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

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

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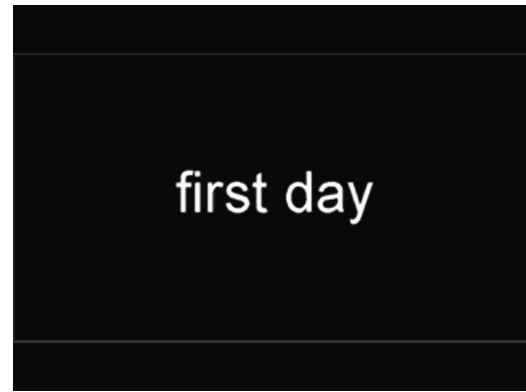
Movie 1. A participant performing the speed-stacking task on the first, second, and last training days.

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).

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
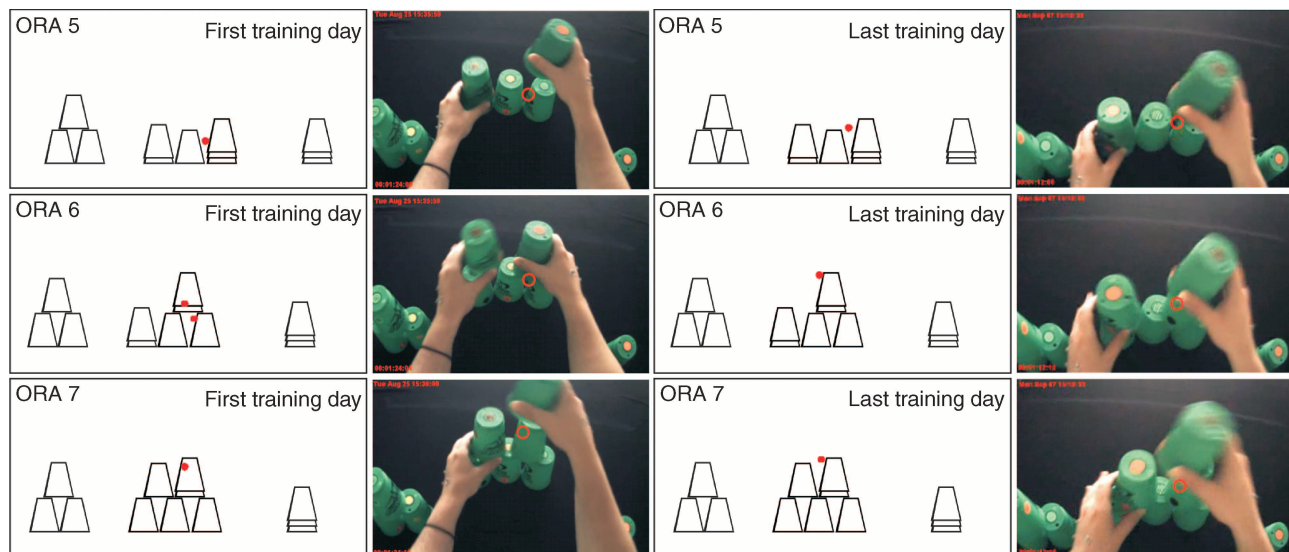

Figure 1. An example of ORA boxes for analyzing the gaze positions. The cup's starting configurations for ORAs 5 to 7 are represented in boxes and in video frames on the (left) first and (right) last training days. In ORAs 5 and 7, the right hand is manipulating a cup or stack. In ORA 6, the left hand 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
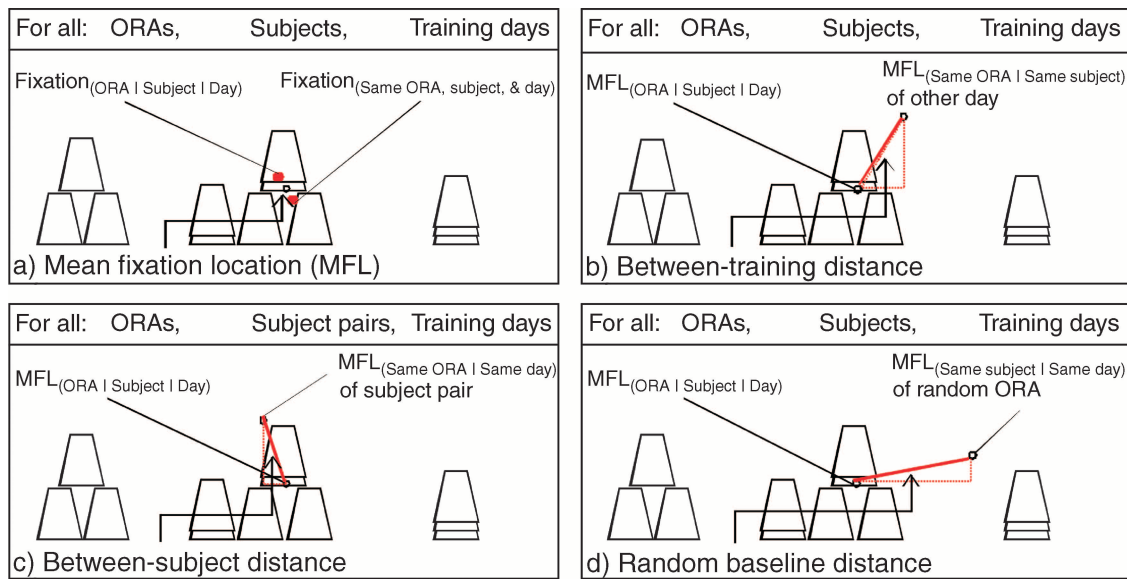
Figure 2. Schematic illustration of the calculations of (a) mean fixation location, (b) between-training distance, (c) between-subject distance, (d), and random baseline distance. (a) Mean fixation locations are the averaged fixation locations within the same ORA, subject, and training day. (b) Between-training 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,

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

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.

