component WM framework, the episodic buffer (Baddeley, 2000) is the bridge between the domain-specific stores, the central executive, LTM, and action. One of the functions that are ascribed to the central executive is attention (Baddeley, 1986, 2000, 2012; Baddeley & Hitch, 1974). The central executive controls the episodic buffer and should therefore also modulate the exchange of internal representations between WM and LTM.

The results of the present study indicate that internal attention and WM processes might be required during automatized processing. We think that attention, WM, and LTM interact in the following way during the execution of automatized tasks. Internal attention selects task-relevant information from LTM structures. The active maintenance of this selected information – by the same internal attention process - corresponds to retention of this information within the domain-specific WM until task execution. This assumption implies that WM encoding and retention is equivalent to selectively activated and maintained LTM contents (e.g., Cowan, 2005). Alternatively, one may assume completely separate WM and LTM stores implying that selected LTM contents have to be transferred to a further WM store, i.e., long-term synaptic weights are transferred into short-term continuous firing in neural circuits (Olivers, Peters, Houtkamp, & Roelfsema, 2011). Importantly, we assume that LTM representations can only be used for action control via WM. Consequentially, a tight interaction should exist between internal attention, domain-specific WM, and LTM processes during the execution of automatized tasks.

Supplementary findings: Cross-domain interference and asymmetry of interference effects.

Two supplementary findings of our study should be discussed. We start with the question, why poem reciting did not only reduce the verbal WM span, but also the visuospatial WM span, although to a smaller degree. This cross-domain interference can either be due to global WM load (within the central executive and the episodic buffer) or to interference within the visuospatial sketchpad or the articulatory loop. Global WM load refers to the involvement of the central executive and the episodic buffer, so that tasks compete for processes within these multidimensional WM domains. Global WM load might be higher during poem reciting than during speed stacking. What justifies this assumption? When performing a sensorimotor task in the real world, humans usually use a just-in-time strategy to gather visual(-spatial) information (Hayhoe et al, 2003). That means that they fixate on relevant visual information just when it is needed to perform a sub-action. This gaze strategy has also been observed in speed stacking (Foerster et al., 2011a). Even if a specific object contains two features relevant for the task at hand, humans process only that feature which is needed for the upcoming sub-action and re-fixate the object later to process the other feature (Droll, Hayhoe, Triesch, & Sullivan, 2005). This strategy of using the “world as external memory” (O‟Regan, 1992) reduces WM load. During high-speed poem reciting, outsourcing of relevant information to the environment is not possible. All relevant information needed for reciting has to come from LTM and has to be maintained in WM. This may cause a higher WM load during reciting than during stacking.

However, it is also possible that the observed interference between reciting and visuospatial WM was due to specific interference within the visuospatial sketchpad. Poem reciting itself might imply visuospatial processing. A visual imagery process of words during reciting could have been introduced because of the visual presentation of the poem during initial learning. If participants imagined words while reciting, they may have also shifted their visuospatial attention with respect to the visual image.

Finally, the interference between poem reciting and visuospatial memory may have been due to processes within the articulatory loop. Participants might have attempted to verbalize the locations they had to memorize. However, it is unlikely that verbalization of the Corsi Block locations was possible. Otherwise, speed stacking would not have disturbed the visuospatial WM span.

An additional supplementary question is why the interference effects between WM spans and automatized tasks were unidirectional. While the WM retention suffered from the concurrent execution of the automatized LTM-driven tasks, these tasks were unaffected by the simultaneous maintenance of information in WM. Participants seem to have prioritized the automatized high-speed tasks over the WM tasks, so that they could maintain at least the performance level of the automatized tasks to the disadvantage of the WM-span tasks.

Summary.

The current study has demonstrated that visuospatial, but not verbal WM was disturbed by an LTM-based, automatized, sensorimotor task. Complementary, verbal but not visuospatial WM was affected by a verbal, LTM-based, automatized task. Moreover, the two automatized tasks were not disturbed by concurrent retention of domain-specific information in WM. This finding of unidirectional and mainly domain-specific interference points to a requirement of internal attention and domain-specific WM processes during the execution of LTM-based automatized tasks. On the one hand, task-relevant LTM information might be retrieved by internal attention allocation. On the other hand, internal attention processes might maintain selected LTM information within domain-specific WM stores until task execution.

References

Awh, E., Anllo-Vento, L., & Hillyard, S. A. (2000). The role of spatial selective attention in working memory for locations: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *12*, 840-847.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*, 119-126.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from PET. *Psychological Science*, *7*, 25-31.

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 780-790.

Awh, E., Smith, E., E., & Jonides, J. (1995). Human rehearsal processes and the frontal lobes: PET evidence. In J. Grafman, K. Holyoak, & F. Boller (Eds.), *Annals of the New York Academy of Sciences: Vol. 769. Structure and functions of the human prefrontal cortex* (pp. 97-119). New York: New York Academy of Sciences.

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*, 201-208.

Awh, E. et al. (1999). Rehearsal in spatial working memory: Evidence from neuroimaging. *Psychological Science*, *10*, 433-437.

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, *4*, 417-423.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829-839.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 12.1-12.29.

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, *49*, 1393-1400.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and therapy* (pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*, 485-493.

Baddeley, A. D., Lewis, V., & Valler, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology*, *36*, 233-252.

Bundesen, C., & Habekost, T. (2008). *Principles of visual attention*. Oxford: Oxford University Press.

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, *112*, 291-328.

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2011). A neural theory of visual attention and short-term memory (NTVA). *Neuropsychologia*, *49*, 1447-1457.

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*, *9*, 613-625.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2010). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73-101.

Cowan, N. (2005). *Working memory capacity*. Hove, UK: Psychological Press.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 761-772.

