MASTER THESIS

A COMPARISON OF METHODS IN THE RESEARCH OF VISUAL WORKING MEMORY CAPACITY

Mentor: Dr.sc. Marko Polič

Co-mentor: Dr.sc. Dražen Domijan

LJUBLJANA, 2008

DOMAGOJ ŠVEGAR
MAGISTRSKA NALOGA

PRIMERJAVA RAZLIČNIH METOD RAZISKOVANJA KAPACITETE VIZUALNEGA DELOVNEGA SPOMINA

Mentor: Dr.sc. Marko Polič

Somentor: Dr.sc. Dražen Domijan

LJUBLJANA, 2008

DOMAGOJ ŠVEGAR
I am grateful to:

Marko Polič and Dražen Domijan – my mentors,
Valentin Bucik and Marko Polič for administrative help,
Ljiljana Perković and Marko Polič for translation into Slovene,
Ljiljana Švegar for lectoring the text,
Marko Maliković for technical assistance,
Zvezdan Pirtošek for outstanding lectures in the field of neuroscience,
my family and colleagues for support.

I dedicate this work to my baby daughter Lucija.

In Ljubljana, 3 June 2008

Domagoj Švegar
1 ABSTRACT

1.1 INTRODUCTION

1.2 METHOD

1.3 RESULTS AND DISCUSSION

1.4 CONCLUSION

2 INTRODUCTION

2.1 DEFINITION OF THE WORKING MEMORY

2.2 DEVELOPMENT OF VISUAL WORKING MEMORY CAPACITY RESEARCH

2.2.1 THE EMBARCATION – SPERLING’S WORK

2.2.2 METHODOLOGICAL IMPROVEMENT: PHILLIPS’S CONTRIBUTION

2.2.3 QUANTITATIVE APPROACH BY PASHLER

2.2.3.1 Purdy et al.’s research

2.2.3.2 Pashler's formula

2.2.4 BEGINNINGS OF MODERN RESEARCH: VOGEL ET AL.’S EXPERIMENTS

2.3 REDUCING THE CONTRIBUTION OF VERBAL WORKING MEMORY

2.4 GENERALITY OF THE VISUAL WORKING MEMORY CAPACITY ESTIMATE

2.5 VISUAL WORKING MEMORY FOR MULTIFEATURE OBJECTS

2.5.1 VOGEL ET AL.’S THEORY

2.5.1.1 Color-color conjunction

2.5.2 WHEELER & TREISMAN'S THEORY

2.5.3 RESOLUTION OF "VOGEL ET AL. VERSUS WHEELER & TREISMAN" DEBATE

2.5.3.1 Wheeler & Treisman's research

2.5.3.2 Research of Alvarez & Cavanagh
2.5.3.3 Švegar & Domijan's experiments______________________________47

2.6 IS VISUAL WORKING MEMORY CAPACITY INDEED SO LIMITED? __________49
2.6.1 ACCUMULATED DECISION ERRORS ______________________________49
2.6.2 ENCODING LIMITATIONS ______________________________________51
2.7 PROBLEM OF THE PRESENT STUDY________________________________52

3 METHOD____________________________________________________55

3.1 PARTICIPANTS________________________________________________55
3.2 INSTRUMENTS ______________________________________________55
3.3 STIMULI AND PROCEDURE ____________________________________55
3.3.1 PRESENTATION MANIPULATION ______________________________56
3.3.2 CHANGE __________________________________________________56
3.3.3 TEST MANIPULATION ______________________________________61
3.3.4 PARTICIPANTS’ TASKS ______________________________________61

4 RESULTS AND DISCUSSION____________________________________63

4.1 THE EFFECTS OF THE TYPE OF INITIAL PRESENTATION AND THE TYPE OF TEST
DISPLAY ON PARTICIPANTS' PERFORMANCE _______________________63
4.1.1 DESCRIPTIVE RESULTS ______________________________________63
4.1.2 ACCURACY PERCENTAGE AS A DEPENDENT VARIABLE __________64
4.1.3 PASHLER'S CAPACITY ESTIMATE AS A DEPENDENT VARIABLE ______65
4.1.4 REACTION TIME AS A DEPENDENT VARIABLE __________________66
4.2 IS VISUAL WORKING MEMORY CAPACITY INDEED SO LIMITED, OR IS IT
UNDERESTIMATED DUE TO ENCODING LIMITATIONS? _________________68
4.3 **DRAWBACKS OF PASHLER’S (1988) VISUAL WORKING MEMORY CAPACITY ESTIMATION PROCEDURE**

4.3.1 **FULL VERSUS PARTIAL TEST DISPLAY: IS THERE A DIFFERENCE IN PARTICIPANTS' PERFORMANCE REGARDING THE TYPE OF TEST DISPLAY?**

4.3.1.1 *Hypothetical situation A: Lenient criterion*

4.3.1.2 *Hypothetical situation B: Strict criterion*

4.3.1.3 *Hypothetical situation C: A realistic situation*

4.3.2 **ANALYSES OF HIT RATE AND FALSE ALARM RATE OBSERVED IN THE PRESENT STUDY**

4.3.2.1 *A change in criterion of responding*

4.3.3 **A BETTER PROCEDURE OF VISUAL WORKING MEMORY CAPACITY ASSESSMENT**

4.3.3.1 *Alvarez & Cavanagh's method: 75% threshold divided by 2*

4.3.3.2 *The best method of visual working memory capacity estimation*

5 **CONCLUSION**

6 **REFERENCES**

7 **POVZETEK**

7.1 **UVOD**

7.2 **METODA**

7.3 **REZULTATI IN RAZPRAVA**

7.4 **SKLEP**
1 ABSTRACT

1.1 Introduction

The subject of the present study was a review and a comparison of behavioral research methods in the field of visual working memory capacity research. The most common paradigm in that field is the change detection paradigm - a procedure in which two sets of stimuli are presented. These two displays are separated by an interstimulus interval, during which a change may occur, usually on one of the stimuli. Participants' task is to answer if change had occurred or not, and visual working memory is then estimated through analyses of their performance. Phillips (1974) was one of the first authors that used such a method to investigate visual working memory. He tachistoscopically presented pairs of displays filled with dots, separated by a variable inter-stimulus interval. Two displays within a trial were either identical, or differed in the addition or the removal of one dot. He found that participants' performance was excellent with short inter-stimuli intervals (approximately 100 ms or shorter), but prolongation of retention interval resulted in a decreased change detection accuracy. These experiments have brought detailed insight into the nature of iconic and working memory, and clarified boundaries between them: Phillips (1974) concluded that iconic memory has vast capacity, but very short duration, in contrast to visual working memory, which is capable of storing only a small amount of information, but can maintain them for a longer period of time.

Pashler (1988) was the first author who managed to develop a reasonably valid procedure of visual working memory capacity quantification. His method is derived from signal detection theory and is based on hit rates ($H$) and false alarm rates ($FA$), and also takes account of the number of initially presented items ($IP$). In trials that contain a change, a hit occurs if participant answers "change present", while the answer "change absent" results with a miss. Hit rate is a proportion of "change present" answers in all trials containing a change. In trials that do not contain a change, a false alarm occurs if participants answers "change present", while the answer "change absent" results in a correct rejection. False alarm rate is a proportion
of "change present" answers in all trials that do not contain a change. As a hit cannot occur in trials that do not contain a change, a false alarm cannot occur in trials that do contain a change. The logic of Pashler's (1988) formula is the following:

Presuming that participants held a certain amount of items in memory \((C)\), when one of them changed, hit would follow, and if the changed item was not stored in memory, miss would occur. When participants did not detect a change, they would answer "change absent", but it must be taken into account that in a certain proportion of the trials they were guessing. Guessing relates to answering "change present" in the cases where there was actually no change (false alarms). Along with the variables mentioned and set size, Pashler (1988) gave the formula for determining visual working memory capacity \((C)\):

\[
H = \frac{C}{IP} + \frac{IP - C}{IP} * FA,
\]

where

- \(C\) is the number of presented objects that are stored in memory (visual working memory capacity);
- \((C / IP)\) is the proportion of trials in which an item that is stored in memory changes
- \((IP - C)\) is the number of presented objects that are not stored in memory
- \((IP - C) / IP\) is the proportion of trials in which an item that is not stored in memory changes
- \(FA\) is the guessing rate (guessing rate is estimated from the false alarm rate – from the proportion of no-change trials in which the subject guesses there was a change).

When the equation above is transformed, a formula for visual working memory capacity assessment can be extracted:

\[
C = \frac{IP * (H - FA)}{1 - FA}
\]
In the majority of recent studies (e.g. Vogel & Luck, 1997; Vogel, Woodman, & Luck, 2001; Wheeler & Treisman, 2002), visual working memory capacity was assessed by this procedure.

To assure that visual working memory capacity estimate is not contaminated by the contribution of verbal working memory, three types of procedures were commonly used in these experiments:

a) stimuli that are difficult to name
   Non-verbal stimuli, such as Phillips's (1974) dots, or Alvarez & Cavanagh's (2004) random polygons, are perfect for visual working memory capacity research because they have no corresponding auditory code. However, these stimuli are rare, difficult to construct and inadequate for some experimental designs.

b) brief presentation of stimuli
   Since phonological coding is a slower process than visual coding, very short presentation of visual material minimizes the contribution of verbal working memory

c) verbal load
   This procedure is also known as auditory or articulatory suppression, and it refers to meaningless repetitive speech that disrupts auditory coding and has almost no other effects. Since it does not impair performance on tasks that require visual coding (e.g. Murray, 1967; Klapp, Marshburn, & Lester, 1983; Richardson, 1984; Baddeley, 1986), this procedure allows testing visual working memory by longer display durations (Frick, 1985).

Commonly, two of these three procedures are used to minimize the contamination by verbal working memory.

Although there are debates about the nature of visual information storage and maintainance, all prominent researchers have reached the consensus that visual working memory is surely limited to not more than 4-5 items.

Some of their studies contained several control experiments (e.g. Vogel et al., 2001; Wheeler & Treisman, 2002) to see whether small observed capacity is the consequence of limitations in visual working memory system, or a result of some other process.
As the contribution of verbal working memory can cause overestimation of true visual working memory capacity, accumulated decision errors can lead to its underestimation. Specifically, at change detection tasks with large set sizes, participants have to compare all memory representations of the initial display to all memory representations of the test display. If displays contain 12 items, then 12 "same-different" decisions have to be made, and each such decision is subject to a certain degree of error. To check whether poor performance in change detection tasks are an artefact of these errors or the consequence of limitations in visual working memory, several control experiments were conducted. In these experiments, performance in classical change detection tasks was compared to performance in conditions in which only one decision had to be made at test. In one such experimental condition, an item that could have changed was cued in test display, and participants were informed that a change surely did not occur on any other item except the cued one. Performance in that condition was the same as in the condition with no cue, but another control experiment was conducted, because it was hypothesized that the mere presence of uncued items may have impaired performance by triggering automatic decision processes. To investigate that possibility, in one experimental condition, only one item was presented after retention interval. Since complete removal of unchanged items disrupts memory performance (Klemer, 1963; Vogel et al., 2001), outlined squares (placeholders) were placed in the locations of irrelevant stimuli. However, even in that condition, in which participants were forced to make not more then only one "same-different" decision, performance was no better, when compared to performance in classical change detection tasks with whole test displays. Therefore, it was concluded that visual working memory for single-feature objects, measured in change detection tasks, is not underestimated due to accumulated decision errors (Vogel et al., 2001; Wheeler & Treisman, 2002).

Apart from accumulated decision errors, encoding limitations were another possible cause of underestimation of visual working memory capacity in change detection tasks. Since in almost all experiments initial presentation of stimuli is tachistoscopyal (in order to prevent verbal encoding of stimuli), it is possible that such short presentation intervals (they usually last 100 to 150 ms) are insufficient to adequately encode all stimuli, especially at large set sizes. The possibility that such limitations are the cause of underestimation in assessing visual working memory
capacity was discarded after two experiments. In the first of them visual search tasks were used. A colored square was presented, and after 500-ms delay, an array of 4, 8 or 12 colored squares was presented, in duration of 100 ms. Participants had to answer whether that array contained the target color or not. Accuracy of the participants' responses was near 100%, even with the largest set sizes, so it was concluded that 100 ms is enough time for participants to identify colors of all squares (Vogel et al., 2001).

In another experiment, initial presentation interval was extended to 500 ms, and participants' performance was then compared to the condition with 100-ms intervals of the initial presentation, while all other variables were held constant. Since participants' performance was approximately the same in these two conditions, it was concluded that 100-ms interval is enough time to adequately encode up to 12 items (Vogel et al., 2001).

Thus, the results of all these experiments suggest that it is certain that visual working memory capacity for single-feature objects indeed equals around 4-5 items (or less, depending on the type of material to be remembered), and it was also found that that estimate is neither overrated by the contribution of verbal working memory, nor underestimated by accumulated decision errors, encoding limitations or anything else.

That belief was challenged in the present study. In previous studies, different kinds of variables were varied (e.g. set size, type of test display, type and complexity of visual material, duration of interstimulus interval, etc.), but all the experiments were based on the same template:

a) short simultaneous exposition of stimuli
b) retention interval during which a change may occur
c) test.

Since all relevant behavioral research was based on the classical change detection paradigm, the goal of this experiment was to test the memory for simple single-feature objects by a different procedure.

Encoding limitations were in the focus of the present study. Although the results of several control experiments (which were described earlier) suggested so, these evidence were yet insufficient to conclude that visual working memory capacity estimates were not underrated due to encoding limitations. One very important issue
was never considered in them. Firstly, attention is crucial for encoding visual material. If it is not focused to a certain stimulus, then that stimulus has a very low probability of being encoded into memory (e.g. Coltheart, 1980; Irwin, 1991; Rensink, O'Regan, & Clark, 1997). Secondly, attention can be simultaneously directed to not more than 4 objects (e.g. Pylyshyn & Storm, 1988; Yantis, 1992), and since more than four objects cannot be memorized, that may either be a striking coincidence or it may mean that participants could not maintain more than four objects in their visual working memory simply because these objects were never encoded there, since they were presented simultaneously in all these experiments. The primary goal of the present study was to answer that question.

In order to do that, an experimental design was constructed, in which the type of initial presentation of stimuli was manipulated. Participants' performance in classical change detection tasks was compared to performance in modified change detection tasks with successive initial presentation of stimuli.
1.2 Method

Thirty-six subjects participated in a 2 x 2 mixed experiment, with the type of initial presentation of stimuli (simultaneous / successive) as a between subject factor and the type of test display (full / partial) as a within subject factor. Each participant went through 60 trials, in all of which squares, defined only by color, were used as stimuli.

All relevant factors besides the type of initial presentation and the type of test display were held constant:

a) set size was not varied – all trials consisted of 8 squares
b) total interval of initial presentation was always 500 ms
c) colors and locations of stimuli were selected randomly for each trial, but in order to assure that tasks are equally demanding in all conditions, the same 60 trials were used in every experimental condition (only their order of presentation was randomly varied across participants)
d) verbal load procedure was identical in all conditions

Therefore, it was ensured that the differences in participants' performance between experimental conditions can be attributed only to the effects of independent variables.

Each trial began with a presentation of a fixation mark, which was followed by a presentation of eight squares. The colors of these squares were chosen randomly among six different colors (yellow, green, red, blue, violet and black) for each trial. The location of stimuli was also randomly selected for every trial. They were presented in pseudorandom positions within a 3 (rows) x 5 (columns) grid. After initial presentation, 900-ms retention interval followed, during which a color change could occur on one of eight squares (the color of the changed stimuli was randomly chosen among five other colors). Changes were equally distributed across the display – they occurred the same number of times in every column in the grid. Half of all trials within every condition contained such change, and the other half did not. Squares could never change location, nor anything else but color. After the retention interval, a test display was presented, and participants’ task was to answer if a change had occurred or not.
Regarding the type of the initial presentation of stimuli, in the simultaneous presentation condition whole displays were presented at the same time, in duration of 500 ms, while in the successive presentation condition displays were revealed from the left to the right side, in five 100-ms presentational intervals. In the first of them only the first column was presented; in the second interval, the first and the second column were visible, and so on, till the last 100-ms interval, in which the whole picture was visible (same as the entire simultaneous presentation condition). Thus, total presentation time was identical in both conditions – it equalled 500 ms.