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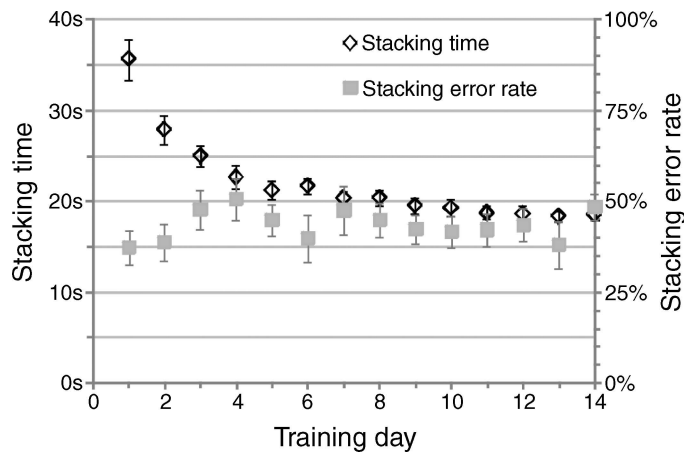
Figure 3. Mean speed-stacking time (dark gray diamonds and left *y*-axis) and error rates (light gray squares and right *y*-axis) with error bars indicating the standard error of the mean per training day.

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 $t(8) =$
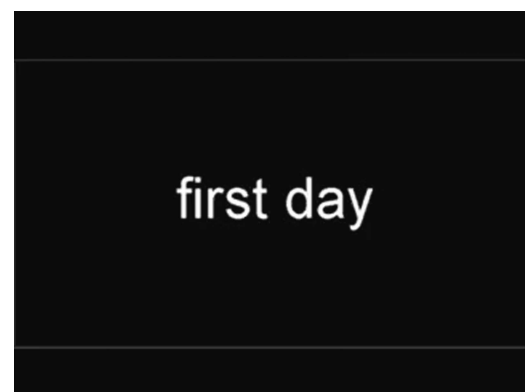
Movie 3. Eye movements during the first, second, and last training days of a participant in slow motion.
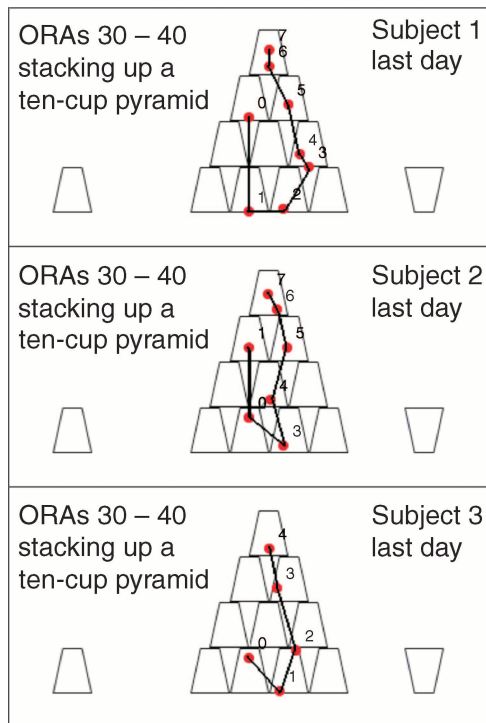
Figure 4. Scan paths of three different participants while stacking up the ten-cup pyramid out of a ten-cup stack on the last training day. Participants' fixations made during 10 successive ORAs (30 to 40) were superimposed on the schematic illustration of the upstacked ten-cup pyramid. Cups are illustrated 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
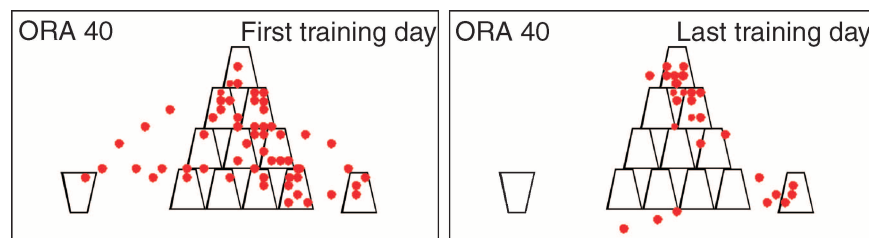

Figure 5. Fixations of all participants for ORA 40 on the (top) first and (bottom) last training days. Cups are illustrated as trapeziums and fixations are illustrated as red dots.

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

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.