De Renzi, E., & Nichelli, P. (1975). Verbal and non-verbal short memory impairment following hemispheric damage. *Cortex*, *11*, 341-354.

Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1416-1438.

Fiehler, K., et al. (2011). Working memory maintainance of grasp-target information in the human posterior parietal cortex. *Neuroimage*, *54*, 2401-2411.

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011a). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision*, *11*(7), 9, 1-16.

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011b). Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task. Manuscript submitted for publication.

Funashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey‟s dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*, 331-349.

Hale, S., Myerson, J., Rhee, S. H., Weiss, C. S., & Abrams, R. A. (1996). Selective interference with the maintenance of location information in working memory. *Neuropsychology*, *10*, 228-240.

Hamamé, C. M. et al. (2011). Reading the mind‟s eye: Online detection of visuo-spatial working memory and visual imagery in the inferior temporal lobe. *Neuroimage*. *59*, 872-879.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*, 49-63.

Lawrence, B. M., Myerson, J., Oonk, H. M., & Abrams, R. A. (2001). The effects of eye and limb movements on working memory. *Memory*, *9*, 433-444.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311-1328.

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology*, *11*, 1215-1220.

Land, M. F., & Tatler, B. W. (2009*). Looking and acting*. Oxford: Oxford University Press.

Logan, G. D. (1988). Towards an instance theory of automatization. *Psychological Review*, *95*, 492-527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*, 1-35.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 883-914.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Visuo-spatial processing in working memory. *Quarterly Journal of Experimental Psychology*, *38A*, 229-247.

Majerus, S., et al. (2011). Attention supports verbal short-term memory via competition between dorsal and ventral attention networks. *Cerebral Cortex*. Advance Access published July 17, 2011.

Neumann, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 255-293). Berlin, Springer.

Neumann, O. (1990). Direct parameter specification and the concept of perception. *Psychological Research*, *52*, 207-215.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*, 327-334.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*, 461-288.

Rose, N. S., Myerson, J., Roediger, H. L., Hale, S. (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span. *Journal of Experimental Psychology - Learning, Memory and Cognition*, *36*, 471-483.

Salame, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 150-164.

Schneider, W., & Shiffrin, R. M. (1977a). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review*, *84*, 1-66.

Schneider, W., & Shiffrin, R. M. (1977b). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127-189.

Smyth, M. M. (1996). Interference with rehearsal in spatial working memory in the absence of eye movements. *Quarterly Journal of Experimental Psychology*, *49A*, 940-949.

Smyth, M. M., & Scholey, K. A. (1994). Interference in immediate spatial memory. *Memory and Cognition*, *22*, 1-13.

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience*, *34*, 259-288.

Theeuwes, J., Kramer, A. F., & Irwin, D. E. (2011). Attention on our mind: The role of spatial attention in visual working memory. *Acta Psychologica*, *137*, 248-251.

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, *54*, 68-80.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104-132.

Vallar G., & Papagno, C. (2002). Neuropsychological impairments of short-term memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), *Handbook of memory disorders* (2nd ed., pp. 249-270). Chichester, UK: John Wiley & Sons.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*, 445-453.

Wolfe, J. M. (2010). Visual search. *Current Biology*, *20*, R346-R349.

Zhijian, C., & Cowan, N. (2009). How verbal memory loads consume attention. *Memory and Cognition*, *37*, 829-836.

Author Note

This research was supported by a grant of the Cluster of Excellence Cognitive Interaction Technology (CITEC) at Bielefeld University. We would like to thank Hendrik Koesling, Thomas Hermann, and Bettina Blaesing for their productive contributions to the speed stacking project.

Correspondence should be addressed to Rebecca Förster, Department of Psychology, Bielefeld University, P.O. Box 100131, D-33501 Bielefeld, Germany, E-mail rebecca.foerster@uni-bielefeld.de