Regarding the type of test display, in the full test display condition all eight squares were presented at test, while in the partial test display condition only one colored square was presented in the test array (in 50% of the trials the color of that square was the same as the color of the square initially presented on that location). Other squares were replaced with placeholders – black outlined squares of the same size, which were placed exactly on the locations of initially presented squares.

In order to minimize the contribution of verbal working memory, besides short presentation of stimuli, verbal load task procedure was used: participants had to repeat the phrase *Coca-cola* aloud during the periods of initial display, memory delay and test display.

Method of the present study is illustrated in Figures 16 and 17.
1.3 Results and discussion

Participants’ performance was measured via the percentage of correct answers, via Pashler’s (1988) capacity estimates, and via reaction time. Mixed two-way ANOVA-s were conducted separately for each of these three dependent variables, with type of initial presentation as a between subject factor, and type of test display as a within subject factor.

Analyses have shown that the main effect of the type of initial presentation was not significant. The interaction between the type of initial presentation and the type of test display was also not significant. That finding does not corroborate the assumption according to which visual working memory capacity estimates, obtained in previous studies, were underrated due to encoding limitations. On the contrary, since successive initial presentation (in which it was ensured for every stimulus to undoubtedly be adequately encoded into visual working memory) resulted with no benefits in comparison to simultaneous initial presentation of stimuli, this result suggests that in classical change detection tasks visual material is adequately encoded, just as Vogel et al. (2001), together with other researchers have claimed.

The main effect of the type of test display was unclear. When measured via percentage of correct answers or via reaction time, it was not significant. However, when Pashler’s (1988) estimates were entered into ANOVA as a dependent measure of participants’ performance, the effect of test display type was significant: visual working memory capacity was higher in partial test display condition, compared to full test display condition. In order to solve this conflicting result, Pashler’s (1988) formula was thoroughly re-analysed through several simulations of participants’ attentional processes, storage of information, reasoning and decision making. For simplicity, simulations were made only for partial test display condition. These simulations were based on two assumptions:

a) set size equals eight items
b) visual working memory capacity equals four items.

In these simulations, the criterion of responding was varied through the instruction to participants:

a) in one simulation, participants were instructed to aim for high hit rate, regardless of false alarm rate – in order to avoid any misses they were told to
attend to and to memorize only two of eight presented items (they were instructed to answer "change present" in all trials in which they are unsure, also as an assurance to avoid any misses)

b) in another simulation, participants were instructed to aim for high false alarm rate, regardless of hit rate – in order to avoid any false alarms they were told to attend to and to memorize only two of eight presented items (they were instructed to answer "change absent" in all trials in which they are unsure, also as an assurance to avoid any false alarms)

c) in the last simulation, participants were given no instruction regarding the criterion of responding.

Rates of hits, misses, false alarms and correct rejections, together with the percentages of correct answers and Pashler’s (1988) capacity estimates were the output of the simulations, and these outputs suggested that Pashler’s (1988) procedure of visual working memory capacity estimation:

a) has a tendency to overestimate true visual working memory capacity

b) is dependent on the criterion of the responding.

If participants make a lot of hits and a lot of false alarms (lenient criterion: \( H = 1; FA = 0.75 \)), then memory capacity estimated by Pashler’s (1988) procedure is higher (\( C = 8 \) items), when compared to the situation in which participants make less hits and less false alarms (strict criterion: \( H = 0.25; FA = 0; C = 2 \) items), in spite of \( H – FA = 0.25 \), and in spite of the fact that in both of these simulations true visual working memory capacity was actually only 2 items. In the realistic simulation (with no instruction), it was shown that Pashler’s (1988) visual working memory capacity estimate equaled 5.33 items, while the true capacity was 4 items.

To see whether participants had changed their criterion of responding between full and partial test condition in the present study, hit rates and false alarm rates were entered into two separate mixed two-way ANOVA-s, with the type of initial presentation and the type of test display as factors. Analyses have shown that both hit rate and false alarm rate were significantly higher in partial test display condition, when compared to the full test display condition. Thus, participants did not use the same criterion of responding in these two conditions, and that is the reason why Pashler’s (1988) estimates of visual working memory were falsely higher in the
partial test display condition. True memory capacity did not differ as a function of the type of test display.

However, this solution of the cause of conflicting results regarding the effects of the test display type resulted with another enigma: why did participants use different criteria of responding in full and in partial test display condition? To answer that question, series of detailed simulations were carried out, in which all possible outcomes were considered separately for full and for partial test display condition (see Figure 23):

a) outcome 1: full test display + change absent + all items presented on test display
b) outcome 2: full test display + change present + one of the items that are held in memory has changed
c) outcome 3: full test display + change present + one of the items that are not held in memory has changed
d) outcome 4: partial test display + change absent + one of the items that are held in memory is presented on test display
e) outcome 5: partial test display + change absent + one of the items that are not held in memory is presented on test display
f) outcome 6: partial test display + change present + one of the items that are held in memory has changed
g) outcome 7: partial test display + change present + one of the items that are not held in memory has changed

The outputs of these simulations were then compared to the data observed in the present study, and it was demonstrated that the pattern of results is similar. In both, the output of simulation, and the results of the present study, it is found that:

1. percentages of correct answers are approximately the same for full and partial test display
2. $H - FA$ is approximately the same for full and partial test display condition, and equals around 0.50
3. Hit rate is higher in partial test display condition.
4. False alarm rate is higher in partial test display condition.
5. Visual working memory capacity, estimated by Pashler’s (1988) method, is higher in partial test display condition.
Therefore, through these simulations, it was explained why participants change their criterion of responding between different types of test display, although they hold 4 objects in both full and partial test display condition (one of the assumptions in all of these simulations is that four items can be stored and maintained in memory in both full and partial test display conditions), and although the percentage of correct answers and $H-FA$ are held constant between these conditions. Also, in these simulations it was also shown that Pashler’s (1988) formula tends to overestimate true visual working memory capacity.

Since it was discovered that Pashler’s (1988) method has several drawbacks, the focus of the rest of the present study was to find a better procedure for visual working memory capacity estimation. Apart from Pashler’s (1988) method being heavily affected by the criterion of responding, and apart from it having a tendency to overestimate visual working memory capacity, it was also shown that it sometimes gives senseless estimates. For example, if participants make at least one correct rejection, and answer "change present" in all other trials, then by his formula, visual working capacity would be estimated to equal set size.

A better procedure of visual working memory capacity assessment is based on the percentage of correct answers, rather than on hit and false alarm rate. Alvarez & Cavanagh (2004) used a classical change detection paradigm, and varied set size from 1 to 15 objects ($i = 2$) in their experiments. To estimate visual working memory capacity, they derived 75% threshold number for each subject, and divided it by two to obtain the capacity estimate. In explanation, they stated that if subjects stored half of all presented objects at, for example set size of 8, then they would obtain 75% of correct answers, because then they would detect 50% of all changes, and in all other trials they would answer "change absent". Thus, participants would obtain 100% of the correct answers in all trials which do not contain a change, and 50% of the correct answers in all trials that do contain a change. However, this procedure also has several drawbacks:

1. since it was created especially for that specific experiment it requires complex experimental designs, and cannot be applied to experiments in which set size is not varied
2. it disregards guessing
3. this procedure in its original form can not be applied to experiments using partial test display tasks, because the essential logic of decision making is completely different. (e.g. In partial test display trials, participants would not answer "change absent" in all the trials that do not contain a change, nor "change present" in 50% of all the trials that do contain a change.)

However, it is possible to derive another procedure, different then Alvarez & Cavanagh’s (2004), from the percentage of correct answers. This procedure was constructed in the present study, on the basis of the final set of simulations, in which the number of items that are stored and held in memory was varied as input. The outputs of the simulations included the percentages of correct answers, Pashler’s (1988) estimates, and rates of hits, false alarms, misses and correct rejections. The simulations indicated that Pashler’s (1988) formula tends to overestimate visual working memory capacity, and more importantly, it was found that the relationship between percentage of correct answers and visual working memory capacity is linear. Accordingly, a new method of assessing visual working memory capacity was constructed:

\[ C_w = \frac{IP^* (PC - 50)}{50} \]

where:

- \( C_w \) is visual working memory capacity estimated via percentage of correct answers
- \( PC \) is percentage of correct answers
- \( IP \) is set size (number of initially presented items).

It was theoretically demonstrated that this formula is valid for different set sizes, and for both (full and partial) types of test display.

When Pashler's (1988), Alvarez & Cavanagh (2004) and this procedure are compared, it is clear that this new procedure is better then the other two:

1. while Pashler's method overestimates true visual working memory capacity, this method does not
2. while Alvarez & Cavanagh's method requires complex experimental designs (that include varying of set size) this method can be applied in
more economical experiments, but it can still be applied in experiments in which set size is varied (in these experiments, the mean capacity for different set sizes can be calculated, and that would be more reliable capacity estimate).

3. while Pashler's formula sometimes gives senseless estimates, this formula does not

4. while Pashler's procedure is heavily affected by the criterion of responding, this procedure is not

5. while Alvarez & Cavanagh's procedure is not valid for both partial and full test display condition, this procedure is.

When the data from the present study are entered into the new formula, then it is estimated that visual working memory equaled between 3.6 to 4.3 items, depending on experimental condition. When the same data are entered into Pashler's (1988) formula, then it estimated visual working memory capacity to equal between 4.6 to 5.9 items, depending on experimental condition, but as it was mentioned earlier, Pashler's (1988) formula overrates memory capacity. Thus, it can be concluded that it was found that visual working memory capacity observed in the present study equals round 4 items.

The new visual working memory assessment procedure is valid for stimuli defined by one feature, while for multi-feature stimuli the situation is more complex. However, since the research of visual working memory capacity limitations is extremely important for the understanding of higher cognitive functions, the construction of new assessment procedure that can be applied for multifeature objects can soon be expected.
1.4 Conclusion

Successive presentation of stimuli had no positive effects on participants' performance, when compared to simultaneous presentation of stimuli. Therefore, in classical change detection tasks, visual working memory capacity is not underestimated due to an inadequate encoding of stimuli.

The effect of the test display type was doubtful. When it was measured by the percentage of correct answers or by reaction time, it was not significant, but when Pashler's (1988) capacity estimates were analysed as a dependent measure, then it was found that compared to full test display condition, in partial test display condition participants had memorized significantly larger number of items. It was discovered that the change in criterion of responding is the cause of these noncongruent results. Specifically, it was demonstrated that Pashler's (1988) procedure of visual working memory assessment is heavily affected by the criterion of responding. Since hit rates and false alarm rates observed in the present study, were significantly higher in partial test display condition, Pashler's (1988) visual working memory capacity estimates were also higher in that condition, although true capacities were approximately the same in both, full and partial test display condition.

It was demonstrated that Pashler’s (1988) method, besides being heavily affected by the criterion of responding, tends to overrate true memory capacity, and has some other drawbacks.

Finally, the new method of visual working memory capacity assessment was constructed. That method estimates the memory capacity from the percentage of correct answers, rather than from hit rate and false alarm rate:

\[ C_\% = \frac{IP^*(PC - 50)}{50} \]

When the data from the present study were entered into that formula, it was found that visual working memory capacity observed in the present study equals circa 4 items.

The new method was compared to other common procedures of visual working memory assessment, and it was demonstrated to be better than the others,
because it does not overestimate the true capacity, is not affected by the criterion of responding and can be used in various experimental designs.
2 INTRODUCTION

2.1 Definition of the working memory

Memory can be defined as ability to encode, store, retain and retrieve information. Based on duration, it can be divided into sensory, short-term (or working) and long-term memory.

Working memory can further be divided into three components (Figure 1): the central executive system, and two slave systems - an auditory short-term store, and a visual short-term store (Baddeley & Hitch, 1974; Baddeley, 1986, 1992, 1996).

![Figure 1. Components of working memory. Adapted from "Working memory: Looking back and looking forward." by A.D. Baddeley, 2003, Nature Reviews Neuroscience, 4, p. 829.](image)

The central executive is the most important, but least understood component of visual working memory (Baddeley, 2003). Originally, it was considered to be responsible for coordinating its slave systems (Baddeley & Hitch, 1974). Also, central executive is believed to be in charge of directing attention to relevant information, for suppressing irrelevant information and improper actions, and for the coordination of cognitive behavior in situations when several tasks have to be done simultaneously (Baddeley & Hitch, 1974; Baddeley, 1986, 1992, 1996). Baddeley (2000) also believes that the central executive system possesses one more component – the
episodic buffer\(^1\), which holds information that are not under the authority of slave systems (e.g. musical information), and representations that integrate phonological, visual and spatial representations.

Auditory short-term store is also known as phonological loop, whereas visual short-term store is usually called visuo-spatial sketchpad. Phonological loop prevents the decay of verbal information by refreshing it, or in other words, by continuously articulating its contents. Similarly, the function of visuo-spatial sketchpad, which can further be divided into a visual subsystem and a spatial subsystem (Baddeley & Hitch, 1974; Baddeley, 1986, 1992), is to hold visual and visuospatial information while they are manipulated by the central executive system.

\section*{2.2 Development of visual working memory capacity research}

The capacity of visual working memory is yet not well investigated nor comprehended, especially when compared to the capacity of verbal working memory. For example, while the history of verbal working memory research reaches to 19th century, to the era of Ebbinghause and his famous experiments, one of the first considerable researchers of visual working memory was George Sperling. Therefore, it would not be wrong to say that the visual working memory research began with Sperling, about only 50 years ago. For illustration, by that time Miller (1956) had already discovered that verbal working memory capacity equals approximately seven items. There is another striking indicator showing that the capacity of visual working memory is not nearly as well understood as the capacity of verbal working memory: Sperling's (1960) methodology and inferences about visual working memory capacity were disapproved soon, while the majority of Ebbinghause's findings are now valid for almost a century and a half, and have been corroborated over and over again.

\footnote{\textsuperscript{1} It is important to distinguish Baddeley's (2000) concept of episodic buffer from the classical comprehension of episodic memory: Baddeley's (2000) episodic buffer is a temporary store.}
Similarly, Miller's (1956) conclusions have still not been disproven. Sperling's (1960) research, along with mistakes in his reasoning will be shown in the next paragraph.

**2.2.1 The embarcation – Sperling's work**

In Sperling's experiments, participants viewed displays composed of 3 to 16 alphanumeric characters. In whole report condition, participant's task was to recall as many presented objects as possible, whereas in partial report condition they were required to report only a subset of characters within the visual display. By the frequency of presented tone, experimenter indicated to observers which set of characters he wanted them to report (see Figure 2).

In whole report condition, participants were unable to recall more then four to five characters, regardless of the number of letters presented. This result suggested that visual memory capacity is limited to 4-5 objects. However, partial report condition revealed that participants had much better memory. For example, after they had viewed a matrix of 3 lines of four letters each, they could correctly recall three out of four characters. (In partial report condition, participants were unable to foretell what line they would be asked for to report, and that must be taken into consideration).

![Figure 2. One of Sperling's (1960) experiments. For example, if participants hear a low frequency tone, that means that they have to report characters from bottom row.](image-url)
Sperling (1960) also varied the retention interval, and discovered that the increment of interval between the offset of the display and the auditory cue, greatly impairs performance in the partial report condition. When that interval had reached 1000 ms, performance in the partial report condition was approximately the same as in the whole report condition. These findings led Sperling to conclude that sensory trace remains accessible for a very short time after the presentation. This type of store is now called iconic memory.