Figures

Figure 1

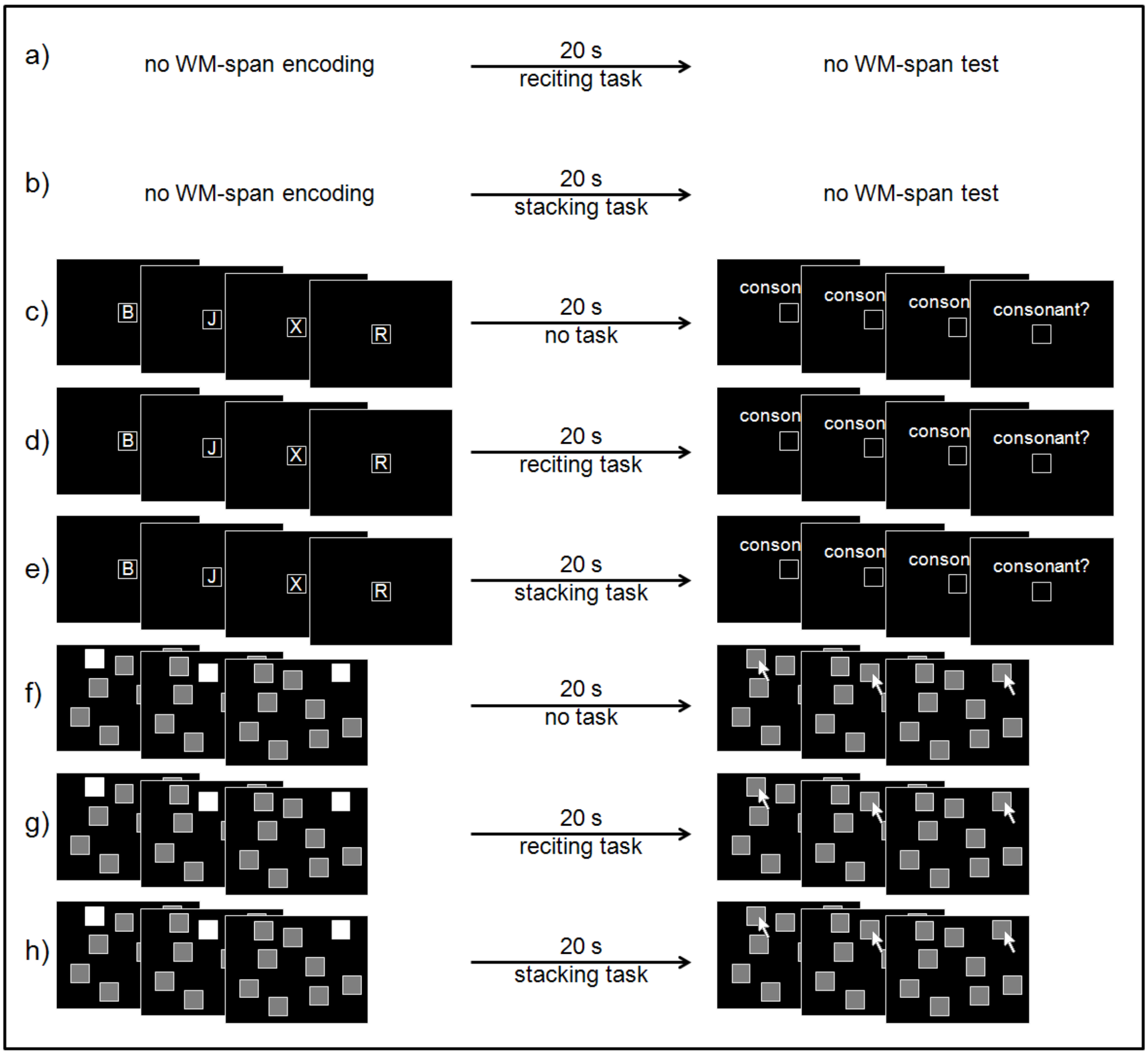

Figure 1. Trial sequences for the eight different experimental combinations of conditions: a) No WM-span task with reciting as automatized task. b) No WM-span task with stacking as automatized task. c) Verbal WM-span task with reciting as automatized task. d) Verbal WM-span task with stacking as automatized task. e) Verbal WM-span task without automatized task. f) Visuospatial WM-span task with reciting as automatized task. g) Visuospatial WM-span task with stacking as automatized task. H: Visuospatial WM-span task without automatized task.

Figure 2

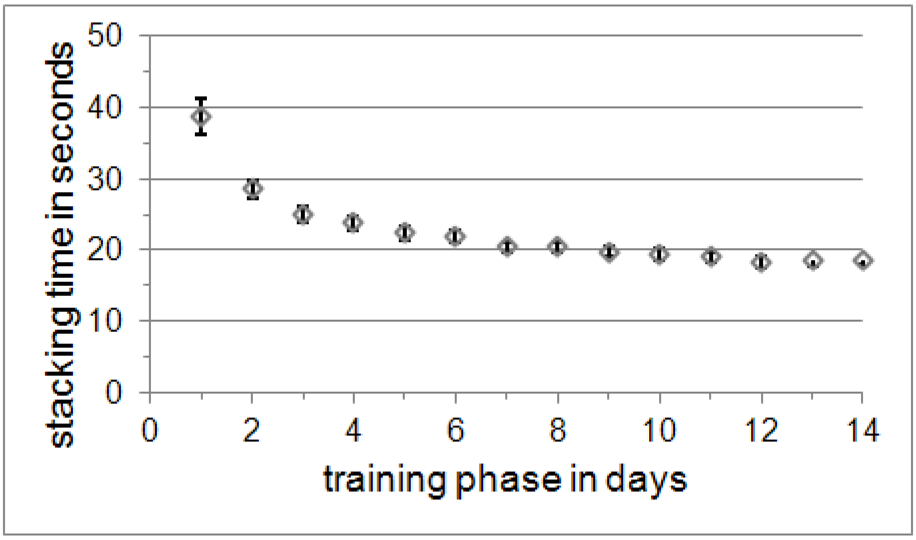

Figure 2. The mean stacking time in seconds with standard error of the mean per training phase in days.

Figure 3

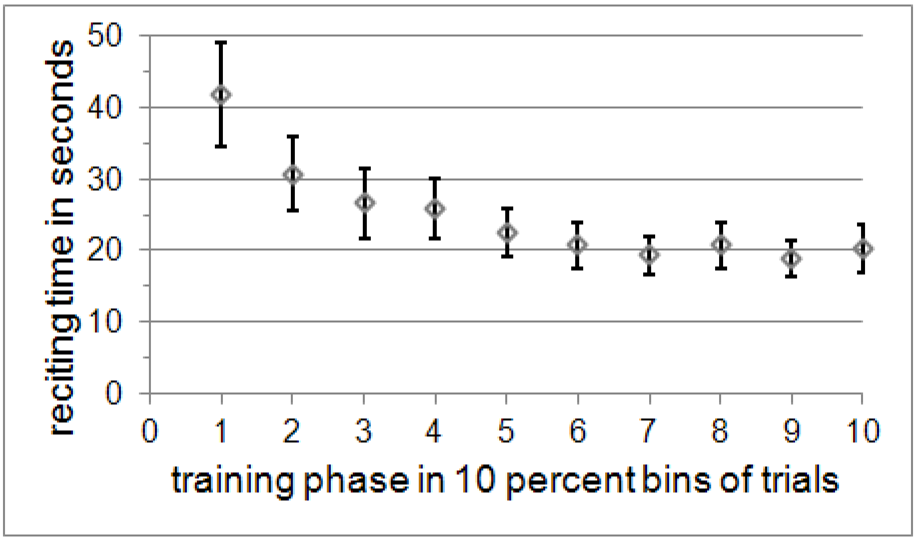

Figure 3. The training trials of each participant‟s reciting automatization day were split into ten percent bins, and means of reciting times were calculated. The figure shows the mean reciting time in seconds with standard error of the mean per training phase in 10 percent trial bins, in temporal order.