Anyhow, Sperling's (1960) process of assessing visual memory capacity had several serious irregularities. Primary, Sperling (1960) used letters and numbers as stimuli, and these items were probably coded not only visually, but also verbally. Beside that, participants had to respond by writing down the characters they had just memorized, and that kind of answering surely forced them to verbally code the stimuli. Verbal contribution has very probably falsely increased visual memory capacity estimate. Suprisingly, regardless of that distortion, Sperling (1960) concluded that observers could remember 4-5 visual items, which (as will be shown later) happens to be the correct estimate of visual memory span.

A different, better method for assessing the capacity of visual working memory was developed by Phillips (1974). Unlike Sperling's (1960), Phillips's paradigm did not rely on verbal report – he used unfamiliar visual stimuli.

### 2.2.2 Methodological improvement: Phillips's contribution

Phillips (1974) was one of the first scientists who investigated visual working memory by applying change detection paradigm. He had successfully distinguished visual from verbal working memory, and also managed to accurately and precisely separate visual working memory from iconic memory.

Phillips (1974) conducted several experiments in which he had tachistoscopically presented displays filled with dots. Trials consisted of two displays containing dots, separated by a variable inter-stimulus interval. Displays within a trial were either identical, or differed in the addition or the removal of one dot. Participants' task was to detect occurrence of change in displays. Their performance was excellent with short inter-stimuli intervals (approximately 100 ms or shorter), but
prolongation of retention interval resulted in decreased change detection accuracy. Another important finding had also emerged: inserting a mask between two displays caused impairment in subjects' performance only in conditions with short inter-stimuli intervals, while with longer inter-stimuli intervals masking had practically no effect. Accordingly, Phillips (1974) concluded that excellent performance with short retention intervals can be attributed to iconic memory, while visual working memory is responsible for longer retention interval performance. These experiments have brought detailed insight into the nature of iconic and working memory, and clarified boundaries between them, and boundaries between visual and verbal working memory also.

Phillips (1974) concluded that iconic memory has vast capacity, but very short duration, in contrast to visual working memory, which is capable to store only small amount of information, but can maintain them for a longer period of time. These conclusions were precious to Pashler, who made important contribution in the field of visual working memory-span research.

### 2.2.3 Quantitative approach by Pashler

In order to illustrate and clarify Pashler's (1988) contribution, Purdy et al.'s experimentation must be described first.

#### 2.2.3.1 Purdy et al.'s research

Purdy, Eimann, & Cross (1980) explored change detection performance by presenting pairs of displays containing 16 letters, with 100 ms-lasting initial display presentation. Displays were separated by variable inter-stimulus interval, ranging between 100 and 5000 ms. In one part of all trials, change could occur on 1, 4 or 16 positions during the inter-stimulus interval, while the rest of the trials did not contain a change between displays. Results were astonishing: when all 16 letters were changed, the prolongation of inter-stimulus interval did not result with impairment of participants' response accuracy – their performance was almost faultless.
At first sight, it seemed that Purdy et al. (1980) had disproven Phillips's (1974) findings. With regard to the results mentioned, they concluded that sensory memory is long-lasting. However, Pashler (1988) questioned and disputed that proposal.

### 2.2.3.2 Pashler's formula

Pashler (1988) stated that the results of Purdy et al. (1980) were not as fascinating as they seemed. He pointed out a cardinal mistake in their reasoning: their finding that participants can easily solve change detection tasks when all 16 positions had changed, did not justify the conclusion that the positions of all items were stored in memory during the retention interval. In that experimental condition, it was enough to pay attention to only one position and to check whether it had changed or not. Thus, participants only had to store one position in memory in order to give correct answer.

Pashler (1988) suggested that condition in which only one character changes position during the retention interval, would be much better indicator of memory capacity. He elaborated this idea and developed a new method of estimation of visual memory span. This method was frequently used, and is now generally accepted as the best way to calculate the capacity of visual working memory.

<table>
<thead>
<tr>
<th>Change</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Absent”</td>
<td>“Present”</td>
</tr>
<tr>
<td>Change present</td>
<td>Miss</td>
</tr>
<tr>
<td>Change absent</td>
<td>Correct rejection</td>
</tr>
</tbody>
</table>

Figure 3. *Four possible outcomes in a change detection task.*
Pashler (1988) had reexamined the experimentation of Purdy et al. (1980). Specifically, he took the condition in which only one character was changed between two displays. In that condition Purdy et al.’s (1980) participants correctly detected 42% of the changes, and at the same time, in 15% of trials that did not contain a change, they responded inaccurately. Pashler’s formula for memory capacity estimation derives from signal detection theory, so speaking in that terminology (look at Figure 3), Purdy et al.’s (1980) participants had 0.42 hit rate, and 0.15 false alarm rate in related experimental condition.

Presuming that participants held a certain amount of items in memory ($C$), when one of them changed, hit would follow, and if the changed item was not stored in memory, miss would occur. When participants did not detect a change, they would answer "change absent", but it must be taken into account that in a certain proportion of the trials they were guessing. Guessing relates to answering "change present" in the cases where there was actually no change (false alarms). Along with the variables mentioned and set size, Pashler (1988) gave the formula for determining visual working memory capacity:

\[
H = \frac{C}{IP} + \frac{IP - C}{IP} \cdot FA
\]

where: $C$ is number of objects stored in memory (visual working memory capacity);
$H$ is hit rate;
$IP$ is total number of items presented in a display;
$FA$ is false alarm rate.

Viewed from another angle:

$C$ is number of presented objects that are stored in memory (visual working memory capacity);

$\left(\frac{C}{IP}\right)$ is proportion of trials in which an item that is stored in memory changes

$(IP - C)$ is number of presented objects that are not stored in memory

$\left(\frac{IP - C}{IP}\right)$ is proportion of trials in which an item that is not stored in memory changes
FA is guessing rate (guessing rate is estimated from the false alarm rate – from the proportion of no-change trials on which the subject guesses there was a change).

Thus, this model assumes that the probability of a hit (H) equals the proportion of correct detections of a change in cases when changed item was stored in working memory (C/IP) added up with guessing (guessing here refers to situations in trials containing a change, in which participants did not store the changing item to memory, but regardless to that correctly responded "change occurred").

Formula for visual working memory capacity can be easily extracted from the equation above:

$$H = \frac{C}{IP} + \frac{IP - C}{IP} \times FA$$

$$H \times IP = C + (IP - C) \times FA$$

$$H \times IP = C + IP \times FA - C \times FA$$

$$C - C \times FA = H \times IP - IP \times FA$$

$$C \times (1 - FA) = H \times IP - IP \times FA$$

$$C = \frac{H \times IP - IP \times FA}{1 - FA}$$

$$C = \frac{IP \times (H - FA)}{1 - FA}$$

When data from Purdy et al.'s (1980) study are inserted into this formula, then visual working memory capacity of their participants equals:
Although Pashler (1988) himself admits that this model is simple and crude (for example, it attributes all errors to the maintenance of representations, rather than the comparison process), it is probably the best method of visual working memory capacity estimation. This model was later used by Vogel et al. in their detailed experimentation.

### 2.2.4 Beginnings of modern research: Vogel et al.'s experiments

Luck & Vogel (1997) estimated visual working memory capacity using change detection paradigm, similar as Phillips (1974). Their goal was to measure visual working memory for simple features and for conjunctions. To estimate capacity for simple features, they flashed arrays of 1, 2, 3, 4, 8 or 12 colored squares (see Figure U3) to participants for 100 ms. Each trial contained two displays of stimuli, separated by 900-ms blank interval, with a restriction that the number and location of stimuli could never change within a trial. Two arrays within each trial could either be the same (in 50% of all trials), or could differ in color of only one square (in the other 50% of trials).

Participants' task was to detect this kind of changes, and they managed to perform almost faultlessly for displays containing three or less squares, but their accuracy started to decline systematically at set size of four items. In their study, Luck & Vogel (1997) conducted several more experiments using similar methodology, some of which with the use of a different kind of stimuli. According to Pashler's (1988) method they calculated that observers' visual working memory capacity equaled round four objects.

A few years later, Vogel and his associates published a more extensive article on the same topic (Vogel, Woodman, & Luck, 2001). They obtained similar results as Luck & Vogel (1997). In both these reports, Vogel and his colleagues stated that they had used colored squares as stimuli (among other), but in spite of that, minimization
of the contribution of verbal working memory was ensured in their experiments. They used a verbal load task and short presentation intervals. Possible approaches of reducing the contribution of verbal working memory are discussed in the next section.

Figure 4. Example of displays used in one of Luck & Vogel’s (1997) and Vogel et al.'s (2001) experiments.

2.3 Reducing the contribution of verbal working memory

There are three main methods of minimizing the contribution of verbal working memory:

a) use of stimuli that are difficult to name

Non-verbal stimuli have no corresponding auditory code, and because of that they are adequate for the research of pure visual memory. However, this kind
of stimuli are rare, so it is difficult to find stimuli that have no chance to be verbally coded by the participants. The best examples of adequate non-verbal stimuli are Phillips's (1974) dots and random polygons (Figure 5) used by Alvarez & Cavanagh (2004). Wheeler & Treisman's (2002) colored squares (Figure 6) are also rather adequate. Except that it is difficult to construct proper kind of stimuli, this approach has one more drawback – it limits the variety of experimental designs.

Figure 5. *An example of non-verbal stimuli – random polygons used by Alvarez & Cavanagh (2004).*

Figure 6. *An example of Wheeler & Treisman's non-verbal stimuli: colored squares.*

b) *brief presentation of stimuli*
This approach aims to deprive the time required for phonological coding. Phonological coding is a slower process then visual coding, and therefore, tachistoscopic presentation of stimuli reduces the possibility of verbal coding and consequently minimizes the contribution of verbal memory (see Frick, 1988).
c) verbal load

Verbal load refers primarily to meaningless repetitive speech that prevents auditory coding (Murray, 1967). This procedure is also known as auditory or articulatory suppression, and it does not impair performance on tasks that require visual coding. Besides disrupting auditory coding, auditory suppression has almost no other effects (Klapp, Marshburn, & Lester, 1983; Richardson, 1984; Baddeley, 1986). Also, this method allows testing visual working memory by longer display durations (Frick, 1985). Besides articulatory suppression, verbal loads can consist of shadowing procedures, as well as having participants remember a verbal item throughout a trial (Vogel et al. 2001).

2.4 Generality of the visual working memory capacity estimate

Using a similar method as Luck & Vogel (1997), Vogel et al. (2001) demonstrated that it is possible to retain information about only 3-4 objects. They obtained that result with several experiments, in which colored squares were used as stimuli. Also, retention interval was 900 ms in all of these experiments. Therefore, Vogel et al. (2001) decided to investigate whether the results acquired by their initial experiments can be generalized to longer delay intervals and to different kinds of stimuli. To answer that question, they conducted two experiments.

In the first of them, they varied the retention interval. Besides 900 ms interval, they applied 2900 ms and 4900 ms intervals. Also, they manipulated set size. Everything else was identical to their initial experiments. Results have shown that accuracy had declined with the increase of set size, yielding a significant effect of set size, while there was no significant effect of retention interval, and no significant interaction between retention interval and set size. There was no decay of the memory representation (Vogel et al., 2001).

After that experiment, Vogel et al. (2001) conducted another one, in which memory for color was compared to memory for orientation. To maximize the
comparability of these two experimental situations, the same stimuli were used in both conditions (Figure 7). Presented stimuli were defined by color (blue, green, red or white) and by orientation (0º, 45º, 90º and 135º). In one condition, only color of one item could have changed during the retention interval, and in the other condition, only orientation of one item could have changed. In both conditions, set size equaled 4 or 8 items. Again, only the effect of set size was significant. Memory for color did not differ from memory for orientation.

Therefore, the finding from this experiment suggests that visual working memory capacity estimate does not reflect specific memory limitation for color. It highly probably reflects general capacity limitation for simple features.

Figure 7. Oriented colored lines. (In Vogel et al.'s (2001) paper, stimuli used in control experiment concerning the generality of the capacity estimate were not illustrated. Therefore, this picture is taken from Luck & Vogel's article (1997) who used similar stimuli.)

2.5 Visual working memory for multifeature objects

2.5.1 Vogel et al.'s theory

Vogel et al.'s (2001) previous experiment inspired them to take one step further, and to raise another question – a question concerning multifeature objects: Do they consume more visual working memory capacity than single-feature objects?

As verbal memory capacity depends on the complexity of verbal material (e.g. Baddeley, Thomson & Buchanan (1975) and Schweickert & Boruff (1986) discovered that long words consume more memory capacity than short words), visual memory capacity might similarly depend on the complexity of visual material (e.g. number of elementary features that define visual objects). Therefore, it is not unreasonable to assume the following: in comparison to tasks which require memorizing objects defined by several features, it is probably possible to retain more objects when objects to be remembered are defined only by a single feature. Therefore, Vogel et al. (2001) poisted the following problem: Does the storage of multifeature objects occupy more visual working memory capacity than the storage of single-feature objects. Reformulated, they wanted to assess whether visual information is stored as integrated objects or as individual features.

To answer that question, Vogel et al. (2001) used the same stimuli as in their previous experiment – bars defined by color and orientation. They compared memory for simple features to memory for the same objects defined by a conjunction of features. So, in one condition, participants had to memorize only color of bars; in another condition, they had to memorize only orientation; and in the last (conjunction) condition, they were required to memorize both, color and orientation. In the conjunction condition, either the color or the orientation of only one item could change, while in other two conditions, only color or only orientation of one item could alter. Therefore, in the conjunction condition, in sets of four objects, participants had to memorize eight features, while in the other two conditions, they had to store only four features. Vogel et al. (2001) found no difference between these three conditions. This result indicates that objects are stored to visual working memory as integrated units, similarly as working memory stores information as chunks. Accordingly, visual
working memory capacity is not limited by the number of features, but by the number of objects. This conclusion was consistent with the findings from visual attention studies, which indicated that attention is directed to entire objects rather than to individual features (e.g. Duncan, 1984).

Yet, some doubts remained. Results of the experiment described can however be alternatively explained. Specifically, it is possible that the participants were unable to selectively attend only to color in the condition which required them to store only color (or to orientation in the other condition). Consequently, participants might have encoded irrelevant features, although it was not required to do so. If that was the case, that would mean that the observed performance in two single-feature conditions was underestimated due to methodological imperfection. To examine that, Vogel et al. (2001) conducted another experiment.

In that control experiment, authors wanted to see what is going to happen when they eliminate information about irrelevant features. To answer that question, Vogel et al. (2001) had to modify stimuli. In order to eliminate information about orientation, they used colored vertical bars, and in the orientation condition, they used black bars which could differ only in their orientation. The conjunction condition was identical as the related condition in the previously described experiment. Analysis have shown no differences between these three conditions. In the conjunction condition, participants were able to store information about both color and orientation equally good as information about only color or only orientation. Since the elimination of irrelevant feature did not lead to improved performance, this experiment corroborated Vogel et al.'s (2001) hypothesis about storage of objects as integrated units.

In one of their experiments, Vogel et al. (2001) demonstrated that participants are able to store information about 16 features, distributed across four objects. They used objects defined by a conjunction of four features as stimuli. Objects varied in color (red/green), size (big/small), orientation (horizontal/vertical), and the presence or absence of a black gap (Figure 8). In the conjunction condition, any feature of any object could have changed, while in the other four conditions, participants were told that only one feature could change. So, in these four conditions, unlike the conjunction condition, participants knew exactly which features they have to attend to and memorize. Regardless of that, their performance in the quadruple conjunction
condition was as good as it was in the individual feature conditions. At a set size of four objects, participants managed to retain 16 features distributed across 4 objects equally well as 4 features distributed across 4 objects.

Figure 8. *Stimuli by which Luck & Vogel (1997) managed to demonstrate that participants can store 16 features into visual working memory. Vogel et al. (2001) used similar stimuli for the same purpose. (In Vogel et al.'s (2001) paper, stimuli used in related experiment were not illustrated. Therefore, this picture is taken from Luck & Vogel (1997) who used similar stimuli.)* Adapted from “The capacity of visual working memory for features and conjunctions,” by S.J. Luck and E.K. Vogel, 1997, *Nature, 390*, p. 280.

2.5.1.1 Color-color conjunction

Within the same study, Vogel et al. (2001) took another step further. They conducted an experiment in which objects were defined with two features within the same dimension. Their goal was to reexamine the finding that objects are stored as integrated units. Particularly, excellent performance in conjunction conditions from
previous experiments could be plausibly explained by the existence of parallel independent memory systems for different features. To rule out that possibility Vogel et al. (2001) used bicolored squares as stimuli (Figure 9).


That experiment consisted of three conditions. In the first of them, Vogel et al. (2001) presented only small squares, and in the second condition, only large squares were presented (without gap in the middle). In the third condition, they presented small squares, which were positioned within the big ones, as illustrated in Figure 9. 50% of all trials did not contain a change. Regarding trials from third condition that contained a change, in half of them, only inner square could change color, and in the other half of change trials, only outer square could alter its color. Statistical analysis showed that participants had memorized items equally good in all three experimental
conditions. In the conjunction condition, they were able to memorize twice as many colors than in the other two conditions. This result corroborated the hypothesis that objects are stored into visual working memory as integrated units.

Decrement in performance for conjunction condition was not observed, and that finding is not consistent with the hypothesis of separate memory systems for different features. It might still be hypothesized that independent memory systems exist for inner and outer squares, but that is simply not plausible.

Although this experiment could justifiably be criticized because conjunction condition was perceptually more demanding then the other two condition, that could only distort results in the opposite direction. In other words, the elimination of that kind of bias could result with the increase of performance in the conjunction condition (Vogel et al., 2001), and that would be an even stronger corroboration of the hypothesis that objects are stored into visual working memory as integrated units, and not in parallel stores.

![Diagram](image)

**Figure 10.** Vogel et al.'s control experiment in which inner squares were shifted diagonally.

As a control of this experiment, Vogel et al. (2001) shifted inner squares diagonally (see Figure 10). After that manipulation, pairs of features were not integrated into single objects any more, but the overall number of features remained unchanged. Diagonal shifting of inner squares impaired participants' performance – they had difficulties with the storage of that kind of stimuli, and that finding also supported Vogel et al.'s (2001) hypothesis of integrated-object storage.

2.5.2 Wheeler & Treisman's theory

Wheeler & Treisman (2002) disagreed with Vogel et al.'s (2001) theory that visual memory stores integrated objects rather then individual features. Specifically, one of Treisman's studies had shown that representations of integrated objects tend to break up into their separate features in memory (Treisman, Sykes, & Gelade, 1977). In that study, Treisman et al. (1977) presented arrays of only two complex stimuli. After short retention interval, second array was presented, which could be the same, or could differ from the initial array. The analysis of participants' errors had shown that the majority of errors occurred in cases when two features switched places. In other words, participants were unable to detect changes on trials in which features changed only their positions, but not their values. For illustration, in one experimental condition, faces with different noses, eyes and mouths were used as stimuli. Initial display could for example consist of a smiling face with round eyes and a frowning face with square eyes. If that display was followed by a presentation of a frowning face with round eyes and a smiling face with square eyes, then majority of participants were incapable of detecting that change. Thus, Treisman et al.'s (1977) participants were unable to store two objects defined by three features each.

Several other authors had also found that features can not remain bound in visual working memory, without extra effort (e.g. Horowitz & Wolfe, 1998; Wolfe, 1999; Rensink, 2000).

Considering all these findings, Wheeler & Treisman (2002) were sceptical towards Vogel et al.'s (2001) theory, according to which, objects are stored into memory as integrated units. In contrast, Wheeler & Treisman (2002) believed that individual features are stored in paralell feature stores, each of which possesses their own independent capacity. Within feature stores, values compete for limited capacity
representation, but between feature stores, there is no competition. According to Wheeler & Treisman's (2002) model, binding information can also be maintained in memory, but they are vulnerable to oscillations in limited attentional resources. Anything else that competes with attentional resources interferes with retention of information about bindings.


They tried to replicate crucial findings from Vogel et al.'s (2001) color-color conjunction experiment, but contradictory results were obtained. Even when Wheeler & Treisman (2002) used procedure practically identical to Vogel et al.'s (2001), they found no evidence for binding between the colored parts of squares. Performance was worse for bicolored squares (conjunction condition) than for single-colored squares. Wheeler & Treisman (2002) stated that they cannot explain the reason for the discrepancy between their and Vogel et al.'s (2001) findings.

### 2.5.3 Resolution of "Vogel et al. versus Wheeler & Treisman" debate

It seems that Wheeler & Treisman (2002) were right: although some studies (e.g. Ceraso, 1985; Irwin, 1992) suggest that objects are stored into memory as integrated units, majority of recent studies support Wheeler & Treisman's (2002) theory. In this paragraph, three prominent studies, the aim of which was the resolution of the debate between Vogel et al. (2001) and Wheeler & Treisman (2002), will be shown. In all three of them, investigators had independently come to similar conclusions, using diverse experimental approaches.

### 2.5.3.1 Wheeler & Treisman's research

After failing to directly replicate Vogel et al.'s (2001) results, Wheeler & Treisman (2002) designed another experiment to investigate the nature of information...
storage. The experiment was composed of four conditions. In all of them, objects (single-colored squares) were initially presented in duration of 150 ms, and all inter-stimuli intervals lasted 900 ms. Half of all trials in all four conditions contained no change during the retention interval. In the rest of the trials, the type of change depended on experimental condition. In the first condition, only the color of two squares could change. In the second condition, only the location of two squares could change. In the third condition, either the color of two squares, or the location of two squares could change (naturally, serial order of trials was randomly intermixed), and in the last condition (binding condition), all colors presented, and all locations occupied remained the same, but colored squares could swap location. Participants were informed about all these possibilities. These four conditions were blocked, and each block of trials was presented in counterbalanced order. Design of this experiment is illustrated in Figure 11.


Analysis indicated that participants' performance was the worst in the condition in which colors swapped their locations, and therefore, it is certain that information about
color and location were not automatically bound together - they were stored into separate parallell stores. This result corroborates Wheeler & Treisman's (2002) theory and is not consistent with Vogel et al.'s (2001) model of visual working memory. In other three conditions, participants performed equally adequate, so accordingly, memory for location did not interfere with memory for color (otherwise, in the condition in which either color or location could change, performance would be worse). Although this finding is consistent with Vogel et al.'s (2001) theory, it is also consistent with Wheeler & Treisman's (2002) theory of parallell storage of features.

Another experiment of Wheeler & Treisman (2002) was the same as the previous one, with one exception: at test, only one square was presented. Conditions were equivalent with conditions of previously described experiment.

- In the first condition, in which only color could change, test square was presented in the middle of test display (neutral location). In half of the trials, color of that square was the same as the color of one of the initially displayed squares, and in the other half of trials, the square in the center was of a color not previously presented in the initial display.

- In the second condition, in which only location could change, black square (neutral color) was presented in a location previously occupied in initial display (in half of trials), or in a previously unoccupied location (in other half of the trials).

- The third condition was again a combination of trials from the first two conditions. In half of the trials, memory for location was tested, while the rest of the trials tested memory for color. The serial order of the trials was randomly intermixed, so the participants did not know what feature would be tested.

- In the last condition, a colored square was presented at a location that had previously been occupied, with the restriction that the color of that square had been present in the initial display. Participants had to answer whether that colored square appeared in this particular location in the initial display, or in some other location.

The results obtained were compared to the results from the previous experiment, in which memory was tested with whole display-tests. The pattern of results for the first three conditions was similar, but the accuracy of responses in the last (binding)
condition was higher in this experiment, in which partial test-displays were used. The results obtained suggest that binding information is accurately retained when memory is tested with partial test-displays. Therefore, from these two experiments, it can be concluded that Wheeler & Treisman's (2002) theory is valid: visual information are stored into parallel feature stores, but also, another mechanism for binding information exists. That mechanism relies on limited attentional resources, and because of that, binding information were lost in the whole display-test conditions. In these situations, attentional resources were redirected to perceive objects that could not have changed. When single-feature stimuli were used, performance was not impaired in the whole display-tests, and that is also consistent with Wheeler & Treisman's (2002) theory, because, additional attentional resources are required only to retain binding information\(^2\).

In one similar control experiment, Wheeler & Treisman (2002) used multifeature objects defined by color and shape, and same results were obtained. Testing condition had no effect when memory for single-feature objects (color only, or location only) was analysed, and memory for multifeature objects was again worse when it was tested by whole displays.

### 2.5.3.2 Research of Alvarez & Cavanagh

Besides Wheeler & Treisman (2002), Alvarez & Cavanagh (2004) also reexamined Vogel et al.'s (2001) model. Specifically, they checked Vogel et al.'s (2001) hypothesis that the total number of features that define objects to be remembered, is not the factor which determines visual working memory capacity for multifeature objects.

Alvarez & Cavanagh (2004) tested whether visual working memory capacity is limited by the number of objects that need to be stored, or by the total amount of information that needs to be memorized. For that purpose, five classes of stimuli were used: colors, polygons, Chinese characters, shaded cubes and letters (Figure U14).

\(^2\) Wolfe (1999) also found that attention is necessary not only to encode, but also to retain integrated objects in visual working memory.
The amount of visual information was assessed via measuring processing rate in visual search tasks, for different kinds of stimuli. Authors also wanted to investigate the relationship between visual search rate and memory capacity.

Figure 12. *Stimuli used in Alvarez & Cavanagh’s (2004) study.*


The results obtained indicated that in terms of the number of objects, visual working memory capacity differs across different categories of stimuli – it varies from 1-2 objects for shaded cubes to 4-5 objects for colored squares. This finding directly contradicts Vogel et al.’s (2001) model of visual working memory, which proposes that its capacity is fixed in terms of the number of objects. Alvaraz & Cavanagh (2004) obtained another crucial result: they found inverse relation between the information load per object and the number of objects that can be stored into visual working memory. In other words, they found that visual working memory capacity is limited by the total amount of information (Alvarez & Cavanagh, 2004). This kind of
a trade off between the complexity of objects and a total number of objects that can be stored in memory, is also contrary with Vogel et al.'s (2001) theory.

2.5.3.3 Švegar & Domijan's experiments

The study of Švegar & Domijan's (2007) had also greatly contributed to the resolution of the debate between Vogel et al. (2001) and Wheeler & Treisman (2002). Although methodological differences between Vogel et al.'s (2001) original color-color conjunction experiment and Wheeler & Treisman's (2002) replication were minimal and probably irrelevant, they were still present. After thorough overiew of both studies, Švegar & Domijan (2007) concluded that the main methodological difference between Vogel et al.'s (2001) original experiment and the replication of Wheeler & Treisman (2002), was in the duration of initial displays. While Vogel et al. (2001) used 100-ms, Wheeler & Treisman used 150-ms initial displays. Therefore, in one of their exsperiments, Švegar & Domijan (2007) examined whether this methodological difference could have caused contradictory results. In that experiment, lines defined by length (long/short), orientation (0º, 45º, 90º and 135º), and color (red, green, blue), were used as stimuli. Pairs of displays containing four objects, divided by 900-ms retention interval, were presented to participants. Retention interval, and the number of changing features were experimentally manipulated, with a restriction that during the retention interval, only one or no objects could have changed. However, the cause of conflicting results remained unknown after that experiment: variations in duration of initial display had no effects on participants' performance. Only the effect of the number of changing features was significant: participants’ performance increased systematically with the increase of number of changing features.

Švegar & Domijan (2007) then took one step further, and constructed another experiment in order to directly test Vogel et al’s (2001) versus Wheeler & Treisman’s (2002) theory. As in the previous experiment, the same multi-feature objects were used again. This experiment consisted of two conditions. In both of them, in all trials only two or none features could have changed during the retention interval. In the first condition, two features of only one object could have changed, and in the second
condition, change was distributed over two objects (Figure 13). If objects are stored into memory as integrated units, as Vogel et al. (2001) had suggested, then participants should have made less errors in the condition when two objects had changed. Analysis showed no differences in performance between these two conditions, and that result suggested that features are separately stored into visual working memory, in conformity with Wheeler & Treisman’s (2002) theory. Also, this result corroborates the model of Alvarez & Cavanagh (2004), according to which, visual working memory capacity is limited by the total amount of information, rather than by the total number of objects.

Figure 13. Illustration of Švegar & Domijan's (2007) study. In Condition 1, one object changed its color, and the other one changed its length. In Condition 2, both feature changes (length and color) occurred on the same object.

\[3\] Interestingly, performance in these two conditions was practically identical.
2.6 Is visual working memory capacity indeed so limited?

2.6.1 Accumulated decision errors

Although there was not a single evidence that it is possible to retain information about more then 4-5 objects (as can be seen from all studies described so far), it would however be very doubtful to claim that visual working memory capacity is so limited. To conclude that, the possibility that observed visual working memory capacity was underestimated due to the accumulation of errors at a decision stage, must be ruled out first.

Specifically, in methods in which whole test-displays are used, participants have to compare all memory representations of the initial display with all memory representations of the test display. For every item they have to decide whether it has changed or not, and each such decision is subject to a certain degree of error. Therefore, it is highly likely that set size and the probability of mentioned error correlate positively (e.g. Shaw, 1982; Palmer, Ames, & Lindsey, 1993), and accordingly, it is uncertain if the participants' poor performance at large set sizes is the consequence of their memory limitation, or the consequence of accumulated decision errors. This is supported by excellent performance at set sizes 1 or 2 (e.g. Vogel et al., 2001), according to which, it could be reasoned that the rate of individual-item errors is very small.

To investigate this possible bias, Vogel et al. (2001) conducted few control experiments. First of them was the same as their initial experiment, with one difference – the introduction of the cue condition: in the test display, round one of the squares, they placed neutrally colored outlined square, and only that item might have changed color. In other words, if any square had changed color, it was the cued square. In the no-cue condition there was no outlined square. These two conditions were otherwise identical. Better performance in the cue condition would suggest that the capacity of visual working memory is indeed higher then 3-4 items. Especially, if the superiority of the cue condition against the no-cue condition would systematically increase with the increase of set size - that would strongly point to conclusion that the
accumulation of errors at the decision stage had biased visual working memory capacity estimation. Analysis had shown that only the effect of set size was significant. Neither the effect of cue-preservation nor interaction were significant. Although this result directly suggested that the accumulation of errors at decision stage did not bias the capacity estimate, Vogel et al. (2001) conducted two more control experiments to be certain.

In the first of them, they demonstrated that participants were capable of using the cue, and in second one, they tested the possibility that the mere presence of uncued items may have impaired their performance by launching automatic decision processes. To investigate that possibility, they conducted an experiment in which they presented only one square after the retention interval, and participants had to answer whether the color of that square was identical to the color of a square which was earlier presented at same location along with other squares. Instead of complete removal of the irrelevant (uncued) items from test display, Vogel et al. (2001) used placeholders – outlined squares placed in the locations of irrelevant items (Figure 14). The use of placeholders was necessary because Klemer (1963) and Vogel et al. (2001) themselves found that the complete removal of irrelevant items disrupts memory performance, probably as a result of Gestalt alteration. Performance in that condition, in which it was certain that participants made only one "same-different" decision per trial, was compared to the performance in the cue condition from the control experiment described in the previous paragraph, in which the item that could have changed was cued in test-display. Analysis have shown that performance was equally good in both conditions at set sizes 8 and 12, while at set size 4, performance was actually slightly but significantly worse in the condition with placeholders (single-test condition), then in the test-cue condition. Authors had no explanation for that effect. From this experiment, it is obvious that accumulated decision errors did not bias visual working memory capacity, which was observed in Vogel et al's (2001) first set of experiments with full test displays.
Figure 14. Vogel et al.’s (2001) placeholders. With the use of placeholders, color information of all but one item are eliminated from test display, while spatial distribution of objects remains unchanged during the retention interval.