Figure 4

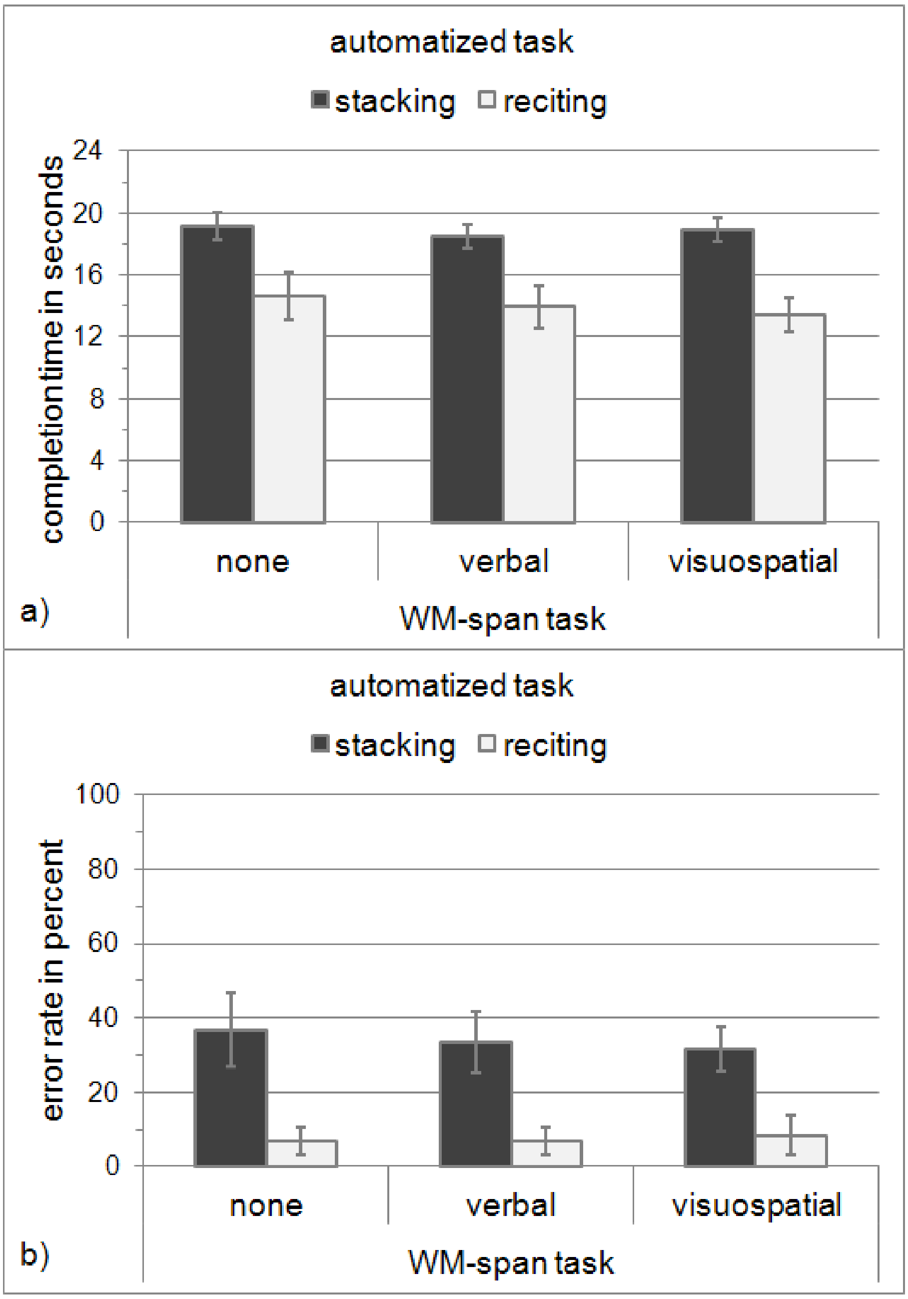

Figure 4. a) Completion time of the automatized tasks (stacking and reciting) in seconds with mean square errors during the WM-span tasks (none, verbal, and visuospatial). b) Error rate of the automatized tasks (stacking and reciting) in percent with mean square errors during the WM-span tasks (none, verbal, and visuospatial).

Figure 5

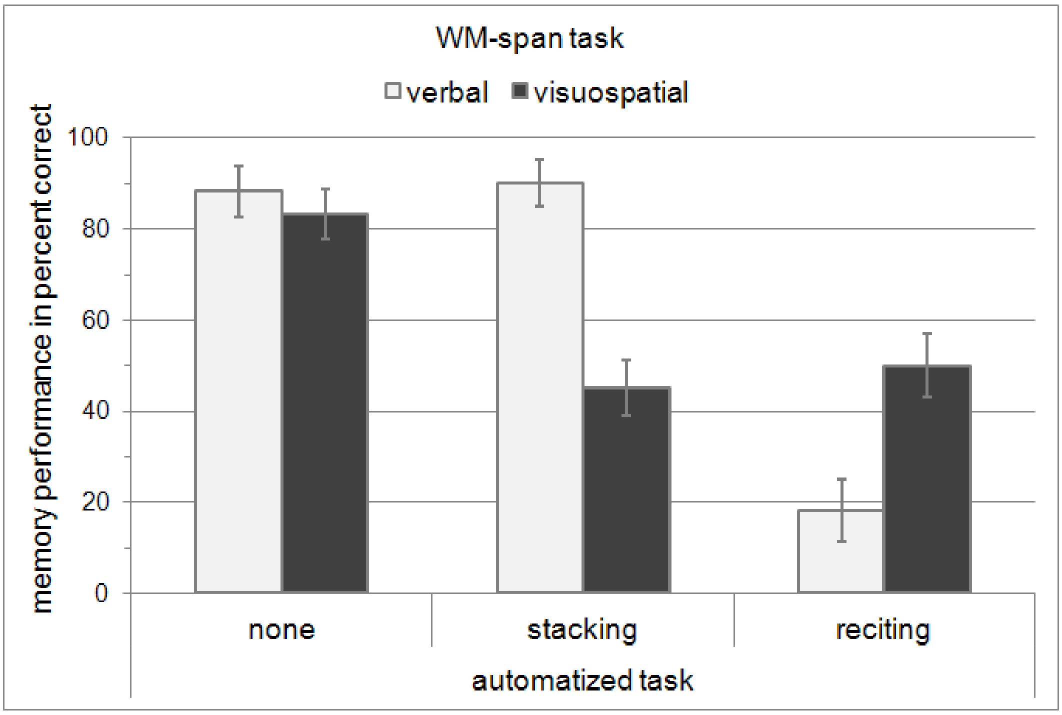

Figure 5. Memory performance of the WM-span tasks (verbal and visuospatial) in percent correct with mean square errors during the automatized tasks (none, stacking, and reciting).

Appendix

The German poem and its English translation.

Nun bin ich hier und soll	Now here I am and shall
find ich es nun ganz toll	do I find it very nice
will ich mich nun beklagen	I want to complain
ein Kurzgedicht aufsagen	and recite a short poem
Man hat mich nicht gefragt	No one asked me
ob ich Gedichte mag	whether I like poems
ob ich Gedichte hasse	whether I hate poems
ob ich mich drauf einlasse	whether I take up with them
Ich halte tapfer aus	I hold out bravely
auch ohne Standapplaus	not longing for standing ovations
ich werd„s zu Ende bringen	I will bring it to an end
die Sache schon bezwingen	and conquer the thing
Nun folgt der letzte Part	Now the final part is come
mir bleibt auch nichts erspart	and I am not spared a thing
nun soll ich auch noch sagen	now I shall even say
ich hab es gut ertragen	I have taken it well

Betreff: Manuscript Submitted
Von: amedhin@apa.org
Datum: 07.11.2011 17:50
An: rebecca.foerster@uni-bielefeld.de

Dear Dr. Foerster-

We are in receipt of your manuscript titled The interplay of attention working memory and long-term memory: An interference study with automatized tasks. You should receive some notice of the status of your manuscript within 60 to 90 days. During this period you should not submit your manuscript to another journal.

Your manuscript number is 2011-[redacted]. To receive an e-mail detailing the history of your manuscript visit http://www.jbo.com/jbo3/dsp_checkhistory.cfm?journal_code=xhp2 and enter your [redacted] as username and your manuscript number as password.

If your manuscript is accepted for publication, you will be required to transfer your copyright to APA, provide full disclosure of any conflict of interest, and certify compliance with APA ethical principles. Also, please read the APA's Open Letter to Authors located at http://www.apa.org/journals/authors/.

Sincerely,

Journal of Experimental Psychology: Human Perception and Performance Editorial Office

Universität Bielefeld

Erklärung zur Urheberschaft - Kumulative Dissertation von Rebecca M. Förster

Hiermit bestätigen wir, dass das Manuskript „*The interplay of attention, working memory, and long-term memory: An interference study with automatized tasks*“ selbstständig von Frau Förster erstellt wurde. Frau Förster war hauptverantwortlich für die Planung, Durchführung und Auswertung dieser empirischen Arbeit. Außerdem verfasste und überarbeitete sie eigenständig alle Versionen des Manuskripts.

(Elena Carbone)

(Werner X. Schneider)

(Rebecca M. Förster)

3 General Discussion

Based on the experiments reported in manuscripts 1, 2, and 3 (Foerster et al., 2011a, 2011b, 2011c), insights were gained into attention, eye movement, and memory control during and after the learning and automatization of a sequential high-speed sensorimotor task. In Foerster et al. (2011a), eye movements that occurred during sensorimotor learning were recorded in order to reveal possible changes in overt visual selection while the participants learned and automatized the task. Throughout the learning process, the participants primarily fixated on placement locations for objects held in the right hand shortly before placing them. After 14 consecutive days of practice, the absolute eye-hand time span (the time delay between eye and hand reaching a task-relevant location) was shorter in comparison to the first training day, while the eye-hand unit span (the delay in task steps between eye and hand reaching a task-relevant location) was not affected and the relative eye-hand time span (absolute eye-hand time span corrected by trial duration) had even increased. In addition, fewer fixations were performed on the last training day (high degree of automatization) compared to the first training day (low degree of automatization). However, scan paths were very similar across training days and between participants, i.e., the same task-relevant locations were fixated in a similar order by all participants across training days. These results suggest that sequences of saccades (scan paths) are transferred to LTM during learning and are retrieved from LTM after automatization.

In Foerster et al. (2011b), trained participants from Foerster et al. (2011a) performed the automatized high-speed bimanual sensorimotor task under normal illumination conditions and also in the dark. Eye movements while acting in the dark were compared to eye movements while acting in the light in order to reveal possible similarities across illumination conditions. Participants performed slower in the dark, probably because no visual information could be used to control hand movements. However, despite the longer task durations in the dark, the number of fixations did not increase. Instead, fixation rates decreased and fixation durations increased. While the relative eye-hand time span and the eye-hand unit span were reduced in the dark, the absolute eye-hand time span did not differ between illumination conditions. Moreover, the eye-hand time span was mainly positive not only in the light, but also in the dark, indicating that the eye guided the hand even when no visual input was available. Most important, scan paths were highly similar between light and the dark illumination conditions. This result of highly systematic eye movements in the dark supports the hypothesis established in Foerster et al. (2011a) of a more LTM-based mode of attention and eye movement control in higher stages of automatization.