2.6.2 Encoding limitations

Accumulated decision errors are not the only threat to accuracy of visual working memory capacity estimation. Relatively small capacity observed in Vogel et al's (2001) experiments might still be a consequence of a bias: perhaps stimuli were not adequately encoded into working memory. It is possible that the duration of the initial presentation of stimuli (100 ms) was too short to identify and encode all stimuli, especially at large set sizes. To investigate that, Vogel et al. (2001) conducted two control experiments.

In first of them, they tested whether 100 ms duration of initial display is sufficient for encoding all stimuli. According to Potter's (1976) results, it is doubtful that participants had enough time to identify all colors and transfer the perceptual representations into stable visual representations during such short interval. If that indeed was the case, participants would then be unable to respond correctly not because of a limitation in the capacity of visual working memory, but due to inadequate encoding of stimuli. To investigate this, Vogel et al. (2001) compared the original 100-ms to a 500-ms duration of the initial display, while the retention interval
was held constant. By increasing the initial display duration, they ensured enough
time for perceiving the stimuli and encoding them in memory (Vogel et al., 2001).
However, variations in the initial display duration had no influence on participants' performance.

Second control experiment was focused on the identification of colors of
squares. By the use of visual search tasks, Vogel et al. (2001) again examined if poor
performance at large set sizes was the consequence of insufficient encoding time or
the result of storage limitations. Each trial in that experiment began with the
presentation of one colored square, in duration of 500 ms. After 500-ms delay, array
of 4, 8 or 12 colored squares was presented for 100 ms, and participants had to answer
whether that array contained the target color (color of a square presented at the
beginning of trial), or not. Accuracy of participants' responses was near 100%,
regardless of set size, and that result drove Vogel et al. (2001) to conclude that 100 ms
duration interval of initial stimuli presentation is sufficient for participants to identify
colors of all squares.

Consequently, Vogel et al. (2001) concluded that the results of these two
experiments provide enough evidence to rule out the hypothesis that limitations in
perceiving the stimuli and encoding them into memory led to underestimation of
visual working memory capacity. After all their control experiments, Vogel et al.
(2001) stated that now they were rather sure that relatively small visual working
memory capacity observed in their study was not a result of limitations in any process
other than visual working memory storage.

However, this conclusion might still be incorrect, and that is the central
problem of the current research.

2.7 Problem of the present study

The aim of this study was to answer the following question: Is visual working
memory capacity indeed so extremely limited, or is it erroneously underestimated due
to inadequate encoding?

Vogel et al. (2001), and other investigators along with them, had overlooked
one important issue before they discarded the possibility that stimuli were not
adequately encoded in experiments which assess visual working memory capacity.
Specifically, attention is undoubtedly crucial for encoding visual stimuli. If attention is not directed to visual stimuli, then they cannot be adequately encoded into visual working memory (e.g. Coltheart, 1980; Irwin, 1991; Resnink, O'Regan, & Clark, 1997). Attention can simultaneously be directed to maximally 3-4 objects (e.g. Pylyshyn & Storm, 1988; Yantis, 1992), and on the other hand, exactly 3-4 items can be stored into visual working memory (e.g. Vogel et al., 2001; Wheeler & Treisman, 2002; Alvarez & Cavanagh, 2004).

Thus, when all findings relevant for setting up the main problem of current study are considered, they can be summarized in four facts:

1. Visual object can not be adequately encoded if attention is not directed to it.
2. Attention can simultaneously be directed to not more then four visual objects.
3. Not more then four objects can be stored into visual working memory.
4. All experiments that investigate visual working memory capacity are based on the same change-detection paradigm. In these experiments, objects are simultaneously displayed before retention interval. Displays usually contain up to 16 objects, and their duration is extremely short: usually initial exposition lasts 100-150 ms (except in one of Vogel et al.'s (2001) experiments, where it equaled 500 ms).

The following question supervenes itself from the facts presented above: Is it possible that even in Vogel et al.'s (2001) experiment with 500-ms exposition of initial display, participants did not encode all presented stimuli at large set sizes?

Finding that only 3-4 objects can be attended simultaneously points to affirmative answer to the question above, because the same number of objects can be stored into visual working memory. Thus, perhaps the observed capacity equaled only 3-4 objects, exactly because attention was directed to no more nor less then 3-4 objects. The rest of objects could not reach visual working memory, since they were never adequately encoded because attention was not directed to them during initial presentation.

By prolonging the interval of initial display presentation in one of their control experiments, Vogel et al. (2001) have not ensured attentional shifting towards every
single displayed object, especially at large set sizes. Therefore, their conclusion that objects were adequately encoded remains doubtful. Adequate encoding of stimuli is even more questionable in experiments in which sets of objects were presented in a duration of 100-150 ms, and the majority of studies concerning visual working memory capacity relies on such experiments.

Therefore, the central idea of the present study was to force participants to attend to every single displayed stimulus. If that could be accomplished, then it would be certain that all of the stimuli were successfully encoded into memory. Directing attention to each presented stimulus was ensured by successive initial stimuli presentation. Participants' performance in that condition was compared to their performance in the classical change detection paradigm with simultaneous initial presentation of stimuli.

Besides the initial presentation, testing of memory was also varied. In one condition memory was tested with partial test-displays, and in the other, full test-displays were applied.

Therefore, 2 x 2 experimental design was used, with type of presentation (simultaneous / successive) and type of test (partial test-display / full test-display) as factors. Visual working memory capacities were estimated and compared for each of these four conditions.
3  METHOD

3.1 Participants

Thirty-seven undergraduate psychology students (age range 19-28, 2 male) from The University of Rijeka, Croatia, participated in the experiment. All of them reported to have normal color vision and normal or corrected to normal visual acuity. Due to experimenter’s technical mistake regarding stimuli presentation, one participant had to be excluded from the analyses.

3.2 Instruments

Stimuli were presented on a 17-inch monitor with resolution of 1024 x 768 pixels. Stimuli presentation was controlled by PC-computer and responses were collected by keyboard.

3.3 Stimuli and procedure

Experiment was conducted in a laboratory. Noise was minimized, and illumination was held constant for each participant. Participants sat with their eyes at a distance of 80 cm from a monitor.

Colored squares – yellow, green, red, blue, violet and black – were used as stimuli. Exactly these colors were used in order to maximize discriminability. Every stimulus subtended 2,58 ° x 2,58 ° of visual angle, and all of them were presented on a white background.

Each participant went through both display conditions (full and partial) and through one presentation condition – eighteen participants (1 male) went through the simultaneous presentation condition, and the other eighteen participants (1 male) went through the successive presentation condition. Each of these four conditions consisted of 10 practice trials and 60 experimental trials (practice trials were discarded from the analyses). Colors and locations of stimuli were selected randomly for each trial, but
same trials were used for every experimental condition. The presentational order of all trials within a condition was randomised across participants.

    Every trial began with the fixation mark, in duration of 500 ms, which was followed by the presentation of eight colored squares. Colors of all squares were chosen randomly, without restriction – theoretically, all eight squares could have been the same in color. These squares were distributed across 8 of 15 possible positions on the display, as shown in Figure 1, with important requirement: each column had to contain at least one square – there could be no empty columns. In other words, stimuli were presented in pseudorandom positions within a 3 (rows) x 5 (columns) grid subtending 18.25° x 10° of visual angle. Stimuli were separated from each other by at least 1.07°. Their center to center separation equalled 3.94° (Figure 15).

3.3.1 Presentation manipulation

    In the simultaneous presentation condition, whole pictures were presented at the same time, in duration of 500 ms (Figure 16), while in the successive presentation condition pictures were revealed from the left to the right side. In other words, their presentation was divided into five 100 ms intervals (see Figure 17). In the first interval after the fixation mark, only the first column was presented; in the second interval, the first and the second column were visible, and so on, till the last 100 ms interval, in which the whole picture was visible (same as the entire simultaneous presentation condition). Thus, total presentation time was identical in both conditions – it equalled 500 ms.

3.3.2 Change

    After 500 ms-lasting initial simultaneous or successive presentation, blank screen was presented for 900 ms. During that retention interval, change could occur in one of the colored squares. Half of all trials within every condition contained a color change, and the other half did not. Changes could occur in only one of eight squares, and the color of the changed stimuli was randomly chosen among five other colors,
Figure 15. Size and location of stimuli.
Figure 16. *Simultaneous presentation of stimuli.*
Figure 17. *Successive presentation of stimuli*
with no restriction. During the retention interval, squares could never change size, location, nor anything else but color. Changes were equally distributed across the display – they occurred the same number of times on every column in the grid. Participants were familiar with all these information concerning change.

### 3.3.3 Test manipulation

Retention interval was followed by a presentation of a test display. There were two test-display conditions: full and partial. Full test-display was the same display as the display that was presented just before the retention interval. (Naturally, in trials that contain a change, full test display differed from the display presented just before the retention interval, in color of one of the squares.)

In partial test (or single test) condition, only one colored square was presented in the test array. Other squares were replaced with placeholders – black outlined squares of the same sizes, which were placed exactly on their locations (see Figures 16 and 17).

### 3.3.4 Participants’ tasks

After presentation of a test display, participants’ task was to hit “1” key if there was a color change during the retention interval, or to press “0” key if they believed that there was no change. Also, they were instructed to respond by chance if they were unsure, and to aim for accuracy, not speed.

Answering was followed by immediate feedback. The word “correct” appeared in blue color at the centre of display if the response was correct and the word “incorrect” in red color appeared if the response was wrong.

The experimental session proceeded in a self-paced manner. After presentation of feedback, which lasted for 3000 ms participants had to press the space bar in order to start a new trial.

Participants had one more task. They had to repeat the phrase *Coca-cola* aloud during the initial display and memory delay periods. They were allowed to stop only after they had selected the answer. This verbal task was used in order to reduce the
effect of rehearsal in verbal working memory. Similar procedure was introduced by Wheeler & Treisman (2002).

The experimental session was divided into two blocks – one with full, and the other one with partial test display, for each participant. Between these two blocks, participants were taking a 20 minute break. Since the duration of each block was also round 20 minutes, the entire experimental session (including the break) lasted for about one hour.
4 RESULTS AND DISCUSSION

4.1 The effects of the type of initial presentation and the type of test display on participants' performance

4.1.1 Descriptive results

Participants’ accuracy (percentage of their correct answers) and Pashler’s (1988) capacity estimate were used as dependent measures to investigate whether type of presentation and type of test display have effects on participants’ performance. Also, reaction time was used as a supplementary measure. Median reaction times were first computed across conditions for each participant, and then subjected to further analysis. Obtained descriptive statistics are shown in Table 1.

Table 1. Participants’ performance measured by three dependent variables (percentage correct, Pashler’s (1988) visual working memory capacity estimate and reaction time) across experimental conditions.

<table>
<thead>
<tr>
<th>Type of presentation and test display (experimental condition)</th>
<th>Percentage correct</th>
<th>Capacity estimate</th>
<th>Reaction time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Simultaneous presentation and full test display</td>
<td>76,85</td>
<td>11,19</td>
<td>5,01</td>
</tr>
<tr>
<td>Simultaneous presentation and partial test display</td>
<td>73,61</td>
<td>6,72</td>
<td>5,90</td>
</tr>
<tr>
<td>Successive presentation and full test display</td>
<td>74,54</td>
<td>6,78</td>
<td>4,64</td>
</tr>
<tr>
<td>Successive presentation and partial test display</td>
<td>72,50</td>
<td>6,60</td>
<td>5,52</td>
</tr>
</tbody>
</table>
4.1.2 Accuracy percentage as a dependent variable

After descriptive analysis, accuracy percentages were subjected to a mixed two-way ANOVA with the type of initial presentation (on two levels: simultaneous / successive) as a between subject factor, and the type of test display (on two levels: full test display / partial test display) as a within subject factor. The main effect of the type of presentation was not significant, \( F(1, 34) = .40; p > .05 \). Also, the main effect of the type of test display was not significant, \( F(1, 34) = 3.35; p > .05 \), and there was no interaction between the type of presentation and the type of test display, \( F(1, 34) = .37; p > .05 \). These results are shown in Figure 18.

Figure 18. Change-recognition accuracy percentages as a function of the type of initial presentation and the type of test display.
4.1.3 Pashler's capacity estimate as a dependent variable

After accuracy percentages, Pashler’s (1988) capacity estimates were entered into ANOVA. According to Pashler's (1988) equation, visual working memory capacity (C) can be estimated from the following formula:

\[ H = \frac{C}{IP} + \frac{IP - C}{IP} * FA, \]

which can be transformed to:

\[ C = \frac{IP * (H - FA)}{1 - FA}, \]

where \( H \) refers to hit rate, \( IP \) is total number of items presented in a display, and \( FA \) is false alarm rate.

Capacities computed by this equation were entered into a mixed two-way ANOVA with the type of initial presentation (on two levels: simultaneous / successive) as a between subject factor, and the type of test display (on two levels: full test display / partial test display) as a within subject factor. The main effect of the type of initial presentation was not significant, \( F (1, 34) = 1.05; p > .05 \). The main effect of the type of test display was significant, \( F (1, 34) = 11.85; p < .05 \). In partial test display condition, visual working memory capacities were higher, compared to full test display condition. There was no interaction between the type of presentation and the type of test display, \( F (1,34) = .00; p > .05 \). These results are presented in Figure 19.
Figure 19. *Visual working memory capacity estimated by Pashler’s (1988) formula, as a function of the type of initial presentation and the type of test display.*

<table>
<thead>
<tr>
<th>Type of Initial Presentation</th>
<th>Full Test Display</th>
<th>Partial Test Display</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simultaneous</td>
<td>4,0</td>
<td>4,5</td>
</tr>
<tr>
<td>Successive</td>
<td>5,0</td>
<td>5,5</td>
</tr>
</tbody>
</table>

**4.1.4 Reaction time as a dependent variable**

Besides accuracy percentage and visual working memory capacity, reaction time was also analysed. Median reaction times of each participant across all four experimental conditions were subjected to a mixed two-way ANOVA with the type of initial presentation (on two levels: simultaneous / successive) as a between subject factor, and the type of test display (on two levels: full test display / partial test display) as a within subject factor (Figure 20). The main effect of the type of initial presentation was not significant, $F(1, 34) = .01; p > .05$. The main effect of the type of test display was not significant, $F(1, 34) = .07; p > .05$. Also, there was no significant interaction between the type of presentation and the type of test display, $F(1, 34) = 2.52; p > .05$.

When the results from these analyses are summarized, two main conclusions can be drawn:
1. Type of initial presentation had no effects on participants’ performance. In the successive initial presentation condition, participants perform approximately as accurate and as fast as in the simultaneous presentation condition. Also, manipulation with the type of initial presentation had no effects on participants’ visual working memory capacity. They managed to successfully store approximately the same number of visual objects in successive presentation trials as in the simultaneous presentation trials.

2. Manipulation with the type of test display had no effect on the percentage of participants’ correct answers, nor on their reaction time, but when visual working memory capacities were used as dependent measure, it was found that participants are able to store significantly more objects when their memory is tested with partial displays (compared to testing with full displays).

Figure 20. *Reaction time as a function of the type of initial presentation and the type of test display.*
4.2 Is visual working memory capacity indeed so limited, or is it underestimated due to encoding limitations?

The finding that the type of presentation had no effect on the percentage of correct detections nor on the visual working memory capacity, does not corroborate the assumption according to which visual working memory capacity seems to be limited because of inadequate encoding of visual material. For that assumption to be true, the main effect of the type of initial presentation would have to be significant: participants would have to answer significantly more correctly in the successive initial presentation condition then in the simultaneous initial presentation condition. Thus, this result is consistent with the conclusions of several studies, all of which indicate that visual working memory capacity is indeed limited to no more then 4-5 units (e.g. Luck & Vogel, 1997; Vogel et al., 2001; Wheeler & Treisman, 2002; Alvarez & Cavanagh, 2004).