In Foerster et al. (2011c), trained participants from Foerster et al. (2011a) had to perform the automatized high-speed sensorimotor task as well as an additionally trained high-speed verbal task while they had to maintain either visuospatial or verbal information in WM. Results revealed that neither the automatized sensorimotor nor the automatized verbal task was influenced by an additional WM load. Importantly, the WM performance was selectively disturbed by the execution of the intervening automatized tasks. The automatized sensorimotor task impaired the visuospatial memory span, but not the verbal memory span. The automatized verbal task impaired the verbal memory span more than the visuospatial memory span. Therefore, even automatized tasks recruited internal attention control and domain-specific WM resources.

In the following, the main results of the present dissertation will be discussed with regard to two important research topics. On the one hand, results will be used to evaluate the validity of the traditional automatization theories and a new conception of automatization will be outlined that fits better to the observed results. On the other hand, results have important implications for the understanding of the interplay of attention and memory during and after sensorimotor learning and automatization. A plausible assumption about how these components work together during automatic processing will be provided at the end of the present synopsis.

3.1 Validity of automatization theories

The study reported in manuscript 1 (Foerster et al., 2011a) revealed, in accordance with other eye movement studies on natural tasks, that humans use their eyes to guide their hand movements. This eye-hand guidance is characterized by a fixation on the target location of the next hand movement shortly before the hand actually arrives at the target location. The grasp zone of an object is, for instance, fixated before the object is grasped at that location and a placement region is fixated before an object is placed in it. Executing a saccade to a spatial location requires a prior covert shift of attention to that location (Deubel & Schneider, 1996; Schneider, 1995). Hence, attention allocation to hand-target locations seems to be necessary to perform a sensorimotor task. Directing attention to an object before automatically acting on it is compatible with Logan‟s (1988, 1990, 1992) instance theory and Neumann‟s (1984, 1990) direct parameter specification theory of automatization. Both conceptions of automatization allow for attentional control during automatic processing. In contrast, the necessity to direct attention to an object before performing an automatic process is not compatible with Schneider and Shiffrin‟s (1977a, 1977b) two-process theory of human information processing, unless the additional assumption of an *automatic-attention response* (Schneider & Shiffrin, 1977b, p. 143) is called in.

In speed stacking, placement areas were fixated. These placing areas did not contain any visual stimulus before the placing action was completed. Accordingly, not stimuli, but „empty‟ target locations were attended and fixated in speed stacking. Attention allocation to „empty‟ locations is neither compatible with Schneider and Shiffrin‟s (1977a, 1977b) theory, nor with Logan‟s (1988, 1990, 1992) instance theory. An automatic-attention response is triggered by a specific stimulus in Schneider and Shiffrin (1977a, 1977b), and the instance theory (Logan, 1988, 1990, 1992) assumes that attention is allocated to all stimuli in the environment causing memory encoding and memory retrieval. If attention is allocated to locations that do not contain visual stimuli, LTM retrieval cannot be triggered. However, Neumann‟s (1984, 1990) direct parameter specification theory of automatization is compatible with the finding of attention allocation to „empty‟ locations. Parameter specification by LTM information is not bound to visual stimuli. Not only a specific stimulus, but also a specific feature of a stimulus or a specific location can be marked in LTM as important for an automatized task. Therefore, attention can be shifted to a specific location that is important for the automatized sensorimotor task regardless of whether or not the location contains a visual stimulus.

The observed changes of eye movement patterns during learning and automatization (Foerster et al., 2011a) also argue against Schneider and Shiffrin‟s and Logan‟s conceptions of automatization. As was mentioned above, humans should not shift their attention and should not perform eye movements during automatic processing, and if they do, not to „empty‟ locations (Schneider & Shiffrin, 1977a, 1977b). Eye movement patterns should also not change during learning and automatization under the assumption that the mechanisms of attention allocation do not change with the accumulation of instances (Logan, 1988, 1990, 1992). In contrast, the assumption that less attentional mechanisms are necessary during automatic than during non-automatic processing (Neumann, 1984, 1990) is compatible with changing eye movement patterns during learning and automatization. However, the results can also be explained by the proposal of changed attention control. The additional fixations early in the learning process seem to reflect a more sensory-based mode of attention control. After automatization, LTM information controls where-to-attend-next and hence also where-to-look-next. LTM-based attention control suffices to execute the automatized sensorimotor task reflected by fewer fixations late in the learning process.

The study reported in manuscript 2 (Foerster et al., 2011b) revealed that eye movement characteristics of trained participants during the execution of an automatized sequential high-speed sensorimotor task were very similar in light and dark illumination conditions. Participants fixated on hand-target locations in the dark before the corresponding hand movement was executed. None of the theories of automatization (Logan, 1988, 1990, 1992; Neumann, 1984, 1990; Schneider & Shiffrin, 1977a, 1977b) is compatible with the finding of systematic attention and eye movement control in the dark. According to the theories, attention – if at all – is allocated to sensory input. Visual sensory input is not available in the dark, so the automatization theories would predict that attention is allocated to auditory or haptic sensory input in the dark. In placing actions, auditory and haptic feedback is available slightly after the hand movement is completed. In Foerster et al. (2011b), the eye was always slightly ahead of the hands even in the dark. Hence, attention was not directed according to sensory (auditory or haptic) feedback. Saccade target locations could only have been specified by LTM information in the dark. LTM can direct attention to task-relevant locations in the dark and control the eye movements before the hands are actually moved, so that there is no need to wait for auditory or haptic feedback. This result therefore strongly supports the proposal that LTM information specifies where-to-attend and where-to-look next.