This result also corroborates Vogel et al.'s (2001) claim that there are no encoding limitations in classical change detection paradigm, which could consequently lead to an underestimation of visual working memory capacity. Accordingly, methods of Luck & Vogel (1997), Wheeler & Treisman (2002) and Alvarez & Cavanagh (2004), which are very similar to Vogel et al.'s (2001) procedure, are also valid together with their conclusions.

Therefore, the main question of this study is answered: in numerous investigations using change detection paradigm, visual working memory capacity was not erroneously underestimated due to inadequate encoding of stimuli. Vogel et al. (2001) were indeed right when they concluded that 100-ms initial simultaneous presentation is sufficient for encoding up to 16 visual stimuli. In the present study, introduction of successive initial presentation did not result with an increase of visual working memory capacity, although in that situation it was ensured for every stimulus to undoubtedly be encoded into visual working memory. Even in that situation, participants were able to store no more visual material, compared to classical change
detection condition that is based on simultaneous initial presentation. Thus, visual working memory is limited to round 4 units\(^4\).

### 4.3 Drawbacks of Pashler’s (1988) visual working memory capacity estimation procedure

#### 4.3.1 Full versus partial test display: is there a difference in participants’ performance regarding the type of test display?

Regarding the test display condition, the results acquired are illogical at first sight. When Pashler’s (1988) visual working memory capacity estimation was entered into ANOVA as dependent measure, the main effect of the type of test display was significant: visual working memory capacity was significantly higher in the partial test display condition (Figure 19). In contrast, when percentage of correct answers (Figure 18) or reaction time (Figure 20) was taken as dependent measure, the main effect of the type of test display was not significant. The percentage of correct answers was even somewhat higher in the full display test condition (Figure 18), but that effect was not significant (\(p = .08\)). That result is indeed surprising, and in order to explain it, Pashler’s (1988) equation will have to be thoroughly re-analysed.

In Pashler’s (1988) formula, visual working memory capacity (\(C\)) depends on three variables: number of presented items (\(IP\)), hit rate (\(H\)) and false alarm (\(FA\)) rate:

\[
C = \frac{IP \times (H - FA)}{1 - FA}.
\]

\(^4\) Pashler’s (1988) estimates suggest that observed visual working memory capacity was higher than 4 items in the present study. However, as will be shown later, Pashler’s (1988) procedure overrated true capacity.
In the present study, the number of presented items was held constant (in every experimental condition, eight items were initially presented), so the only two variables that determined Pashler’s (1988) capacity estimation were hit rate and false alarm rate.

Why did the contradictory result occur? Is it possible that participants used different response criteria in different experimental conditions? The logic of this hypothesis will be explained in the following examples. Three hypothetical situations will be shown. (For simplicity, only partial test display condition will be considered in these two examples.)

4.3.1.1 Hypothetical situation A: Lenient criterion

What would happen if in the present study participants were instructed to answer "change absent" only when they were 100% sure that the change did not occur during the retention interval, and to answer "change present" in all other trials?

To assure that, participants would be instructed to attend to and to memorize only two items, because two items can undoubtedly be successfully held in visual working memory. They would be given this instruction: "Eight colored squares are going to be presented in each trial. Your task is to attend to and to memorize two of these eight squares. After a short retention interval, test situation will follow: only one test square will be presented on one location which was previously occupied by one of eight initially presented squares. If the test square has the same color as the square initially presented in its location, then a change did not occur during the retention interval. Otherwise, change did occur. You have to answer "change absent" only if you are 100% sure that the change did not occur. If change did occur, then answer "change present". Also, you have to answer "change present" in all trials in which you are unsure."

If participants strictly followed the instruction, then the following scenario would happen:
1. **Hit rate would be very close to 1, and there would be no misses**

Participants would detect a change in about 25% of trials that contain a change (only trials containing a change are relevant for hit rate calculation), because in that percentage of trials, the location of the test item would be the same as the location of one of two items that they attended to and memorized. In these trials, the participants would naturally respond "change present". In the rest of the trials that contain a change, the location of the test item would be different from the location of one of two items they had memorized. In these trials, the participants would again answer "change present", because they were instructed to do so. Therefore, for every trial that contained a change, participants would answer "change present", so hit rate would equal one.

2. **False alarm rate would equal round 0,75, and the proportion of correct rejections would be about 0,25**

In one quarter of all trials that do not contain a change (only trials which do not contain a change are relevant for calculation of false alarms and correct rejections), the location of the test item would be the same as the location of one of two items which participants attended to and memorized. In these trials, the participants would be 100% sure that there was no change during the retention interval, so they would answer "change absent". In the other 75% of the trials that do not contain a change, the location of test item would not be same as the location of one of two items participants had memorized. In these trials the participants would answer "change present", as they were instructed to. Therefore, the participants would score 25% correct rejections, and their false alarm rate would equal round 0,75.

3. **Percentage of correct answers would equal round 62,5%**.

In all trials that contain a change, the participants would correctly answer "change present", so in half of all trials, the percentage of their correct answers would be 100%. In 25% of the trials that do not contain a change, the participants would correctly answer "change absent", and in the rest of these trials, they would inaccurately answer "change present". Thus, in the other half of all trials, the
percentage of participants’ correct answers would be 25%. Therefore, their overall percentage of correct answers would equal 62.5%.

When these values are entered into equation, visual working memory capacity estimated by Pashler’s (1988) formula equals 8 items!

\[ H = 1 \]
\[ FA = 0,75 \]
\[ IP = 8 \]

\[ C = \frac{IP \times (H - FA)}{1 - FA} \]

\[ C = \frac{8 \times (1 - 0.75)}{1 - 0.75} = 8 \text{ items} \]

### 4.3.1.2 Hypothetical situation B: Strict criterion

Analogously to the previous hypothetical situation, where participants were instructed to answer "change absent" only when they were 100% sure that the change did not occur during the retention interval, what would happen if they were told the opposite – to answer "change present" only if they are 100% sure that a change had occurred?

As in the previous scenario, participants would again be instructed to attend to and to memorize only two of eight items, but they would be given different instruction: "Eight colored squares are going to be presented in each trial. Your task is to attend to and to memorize two of these eight squares. After a short retention interval, test situation will follow: only one test square will be presented on one location which was previously occupied by one of eight initially presented squares. If the test square has the same color as the square initially presented on its location, then a change did not occur during the retention interval. Otherwise, change did occur. You have to answer "change present" only if you are 100% sure that a change did occur. If a change did not occur, then answer "change absent". Also, you have to answer "change absent" in all trials in which you are unsure."

72
Obeying this instruction would lead to the following:

1. **Hit rate would be approximately 0.25, and the proportion of misses would be round 0.75**

Participants would detect a change in about 25% of the trials that contain a change (only trials containing a change are relevant for the calculation of hits and misses), because in that percentage of trials, the location of the test item would be the same as the location of one of two items that they attended to and memorized. In the rest of the trials that contain a change, the location of the test item would be different from the location of one of two items they had memorized. In these trials, the participants would answer "change absent", because they were instructed to do so. Therefore, for only one quarter of the trials that contained a change, participants would answer "change present", so hit rate would equal approximately 0.25.

2. **False alarm rate would be very close to zero, while proportion of correct rejections would be very close to 1.**

Only trials which do not contain a change are relevant for calculation of false alarms and correct rejections. In one quarter of such trials, the location of the test item would be the same as the location of one of two items which the participants attended to and memorized. In these trials, participants would be 100% sure that there was no change during the retention interval, so they would answer "change absent". In the other 75% of the trials that do not contain a change, the location of test item would not be the same as the location of one of two items the participants had memorized. In these trials the participants would again answer "change absent", because they were instructed to do so. Therefore, participants would reach nearly 100% correct rejections, and their false alarm rate would equal around zero.

3. **Percentage of correct answers would equal round 62.5%.**

In all trials that do not contain a change, participants would correctly answer "change absent". In one quarter of the trials that do contain a change, the participants would
correctly answer "change present", while in the rest of these trials, they would erroneously answer "change absent". Thus, their overall percentage of correct answers would be about 62.5% (100% for trials that do not contain a change and 25% for trials that do contain a change).

When these values are entered into Pashler’s (1988) equation, then estimated visual working memory capacity equals 2 items!

$$H = 0.25$$
$$FA = 0$$
$$IP = 8$$

$$C = \frac{IP \times (H - FA)}{1 - FA}$$

$$C = \frac{8 \times (0.25 - 0)}{1 - 0} = 2 \text{ items.}$$

Summarized, these two examples point to several weaknesses of Pashler's (1988) procedure of visual working memory capacity assessment.

Firstly, Pashler's (1988) memory assessment is heavily affected by the criterion of participants' responding. When participants aimed to make as less as possible false alarms, regardless of misses, their visual working memory capacity was estimated to be 2 items. On the other hand, when they were instructed to be careful not to make any misses, regardless of false alarms, then their memory capacity was estimated to be 8 items. Although the percentage of correct answers and $H - FA$ were held constant in two hypothetical situations, the capacity estimates differed greatly.

Secondly, it seems that Pashler's (1988) procedure overestimates real visual working memory capacity. In both of the two given examples the participants held only 2 items in their working memory, but nevertheless, in the case when hit rate and false alarm rate were high (Hypothetical situation A), Pashler's (1988) formula gave an estimate that they were holding 8 items in working memory. In the case when hit rate and false alarm rate were low (Hypothetical situation B), Pashler's (1988) calculation gave correct estimate of memory capacity, so it is still slightly doubtful
that Pashler's (1988) method overestimates memory capacity. These hypothetical situations were indeed extreme, so another, realistic example will be presented to examine the problem of memory capacity overestimation in Pashler's (1988) method.

### 4.3.1.3 Hypothetical situation C: A realistic situation

As previous two examples, this one is also based on the method used in the present study, with initial presentation of 8 objects. It is found that participants can hold about 4 items in visual working memory (Luck & Vogel, 1997; Wheeler & Treisman, 2002; Alvarez & Cavanagh, 2004; Vogel et al., 2001) and that is going to be one of the assumptions in this example. So, if participants can hold 4 objects in memory, what should be expected in the present study, in which participants were given no instruction regarding the criteria of responding? The only information they had was that 50% of all trials contain a change, and the rest of trials do not.

1. **Hit rate would be approximately 0.75, and proportion of misses would be about 0.25.**

In one half of all trials containing a change, the test item would be presented in the location of one of four items which participants had stored and held during the retention interval, so in these trials the participants would correctly answer "change present". In the other half of the trials containing a change, participants would be unsure, so they would guess: probably half of guesses would be "change present", and the other half would be "change absent". For each trial in which participants are guessing, there is a 50% chance to answer correctly. Therefore, in all trials containing a change, participants would obtain hit rate of round 0.75, while the proportion of misses would be about 0.25.

2. **False alarm rate would be about 0.25, while the proportion of correct rejections would be round 0.75.**

In 50% of the trials that do not contain a change, the test item would be presented on the location of one of four items which the participants had stored and held during the
retention interval. In these trials, the participants would correctly answer "change absent". In the other 50% of trials that do not contain a change, participants would be unsure, so they would guess. In about half of the trials in which they are unsure, they would guess "change present", while in the other half of the trials in which they are not sure, they would guess "change absent". For each guess, there is 50% chance of correct answer, thus the participants would score 75% correct rejections, while their false alarm rate would be about 0,25.

3. **Percentage of correct answers would equal round 75%**.

When these values are entered into equation, visual working memory capacity estimated by Pashler's (1988) formula equals 5,33 items!

\[
H = 0,75 \\
FA = 0,25 \\
IP = 8
\]

\[
C = \frac{IP \times (H - FA)}{1 - FA}
\]

\[
C = \frac{8 \times (0,75 - 0,25)}{1 - 0,25} = 5,33 \text{ items.}
\]

Therefore, the inexactness of Pashler's (1988) method is again demonstrated – this time in a realistic simulation: in addition to depending on responding criterion, it overestimates true visual working memory capacity. While participants held 4 items in working memory, Pashler's (1988) method estimated their visual working memory capacity to equal 5,33 items.
4.3.2 Analyses of hit rate and false alarm rate observed in the present study

Since it was demonstrated (in Hypothetical situations A & B) that Pashler's (1988) visual working memory capacity estimation depends on participants' criterion of responding, hit rates and false alarm rates observed in the present study will be analysed. The purpose of these analyses is clarification of conflicting results regarding the effects of the test display type on participants' performance.

Means and standard deviations of hit rates and false alarm rates for each experimental condition are shown in Table 2. Hit rates and false alarm rates were separately subjected to a mixed two-way ANOVA with the type of initial presentation (on two levels: simultaneous / successive) as a between subject factor, and the type of test display (on two levels: full test display / partial test display) as a within subject factor.

Table 2. Hit rates and false alarm rates across experimental conditions.

<table>
<thead>
<tr>
<th>Type of presentation and test display (experimental condition)</th>
<th>Hit rate</th>
<th>False alarm rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simultaneous presentation and full test display</td>
<td>0.70</td>
<td>0.16</td>
</tr>
<tr>
<td>Successive presentation and full test display</td>
<td>0.65</td>
<td>0.16</td>
</tr>
<tr>
<td>Simultaneous presentation and partial test display</td>
<td>0.84</td>
<td>0.37</td>
</tr>
<tr>
<td>Successive presentation and partial test display</td>
<td>0.80</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Analysis of variance with hit rate as dependent variable has shown significant main effect of the type of test display, F (1, 34) = 32.03; p < .05. When memory was tested with partial test displays, hit rates were higher in comparison to condition
where memory was tested via full test displays. The main effect of the type of presentation was not significant, $F (1, 34) = 1.63; p > .05$; and there was no interaction between the type of presentation and the type of test display, $F (1, 34) = .02; p > .05$. These results are presented in Figure 21.

Figure 21. *Hit rates as a function of the type of initial presentation and and the type of test display.*

When false alarm rates were entered into mixed two-way ANOVA (Figure 22), the main effect of test display type was again significant, $F (1,34) = 92.98; p<.05$. False alarm rates were higher when memory was tested with partial test displays. The main effect of the type of presentation was not significant, $F (1, 34) = 0.06; p > .05$; and there was no interaction between the type of presentation and the type of test display, $F (1, 34) = .21; p > .05$. 

![Graph showing hit rates as a function of presentation and test display type](image-url)
Summarized, hit rates and false alarm rates were significantly higher in partial test display conditions (Table 3). Exactly because of that, the main effect of type of test display was significant only when Pashler's (1988) visual working memory capacity estimations were entered into ANOVA as dependent variable, while in other two analyses of variance (with reaction time and percentage of correct answers) that effect was not significant.

As it was demonstrated earlier, capacities estimated by Pashler's (1988) method are dependent on participants' criterion of responding, that is, they are dependent on hit rate and false alarm rate. Specifically, the higher hit rate and false alarm rate are, the higher estimated capacity is (even when $H - FA$ and / or percentage of correct answers is constant). If hit rate reaches 1, then estimated capacity equals set size, regardless of false alarm rate:
\[ C = \frac{IP^*(H - FA)}{1 - FA} \]
\[ C = \frac{IP^*(1 - FA)}{1 - FA} \]
\[ C = IP \]

So, if participants manage to make at least one correct rejection (if not then denominator value would equal zero, and capacity could not be calculated), and answer "change present" in every trial but that single one, then Pashler’s (1988) formula would estimate that they had successfully encoded and memorized every single stimulus in the entire experimental session, regardless of set size. That is obviously senseless.

In the present study, neither percentages of correct answers nor reaction times did not significantly differ between full and partial test display conditions, while hit rates and false alarm rates were significantly higher in partial test display conditions. Two conclusions can be drawn from these findings:

1. Responding criterion was different between full and partial test display conditions. Since higher hit rates and false alarm rates were observed in partial test display condition, it is certain that participants were responding "change present" more frequently in that condition, compared to full test display condition. In full test display condition participants were responding "change absent" more often.