The study reported in manuscript 3 (Foerster et al., 2011c) revealed that the maintenance of verbal and visuospatial memory spans was selectively impaired by the execution of a verbal or a visuospatial automatized high-speed task within the maintenance interval. Therefore, even automatized tasks seem to interfere with other tasks as long as they belong to the same domain. This result is incompatible with the assumption that automatized tasks can be executed in parallel with other tasks without any interference (Schneider & Shiffrin, 1977a, 1977b). Moreover, the result of domain-specific interference is not compatible with direct LTM control of automatic processing (Logan, 1988, 1990, 1992; Neumann, 1984, 1990). Such a direct link between LTM representations (instances or traces) and response output would imply that selected LTM representations do not have to be maintained within a domain-specific WM store until task execution and thus should not interfere with parallel WM retention. The observed interference between automatized processing and WM retention documents the need of domain-specific WM processes and internal attention for LTM retrieval during the execution of automatized tasks.

On the basis of the results revealed by the present work the understanding of attention, eye movements, and memory processes during non-automatic and automatic processing should be modified in the following way (the also last row of Table 1). Attention control is more sensory-based during non-automatic processing, while a more LTM-based mode of attention control characterizes automatic processing. With high degrees of automatization, LTM information determines which sensory input has to be attended and where this relevant sensory input can usually be found in a specific environment. Shifts of visuospatial attention in turn determine where to look next. The gathered visual information then provides updated information for action control. At the same time LTM representations are updated with respect to the validity of their predictions.

3.2 The interplay of attention and memory during and after automatization

The following explications will illustrate how the findings reported in the three manuscripts (Foerster et al., 2011a, 2011b, 2011c) can shed light on the interplay of attention and memory during and after automatization.

The study reported in manuscript 1 (Foerster et al., 2011a) revealed that, with a high degree of automatization, fewer fixations are needed to perform the same number of object-related actions. Moreover, scan paths were highly similar between participants. These results indicate that more saccades were necessary to perform the task in early stages of automatization than in later stages of automatization. The need for many fixations in the beginning of the learning process implies a sensory-based mode of attention control. In early stages of learning, only visual input from the periphery can be used to specify the next saccade-target location because task-related LTM information is not yet available. The decreased number of fixations in later stages of automatization implies a LTM-based mode of attention control. In later stages of learning, it should be possible to retrieve a task-relevant scan path from LTM because the LTM contains the sequence of task-relevant saccade-target locations.

The study reported in manuscript 2 (Foerster et al., 2011b) revealed highly similar scan paths and a similar number of fixations in the dark compared to in the light while the participants performed an automatized sequential high-speed sensorimotor task. In addition, the eyes saccaded to a location in the dark before the hands reached that location. These results strongly support the idea of a LTM-based mode of eye movement control after automatization, i.e., a scan path with a fixed number of fixations is retrieved from LTM to perform the automatized task. Without the consideration of a LTM-based mode of attention control, it is difficult if not impossible to explain why the eyes moved to the hand-target locations in the dark before the locations could have been selected based on sensory (auditory or haptic) information.

The study reported in manuscript 3 (Foerster et al., 2011c) revealed domain-specific disturbance of the maintenance of memory spans while executing highly automatized sequential high-speed tasks. Therefore, even in case of a high degree of automatization, task execution is not capacity-free and needs attention control. Moreover, domain-specific WM processes seemed to contribute to the execution of automatized tasks. The most likely explanation for the domain-specific interference reported in Foerster et al. (2011c) is that internal attention and domain-specific WM processes were needed during the execution of automatized tasks, so that they competed with the memory-span tasks for internal attention and WM processes.

According to the reported results, the following interplay of attention, eye movement, and memory processes during the execution of automatized sensorimotor tasks seems to be plausible. Relevant LTM representations have to be retrieved via internal attention. These LTM selected representations have to be transferred to domain-specific WM stores and maintained until the representations determines where-to-attend and where-to-look in the environment. Relevant information gathered from fixations finally determines the execution of the next sensorimotor sub-action.

In summary, results of the present dissertation shed light onto the mechanisms underlying sensorimotor automatization and the interplay between attention, eye movements, and memory processes during and after automatization. Attention control seems to shift gradually from a more sensory-based to a more LTM-based mode during learning and automatization of sensorimotor tasks. During automatic processing, internal attention seems to be needed for retrieving task-relevant LTM information and to maintain it within domain-specific WM stores. The activated LTM information determines which sensory input has to be attended and where this input can usually be found. These LTM-based shifts of visuospatial attention are accompanied by saccades to the relevant visual input. Visual information is gathered via fixation on the relevant visual input and provides updated information for just-in-time controlled task executions. At the same time the LTM skill content is updated with respect to the gathered task-related information to make even better predictions for future attention and gaze control.

4 References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press.

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 780-790.

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*, 201-208.

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417-423.

Baddeley, A. D. (2003). Working memory: Looking back and locking forward. *Nature Reviews Neuroscience*, *4*, 829-839.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 12.1-12.29.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (47-89). Academic Press.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523-547.

Bundesen, C., & Habekost, T. (2008). *Principles of visual attention*. Oxford, England: Oxford University Press.

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*, *9*, 613-625.

Chun, M. M, Golomb, J. D., & Turk-Browne, N. B. (2010). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73-101.

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *Neuroimage*, *37*, 343-360.

Cowan, N. (2000). *Processing limits of selective attention and working memory: Potential implications for interpreting*. Paper presented at the ASCONA II, Geneva, Switzerland.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*, 1827-1837.

Dodge, R. (1900). Visual perception during eye movement. *Psychological Review*, *7*, 454-465.

Dodge, R. (1903). Five types of eye movement in the horizontal meridian plane of the field of regard. *American Physiological Society*, *8*, 307-329.

Droll, J. A., & Hayhoe, M. M. (2008). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1352-1365.

Droll, J. A., Hayhoe, M. M., Sullivan, B. T., & Triesch, J. (2008). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1352-1365.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active Vision: The psychology of looking and seeing*. New York: Oxford University Press.

Findlay, J. M. (2009). Saccadic eye movement programming: Sensory and attentional factors. *Psychological Research*, *73*, 127-135.

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011a). Saccadic eye movements in a high-speed bimanual stacking task: Changes of attentional control during learning and automatization. *Journal of Vision*, *11*(7), *9*, 1-16.

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011b). Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task. *Manuscript submitted for publication in Journal of Vision*.

Foerster, R. M., Carbone, E., & Schneider, W. X. (2011c). The interplay of attention, working memory, and long-term memory: An interference study with automatized tasks. *Manuscript submitted for publication in Journal of Experimental Psychology: Human Perception and Performance*.

Glöckner, A, & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology*, *34*, 1055-1075.

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

Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*, 43-64.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*, 49-63.

Hebb, D. O. (1949). *Organization of Behavior*. New York: Wiley.

Helmholtz, H. von (1910). *Handbuch der physiologischen Optik III: Die Lehre von den Gesichtswahrnehmungen*. Hamburg: Voss.

Hofmann, W., Gschwendner, T., Frise, M., Wiers, R. W., & Schmitt, M. (2008). Working memory capacity and self-regulatory behavior: Toward an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology*, *95*, 962-977.

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, *137*, 163-181.

Irwin, D. E. (1992). Memory for position and identity across saccadic eye movements. *Cognitive Psychology*, *23*, 420-456.

Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp.125-155). Cambridge, MA: MIT Press.

James, W. (1890). *The Principles of psychology*. Holt, New York.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-hand coordination in object manipulation. *The Journal of Neuroscience*, *21*, 6917-6932.

Küper, K., & Heil, M. (2010). Letter search and relatedness proportion: Further electrophysiological evidence for the automaticity of semantic activation. *Neuroscience Letters*, *482*, 26-30.

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*, 3559-3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*, 1311-1328.

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology*, *11*, 1215-1220.

Land, M. F., & Tatler, B. W. (2009*). Looking and acting*. New York: Oxford University Press.

Logan, G. D. (1988). Towards an instance theory of automatization. *Psychological Review*, *95*, 492-527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, *22*, 1-35.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 883-914.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*, 376-400.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1999). Attention and automaticity: Toward a theoretical integration. *Psychological Research*, *62*, 165-181.

Luck, S. J., & Hollingworth, A. (2008). *Visual memory*. Oxford, England: Oxford University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279-281.

Mennie, N., Hayhoe, M., & Sullivan, B. (2006). Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research*, *179*, 427-442.

Neumann, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz & A. F. Sanders (Eds.), *Cognition and motor processes* (pp. 255-293). Berlin: Springer.

Neumann, O. (1990). Direct parameter specification and the concept of perception. *Psychological Research*, *52*, 207-215.

Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Eds.), *Cognitive skills and their acquisition (pp. 57-84)*. Hillsdale, New Jersey: Erlbaum.

Noton, D., & Stark, L. W. (1971a). Scan paths in eye movements during pattern perception. *Science*, *171*, 308-311.

Noton, D., & Stark, L. W. (1971b). Scan paths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*, 929-942.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: when it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*, 327-334.

O‘Regan, J. K. (1992). Solving the “real” mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461–288.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Palmero, R., & Rhodes, G. (2006). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia*, *45*, 75-92.

Pashler, H. (1988). Familiarity and the detection of change in visual displays. *Perception and Psychophysics*, *44*, 369-378.

Pashler, H., Johnston, J., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, *52*, 629-651.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics*, 16, 283-290.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, HJ: Lawrence Erlbaum Associates.

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye-hand coordination during learning of a novel visuomotor task. *The Journal of Neuroscience*, *25*, 8833-8842.

Saling L. L, & Phillips, J. G. (2007). Automatic behavior: Efficient not mindless. *Brain Research Bulletin*, *73*, 1-20.

Schacter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 501-518.

Schneider, W. X. (1995). VAM: A neuro-cognitive model for visual attention control of segmentation, object recognition and space-based motor action. *Visual Cognition*, *2*, 331-376.

Schneider, W., & Shiffrin, R. M. (1977a). Controlled and automatic human information processing: 1. Detection, search and attention. *Psychological Review*, *84*, 1-66.

Schneider, W., & Shiffrin, R. M. (1977b). Controlled and automatic human information processing: 2. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127-189.

Smyth, M. M. (1996). Interference with rehearsal in spatial working memory in the absence of eye movements. *Quarterly Journal of Experimental Psychology*, *49A*, 940-949.

Smyth, M. M., & Scholey, K. A. (1994). Interference in immediate spatial memory. *Memory and Cognition*, *22*, 1-13.

Squire, L. R. (1982). The neuropsychology of human memory. *Annual Review of Neuroscience*, *5*, 241-273.

“t Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., & Einhäuser, W. (2009). Gaze allocation in natural stimuli: Comparing free exploration to head-fixed viewing conditions. *Visual Cognition*, *17*, 1132-1158.

Theeuwes, J., Kramer, A. F., & Irwin, D. E. (2011). Attention on our mind: The role of spatial attention in visual working memory. *Acta Psychologica*, *137*, 248-251.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97-136.

Tulving, E. (1985). How many memory systems are there? *American Psychologist*, *40*, 385-398.

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). *Trends in Cognitive Sciences*, *9*, 445-453.

Wundt, W. (1903*). Grundzüge der physiologischen Psychologie*, Vol. III, 5th ed. Leipzig: Engelmann.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.