2. Participants’ visual working memory was approximately the same in full and partial test display conditions, because the percentage of correct answers did not differ significantly.
Table 3. Average hit rates, false alarm rates and percentage of correct answers as a function of test display (both conditions of presentation type are included).

<table>
<thead>
<tr>
<th>Type of presentation and test display (experimental condition)</th>
<th>Hit rate</th>
<th>False alarm rate</th>
<th>Pashler’s capacity estimation</th>
<th>Percentage correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>Full test display (simultaneous and successive presentation)</td>
<td>0.67</td>
<td>0.15</td>
<td>0.16</td>
<td>4.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.19</td>
</tr>
<tr>
<td>Partial test display (simultaneous and successive presentation)</td>
<td>0.82</td>
<td>0.09</td>
<td>0.36</td>
<td>5.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.59</td>
</tr>
</tbody>
</table>

Thus, when all this is summarized (Table 3), it is obvious that the finding that visual working memory capacity is higher in partial test display condition, is not valid. It is the consequence of the inaccuracy of Pashler’s (1988) equation, which has a tendency to fallaciously overestimate true memory capacity, especially if the responding criterion is lenient (aiming for high hit rate, at the expense of increasing false alarm rate), as it was in the partial test display condition. Finding that accuracy percentages were nearly the same in partial and in full test display condition also corroborates the conclusion that true visual working memory capacities did not differ as a function of test display type. Finally, this conclusion is consistent with the other two studies in which type of test display was manipulated. Vogel et al.’s (2001) and Wheeler & Treisman’s (2002) test display manipulation resulted with identical findings: performance was equally good in full test display condition as it was in the partial test display condition.

Now, when the question of the effect of test display type is solved, another enigma has emerged: why did not participants use the same criteria of responding in full and partial test display condition?
4.3.2.1 A change in criterion of responding

The answer to that question is not obvious immediately, but when the problem is thoroughly analysed, a logical explanation of this phenomena is found. To make that explanation more understandable, a simulation of participants’ reasoning will be presented. That simulation is rough and simplified, and is schematically shown in Figure 23. Figure 23 consists of three segments:

a) **initial presentation**
   Only one display is presented for initial presentation, because type of presentation had no effects on participants’ performance. Also, last display presented before retention interval is the same in simultaneous as in the successive presentation condition.

b) **memorized items**
   Items that are held in memory during the retention interval are shown in this segment, under assumption that four items can be maintained in visual working memory.

c) **test**
   In the last segment of Figure 23, test displays for every possible scenario are shown:
   - TEST A – full test display + change absent + all items presented on test display
   - TEST B – full test display + change present + one of items that are held in memory has changed
   - TEST C – full test display + change present + one of items that are not held in memory has changed
   - TEST D – partial test display + change absent + one of items that are held in memory is presented on test display
   - TEST E – partial test display + change absent + one of items that are not held in memory is presented on test display
   - TEST F – partial test display + change present + one of items that are held in memory has changed
   - TEST G - partial test display + change present + one of items that are not held in memory has changed
Now, under assumption that visual working memory capacity equals 4 items, all possible outcomes will be considered separately for full and for partial test display condition:

Figure 23. *Simulation of participants’ reasoning in change detection tasks, under assumption that 4 items can be held in memory. Set size = 8.*
1. *In 50% of all trials in partial test display condition, test item is presented in the location of one of 4 items that are held in memory (Test D & Test F situation in Figure 23).* 
   Outcome: Participants would answer "change present" if change had occurred (Test F situation in Figure 23) or "change absent" if change had not occurred (Test F situation in Figure 23). Therefore, for half of all trials, participants would score:
   - Hit rate = 1
   - False alarm rate = 0
   - Miss rate = 0
   - Correct rejection rate = 1
   - Percentage correct = 100%

2. *In 25% of all trials in partial test display condition, change occurs on an item that is not held in memory, and that item is presented at test (Test G situation in Figure 23).* 
   Outcome: Participants would be unsure, and they would guess. They would score:
   - Hit rate = 0,50
   - Miss rate = 0,50
   - Percentage correct = 50%

3. *In 25% of all trials in partial test display condition, test item is presented in the location of one of 4 items that are not held in memory, and is the same as the object that had initially occupied its location (Test E situation in Figure 23).* 
   Outcome: Participants would be unsure, and they would guess. They would score:
   - Correct rejection rate = 0,50
   - False alarm rate = 0,50
   - Percentage correct = 50%.
Overall outcomes for partial test display condition can now be easily calculated:

- Hit rate = 0.75
- False alarm rate = 0.25
- Miss rate = 0.25
- Correct rejection rate = 0.75
- Percentage correct = 75%

**B: FULL TEST DISPLAY**

1. In 25% of all trials in full test display condition, change occurs on one of the items that are held in memory (Test B situation in Figure 23).
   
   Outcome: Participants will detect a change and answer "change present".
   
   Therefore:
   
   - Hit rate = 1
   - Miss rate = 0
   - Percentage correct = 100%

2. In 25% of all trials in full test display condition, change occurs in a location that was initially occupied by an item not held in memory (Test C situation in Figure 23).

   In 50% of all trials in full test display condition, all items are presented at test, and there is no change (Test A situation in Figure 23).

   Outcome: In all of these trials, the participants would not detect a change. When they would compare all memorized items to test items, all comparisons would match, so consequently, it would appear to them that there was no change, so they would answer "change absent". Expected percentage of answers "change absent", would be close to 100%. Thus, following results are expected:

   - Miss rate: 1
   - Hit rate: 0
   - False alarm rate = 0
   - Correct rejection rate = 1
   - Percentage correct = 66.67%.
However, these numbers would be slightly different, because it must be expected that in some proportion of these trials, participants would guess. If participants answered "change absent" in all these trials, then in all full test display trials overall, they would have 75% "change absent" and 25% "change present" answers. Since they were informed that 50% of all trials contain a change and the other 50% do not, and since they are given feedback after every trial, they would surely in time conclude that:

a) if they detected a change, then change had surely occurred
b) if they did not detect a change, then change might have occurred.

Consequently, there would be a few hits, and a few false alarms, but not much, because, participants would start guessing only at the end of their experimental session. Also, in the trials in which participants are guessing, at first sight it should be expected that the number of false alarms would be higher than the number of hits, because two thirds of such trials do not contain a change, and only one third of them does. That would ultimately mean that the overall percentage of correct answers in full test display condition would be lower than 75%. However, in that case, the percentage of correct answers in full test display condition would be lower than in partial test display condition, and that is consistent neither with results of Vogel et al. (2001), Wheeler & Treisman (2002), nor with the results of the present study. Therefore, it can for example be hypothesized that contextual cues from surrounding items have some positive effects on participants’ performance. These contextual cues would slightly increase the percentage of correct answers through increase of hit rate and alarm rate. These increases would be slight, and equal overall (in other words, $H - FA$ has to remain 0.50), because otherwise the percentage of correct answers would be different from the percentage correct in the partial test display condition. Therefore, when guessing is considered, then hit rate and false alarm rate are higher but slightly, while the percentage of correct answers remains unchanged. Arbitrarily, the increase of false alarm rate and hit rate is roughly approximated to be round 0.1.
**Overall outcomes for full test display condition can now be calculated:**

- Hit rate = 0.60
- False alarm rate = 0.10
- Miss rate = 0.40
- Correct rejection rate = 0.90
- Percentage correct = 75%

All these expected results together with results observed in the present study, are summarized in Table 4.

Table 4. *Comparison of simulation output and results observed in the present study.*
*Simulation output is printed in blue color, while observed results are printed in red.*

<table>
<thead>
<tr>
<th>display</th>
<th>$H$</th>
<th>$FA$</th>
<th>$M$</th>
<th>$CR$</th>
<th>% correct</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial</td>
<td>0.75</td>
<td>0.25</td>
<td>0.25</td>
<td>0.75</td>
<td>75%</td>
<td>5.33</td>
</tr>
<tr>
<td></td>
<td>0.82</td>
<td>0.36</td>
<td>0.18</td>
<td>0.64</td>
<td>73.06%</td>
<td>5.71</td>
</tr>
<tr>
<td>Full</td>
<td>0.60</td>
<td>0.10</td>
<td>0.40</td>
<td>0.90</td>
<td>75%</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.16</td>
<td>0.33</td>
<td>0.84</td>
<td>75.69%</td>
<td>4.83</td>
</tr>
</tbody>
</table>

*Display = type of test display*

*H = hit rate*

*FA = false alarm rate*

*M = miss rate*

*CR = correct rejection rate*

*% correct = percentage of correct answers*

*C = visual working memory capacity estimated by Pashler’s (1988) procedure*

As it can be seen in Table 4, the output of this simulation is similar to the results observed in the present study:
1. According to simulation, percentages of correct answers are approximately the same for full and partial test display conditions (75%). The difference observed was only 2.63%, and the percentages correct were 73.06% for partial test display condition, and 75.69% for full test display condition.

2. According to simulation, $H - FA$ is approximately the same for full and partial test display condition, and equals round 0.50. Observed $H - FA$ was 0.51 in full test display condition and 0.46 in partial test display condition. Therefore, observed $H - FA$ difference between full and partial test display conditions was only 0.05.

3. Hit rate is higher in partial test display condition. That result is predicted in simulation, and is obtained in the present study.

4. False alarm is higher in partial test condition. That result is predicted in simulation, and is obtained in the present study.

5. Visual working memory capacity estimated by Pashler’s (1988) method, predicted by simulation, is higher in partial test display condition (for 0.89 units). In the present study, it was also higher in partial test display condition (for 0.88 units).

Therefore, in this realistic simulation, it is shown that participants change their criterion of responding between different types of test display, although they hold 4 objects in both full and partial test display condition. In partial test display test condition, they make more hits and more false alarms, compared to full test display condition. Also, the percentage of correct answers is about the same in full as in partial test display condition. However, although in both conditions participants held the same amount of objects (4 objects) in memory, and although the percentage of correct answers is the same for both conditions, Pashler’s (1988) method estimated that in partial test display condition visual working memory capacity equals 5.33 units, while in full test display condition it equals 4.44 units. The pattern of results observed in the present study is approximately the same as the output of this simulation, as it can be seen in Table 4.

In a short summary of this chapter, it should be emphasized that now it is demonstrated that participants did not use the same criterion of responding in full and partial test display condition. That is why they had higher hit rate and false alarm rate
in partial test display condition. Direct consequence of that is that Pashler’s (1988) equation had incorrectly indicated that visual working memory capacity is higher in partial test display condition, compared to full display test display condition, although percentages of correct answers did not differ as a function of test display type. Also, Pashler’s (1988) estimation of capacity is higher then real memory capacity, which equaled 4 items.

Is there a better way of visual working memory capacity assessment?

4.3.3 A better procedure of visual working memory capacity assessment

A better approximation of visual working memory capacity should be directly based on the percentage of correct answers. An interesting method, which is based on this principle was devised by Alvarez & Cavanagh (2004). Although Alvarez & Cavanagh (2004) themselves claimed that Pashler's (1988) method is more reasonable then their own, they were wrong.

4.3.3.1 Alvarez & Cavanagh's method: 75% threshold divided by 2

The experimentation of Alvarez & Cavanagh (2004) was described in introduction, and now their method of visual working memory capacity estimation will be presented. They used an experimental design, based on the classic change detection paradigm (full test display and simultaneous initial presentation was used in all trials), in which set size was varied – set sizes of 1, 3, 5, 7, 9, 11, 13 and 15 stimuli were presented. When calculating visual working memory capacity, Alvarez & Cavanagh (2004) stated that if subjects store half of all presented objects at, for example set size of 8, then they would obtain 75% of correct answers, because then they would detect 50% of all changes, and in all other trials they would answer "change absent". Thus, participants would obtain 100% correct answers in all trials which do not contain a change, and 50% correct answers in all trials that do contain a change. As Alvarez & Cavanagh (2004) varied set size, they were able to derive the
75% threshold number for each subject, and when they divided that number by 2, then visual working memory capacity estimate was obtained.

That method is better than Pashler's (1988), whose flaws were explained earlier, but still, the method of Alvarez & Cavanagh (2004) has several drawbacks. Most importantly, this method was created especially for that specific experiment, and can not be applied to experiments in which set size is not varied. Therefore, this procedure of Alvarez & Cavanagh (2004) requires complex experimental designs. Secondly, it disregards guessing, and thirdly, this procedure in its original form can not be applied to experiments using partial test display tasks, because the essential logic of decision making is completely different. In partial test display trials, participants would not answer "change absent" in all trials that do not contain a change, nor "change present" in 50% of all trials that do contain a change. However, another procedure, different than Alvarez & Cavanagh’s (2004), can be derived from the percentage of correct answers. That method – the best way of visual working memory capacity assessment, that can be applied in various experimental designs, will be presented in the next chapter.

4.3.3.2 The best method of visual working memory capacity estimation

In the following simulation, the relationship of the percentage correct, and the number of items held in memory is going to be demonstrated.

Partial test display:

1. percentage correct = 50% → visual working memory capacity = 0

Partial and full test display

If participants do not memorize any object, then they will guess on every trial. Each guess would have 50% chance to be correct, so the overall percentage correct would be 50%. Hit rate would be 0,5, and false alarm rate would be 0,5. Pashler's (1988) capacity estimation would equal 0:
\[ C = \frac{IP \cdot (H - FA)}{1 - FA} \]

\[ C = \frac{IP \cdot (0,5 - 0,5)}{1 - 0,5} \]

\[ C = 0 \]

2. percentage correct = 100% → visual working memory capacity = set size

Partial and full test display

If participants memorize all objects, then they will answer correctly on every trial. Their percentage correct would equal 100%, hit rate would equal 1, and false alarm rate would equal 0. According to Pashler's (1988) formula, same conclusion is reached: memory capacity equals set size:

\[ C = \frac{IP \cdot (H - FA)}{1 - FA} \]

\[ C = \frac{IP \cdot (1 - 0)}{1 - 0} \]

\[ C = IP \]

3. percentage correct = 75% → visual working memory capacity = \( \frac{1}{2} \) set size

Partial test display

If participants manage to store half of all initially presented objects, then they will answer correctly in half of all trials. In the other half of the trials, they will guess, and each guess will have 50% chance to be correct. Therefore, overall percentage of correct answers would be 75%, hit rate would be 0,75, and false alarm rate would be 0,25. According to Pashler's (1988) formula, participants' visual working memory capacity would equal two thirds of set size:
If participants manage to store half of all initially presented objects, then they will answer correctly in half of all trials that do contain a change. In all the trials that do not contain a change, they would answer "change absent". So, the overall percentage correct would be 75%, hit rate would be 0.5, and false alarm rate would be 0. According to Pashler's (1988) formula, visual working memory capacity would equal half of set size:

\[ C = \frac{IP*(H - FA)}{1 - FA} \]

\[ C = \frac{IP*(0.75 - 0.25)}{1 - 0.25} \]

\[ C = \frac{2}{3} IP \]

*Full test display*

If participants manage to store half of all initially presented objects, then they will answer correctly in half of all trials that do contain a change. In all the trials that do not contain a change, they would answer "change absent". So, the overall percentage correct would be 75%, hit rate would be 0.5, and false alarm rate would be 0. According to Pashler's (1988) formula, visual working memory capacity would equal half of set size:

\[ C = \frac{IP*(H - FA)}{1 - FA} \]

\[ C = \frac{IP*(0.5 - 0)}{1 - 0.5} \]

\[ C = \frac{1}{2} IP \]

Therefore, under these assumptions, Pashler's (1988) assessment is correct.

But however, in some proportion of the trials participants would surely guess. Only in half of the trials that do contain a change participants would not have to guess, while in some of the remaining 75% of all the trials, participants would guess. Since in Vogel et al.'s (2001), Wheeler & Treisman's (2002) and in the present study it was found that the overall percentage of correct answers for this kind of stimuli is approximately equal in full and in partial test display condition, that means that the number of trial containing a change in which participants are guessing would be approximately the same as the number of trials without a change in which participants are guessing. Otherwise, the percentage of correct answers would differ between full
and partial test display condition. For example, if participants guess "change absent" in 15% of all trials that contain a change, then to obtain 75% of correct answers, they would have to guess "change present" in 15% of all trials that do not contain a change. In that case, hit rate would be 0,65, and false alarm rate would be 0,15, while percentage correct would remain 75%. In that case, Pashler’s (1988) visual working memory capacity estimate would equal 59% of set size, which is higher than the true visual working memory capacity:

\[ C = \frac{IP^*(H - FA)}{1 - FA} \]

\[ C = \frac{IP^*(0,65 - 0,15)}{1 - 0,15} \]

\[ C = 0,59IP \]

When the proportion of the trials in which participants are guessing is incorporated into Pashler's (1988) formula, then:

\[ C = \frac{IP^*(H - FA)}{1 - FA} \]

\[ C = \frac{IP^*[X + H - (X + FA)]}{1 - (X + FA)} \]

\[ C = \frac{IP^*(X + H - X - FA)}{1 - X - FA} \]

\[ C = \frac{IP^*(H - FA)}{1 - X - FA} \]

\( X \) is the proportion of the trials that contain a change on which participants are guessing "change absent", and at the same time it is the proportion of the trials that do not contain a change, in which participants are guessing "change present". \( X \) can not
be higher then 0.5, because participants would then obtain hit rate higher then 1 (without any guessing they already have hit rate 0.5).

So, Pashler’s (1988) estimate is correct only if there is no guessing at all. If participants guess, then capacity estimated by Pashler’s (1988) method is higher then the true visual working memory capacity. The higher the guessing rate is, the higher the memory capacity overestimation is:

\[
C = \frac{IP \cdot (0.5 - 0)}{1 - X - 0}
\]

\[
C = \frac{0.5 \cdot IP}{1 - X}
\]

In the example above, \(X = 0.15\), thus: \(C = 0.59\).

4. percentage correct = 62.5% \(\rightarrow\) visual working memory capacity = \(\frac{1}{4}\) set size

*Partial test display*

If participants manage to store one quarter of all initially presented objects, then they will have 100% percentage correct in one quarter of all trials. In the other three quarters of trials, they will guess, and each guess will have 50% chance to be correct. Therefore, the overall percentage of correct answers would be 62.5%, hit rate would be 0.625 (\(H = 1\) in \(\frac{1}{4}\) of the trials containing a change, and \(H = 0.5\) in the other \(\frac{3}{4}\) of the trials containing a change); false alarm rate would be 0.375 (\(FA = 0\) in \(\frac{1}{4}\) of the trials without a change, and \(FA = 0.5\) in the other \(\frac{3}{4}\) of the trials without a change. According to Pashler's (1988) formula, participants' visual working memory capacity would equal two fifths of set size:

\[
C = \frac{IP \cdot (H - FA)}{1 - FA}
\]

\[
C = \frac{IP \cdot (0.625 - 0.375)}{1 - 0.375}
\]
\[ C = \frac{2}{5} IP \]

**Full test display**

If participants manage to store one quarter of all initially presented objects, and if they do not guess at all, they will answer correctly in one quarter of all the trials that do contain a change. In all the trials that do not contain a change, they will answer "change absent". So, the overall percentage correct would be 62.5%, hit rate would be 0.25, and false alarm rate would be 0. According to Pashler's (1988) formula, visual working memory capacity would equal one quarter of the set size (if participants do not guess at all):

\[ C = \frac{IP \times (H - FA)}{1 - FA} \]

\[ C = \frac{IP \times (0.25 - 0)}{1 - 0} \]

\[ C = \frac{1}{4} IP \]

If participants do guess, and if the percentage of correct answers is held constant, then Pashler's (1988) estimate of visual working memory capacity would be higher than one quarter of the set size:

\[ C = \frac{IP \times (H - FA)}{1 - X - FA} \]

\[ C = \frac{IP \times (0.25 - 0)}{1 - X - 0} \]

\[ C = \frac{0.25 \times IP}{1 - X} \]

Guessing rate cannot be higher than 0.75, because then hit rate would be higher than 1.
5. percentage correct = 87.5% → visual working memory capacity = ¼ set size

Partial test display

If participants manage to store three quarters of all initially presented objects, then they will have 100% percentage correct in three quarters of all trials. In the other quarter of the trials, they will guess, and each guess will have 50% chance to be correct. Therefore, the overall percentage of correct answers would be 87.5%, hit rate would be 0.875 (H = 1 in ¼ of the trials containing a change, and H = 0.5 in the other ¼ of the trials containing a change); false alarm rate would be 0.125 (FA = 0 in ¼ of the trials without a change, and FA = 0.5 in the other ¼ of the trials without a change. According to Pashler's (1988) formula, participants visual working memory capacity would equal 85.7% of set size:

\[
C = \frac{IP\ast(H - FA)}{1 - FA}
\]

\[
C = \frac{IP\ast(0.875 - 0.125)}{1 - 0.125}
\]

\[
C = 0.857 IP
\]

Full test display

If participants manage to store three quarters of all initially presented objects, and if they do not guess at all, they will answer correctly in three quarters of all the trials that do contain a change. In all the trials that do not contain a change, they will answer "change absent". So, the overall percentage correct would be 87.5%, hit rate would be 0.75, and false alarm rate would be 0. According to Pashler's (1988)
formula, visual working memory capacity would equal three quarters of the set size (if participants do not guess at all):

$$C = \frac{IP \times (H - FA)}{1 - FA}$$

$$C = \frac{IP \times (0.75 - 0)}{1 - 0}$$

$$C = \frac{3}{4} IP$$

If participants do guess, and if percentage of correct answers is held constant, then Pashler's (1988) estimate of visual working memory capacity would be higher then three quarters of the set size:

$$C = \frac{IP \times (H - FA)}{1 - X - FA}$$

$$C = \frac{IP \times (0.75 - 0)}{1 - X - 0}$$

$$C = \frac{0.75 \times IP}{1 - X}.$$  

Guessing rate $X$ can not be higher then 0.25, because then hit rate would be higher then 1.

All these results are summarized in Table 5.

Two regularities are observed from the 5 situations above:

a) Pashler's (1988) formula overestimates visual working memory capacity. (The only exceptions are situations in full test display condition when participants are not guessing at all. These situations are extremely rare. For example, in the present study, all participants were guessing in some proportion of the trials, in both test display conditions. Besides that, Pashler's (1988) estimates are correct if participants memorized all stimuli or none of them.)
b) The relationship of true capacity and the percentage of correct answers is linear.

Table 5. Percentages correct and Pashler's (1988) memory capacity estimates, for different true memory capacities.

<table>
<thead>
<tr>
<th>True capacity</th>
<th>Full test display condition</th>
<th>Partial test display condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pashler's C</td>
<td>% correct</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>25% of set size</td>
<td>25% of set size or higher</td>
<td>62,5</td>
</tr>
<tr>
<td>50% of set size</td>
<td>50% of set size or higher</td>
<td>75</td>
</tr>
<tr>
<td>75% of set size</td>
<td>75% of set or higher</td>
<td>87,5</td>
</tr>
<tr>
<td>set size</td>
<td>set size</td>
<td>100</td>
</tr>
</tbody>
</table>

Therefore, visual working memory capacity can be correctly estimated via percentage of correct answers. True visual working memory capacity equals:

\[ C_{%} = \frac{IP*2*(PC-50)}{100} \]

\[ C_{%} = \frac{IP*(PC-50)}{50} \]

where:
When the results of the present study are inputed into this formula, then visual working memory capacities can easily be computed. In Table 6, capacities obtained by this method are compared to capacities estimated by Pashler's (1988) formula. It is found that in all experimental conditions, Pashler's (1988) capacity estimates are higher then the estimates obtained via percentages correct. Pashler's overestimations range between 0.71 and 2.12 units.

Table 6. Capacities estimated via percentage of correct answers and via Pashler's (1988) formula.

<table>
<thead>
<tr>
<th>Type of presentation and test display (experimental condition)</th>
<th>Capacity estimated by Pashler's method</th>
<th>Capacity estimated via percentages correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Simultaneous presentation and full test display</td>
<td>5.01</td>
<td>1.92</td>
</tr>
<tr>
<td>Simultaneous presentation and partial test display</td>
<td>5.90</td>
<td>1.04</td>
</tr>
<tr>
<td>Successive presentation and full test display</td>
<td>4.64</td>
<td>1.04</td>
</tr>
<tr>
<td>Successive presentation and partial test display</td>
<td>5.52</td>
<td>1.13</td>
</tr>
</tbody>
</table>

When capacity estimates via percentage correct are entered into ANOVA as dependent variable, the results are the same as in ANOVA with percentages correct as dependent variable. That is logical since this method of capacity estimation linearly transforms percentages correct. The main effect of the type of presentation is not significant, $F (1, 34) = .40; p > .05$. The main effect of the type of test display is not
significant, F (1, 34) = 3.35; p > .05, and there is no interaction between the type of presentation and the type of test display, F (1, 34) = .37; p > .05. These results are shown in Figure 24.

Figure 24. Visual working memory capacity estimated via percentage correct, as a function of the type of initial presentation and the type of test display.

This method of visual working memory assessment is better than Pashler's (1988), because it neither overestimates the true memory capacity nor it is biased by the criterion of responding, and also it does not give nonsensical estimates. The new procedure is better than Alvarez & Cavanagh's (2004) method, because it does not require the manipulation of set size, so therefore it does not require complex experimental designs, and also because it is valid for partial test display condition. In experiments that include set size manipulation, this method can be used to calculate memory capacities separately for different set sizes (higher than 4 for this kind of stimuli). The mean value can then be calculated, and that value would be more reliable estimate of visual working memory capacity. In the present study, this new procedure of visual working memory capacity assessment is shown to be valid for
simple stimuli – defined by one feature. For multi-feature stimuli, the situation is much more complex, since participants can memorize some objects partially. Also, the percentage of correct answers for multifeature objects differs as a function of test display (Wheeler & Treisman, 2002), and that is another reason why this method of visual working memory capacity estimation can not be applied to that kind of stimuli. As multifeature objects were not the topic of this research, that will be considered in some other study.
5 CONCLUSION

It is found that, compared to simultaneous presentation, successive presentation of stimuli had no positive effects on participants' performance. That means that in previous experiments (e.g. Vogel et al., 2001; Wheeler & Treisman, 2002; Alvarez & Cavanagh, 2004), visual working capacity was not underestimated due to inadequate encoding of stimuli. Although adequate encoding of each stimulus was assured in condition with successive initial presentation, performance in that condition was equally good as in the simultaneous presentation condition. Consequently, that result corroborates that classical change detection paradigm is a valid tool for assessing visual working memory capacity.

The effect of test display type on participants' performance was doubtful. When it was measured by the percentage of correct answers or by reaction time, it was not significant, but when Pashler's (1988) capacity estimates were analysed as dependent measure, then the effect of the type of test display type was significant: analysis showed that, compared to full test display condition, in partial test display condition participants had memorized more stimuli.

It was demonstrated that the change in criterion of responding is the cause of these noncongruent results. Through simulations, it was shown that Pashler's (1988) procedure of visual working memory assessment is heavily affected by the criterion of responding. Specifically, Pashler's (1988) method is shown to be valid only if participants do not make false alarms at all. The higher hit rate and false alarm rate are, the higher the overestimation of Pashler's (1988) assessment is, even when the percentage of correct answers is held constant. In other words, in situations in which participants do not make many false alarms (e.g. \( FA = 0.1 \)) and many hits (e.g. \( H = 0.6 \)), their working memory capacity assessed by Pashler's (1988) method is lower, compared to situations in which they make many false alarms (e.g. \( FA = 0.4 \)) and lots of hits (e.g. \( H = 0.9 \)).

To check this theoretical conclusion, false alarm rates and hit rates from the present study were analysed, and analyses have shown that the inference is valid. Participants had different criteria of responding in partial and in full test display condition: they made significantly more false alarms and significantly more hits in the partial test display condition, compared to the full test display condition. Exactly
because of that, Pashler's (1988) capacity estimates differed as a function of test display type, although true capacities and percentages of correct answers were approximately the same in these two conditions.

In the next theoretical simulation, another drawback of Pashler's (1988) memory capacity estimation procedure was found: in addition to depending on responding criterion, it tends to overestimate true visual working memory capacity. Also, it was demonstrated that Pashler's (1988) method sometimes gives nonsensical estimates (e.g. if participants make at least one correct rejection, and answer "change present" in all other trials, then by that procedure, visual working capacity would be estimated to equal 16 items, if set size is 16).

Solvation of conflicting results regarding the effect of test display type, led to another enigma: Why did participants use different responding criteria in full and in partial test display condition? That question was solved through another thorough simulation of participants' reasoning and decision making, the output of which was similar to the pattern of all results obtained in the present study.

Finally, the new method of visual working memory capacity assessment was constructed. In contrast to Pashler's (1988) which is based on hit rates and false alarm rates, this procedure estimates visual working capacity via percentages of correct answers:

\[ C_{\%} = \frac{IP \times (PC - 50)}{50} \]

The new procedure is better than Pashler's (1988) hence it does not overestimate true memory capacity and it is not at all affected by the criterion of responding. Also, it never gives nonsensical estimates. Although Alvarez & Cavanagh (2004) were one of the first authors that estimated visual working memory capacity via percentages of correct answers, their method was constructed only for the purpose of their specific experiment. The method constructed in the present study is better than Alvarez & Cavanagh's (2004) procedure because it can be applied in all sorts of experimental designs in the field of visual working memory capacity research - in contrast to the procedure of Alvarez & Cavanagh (2004) it does not require the use of uneconomical experimental designs.
At the end, the new method of visual working memory capacity assessment was compared to Pashler's (1988) method (since it requires the manipulation of set size, the procedure of Alvarez & Cavanagh (2004) could not be compared to these two methods). When data obtained in the present study was subjected to analyses, Pashler's (1988) visual working memory capacity estimates ranged between 4.64 and 5.90, while the estimates of the new method ranged from 3.60 to 4.30. Therefore, it can be concluded that visual working memory capacity is indeed limited to not more than 4 objects.
6 REFERENCES


Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science, 8*, 368-373.


7 POVZETEK

7.1 Uvod


spremembe, ne more priti do zadetka, tudi ne more priti do lažnega preplaha v poskusih, ki vsebujejo spremembe. Logika Pashlerjeve (1988) formule je naslednja:
Ob domnevi, da so udeleženci obdržali določeno količino postavk v spominu (C) sledi zadetek, ko se ena od njih spremeni, če spremenjena postavka ni bila shranjena v spomin, sledi pogrešek. Kadar udeleženci niso opazili spremembe, bi odgovorili bi „sprememba odsotna“, toda upoštevati moramo, da so v določenem proporcu poskusov ugibali. Ugibanje se nanaša na odgovore „sprememba prisotna“ v primerih ko dejansko ni bilo nobene spremembe (zmotni preplahi). V skladu z omenjenimi spremenljivkami in nastavljeno velikostjo, je Pashler (1988) podal formulo za ugotavljanje zmogljivosti vizualnega delovnega spomina (C):

\[ H = \frac{C}{IP} + \frac{IP - C}{IP} * FA, \]

Kjer
je C število podanih objektov shranjenih v spominu (zmogljivost vizualnega delovnega spomina);
(C/ IP) je propor poskusov v katerih se je postavka shranjena v spominu spremenila;
(IP/ C) je število podanih objektov, ki niso shranjeni v spominu;
(IP - C) / IP je propor poskusov v katerih se je postavka, ki ni bila shranjena v spominu, spremenila
FA je stopnja uganjevanja (oceni se iz stopnje zmotnih preplahov – iz proporca poskusov brez spremembe v katerih je udeleženec menil, da je prišlo do spremembe)

Ko zgornjo enačbo pretvorimo, lahko izločimo formulo za oceno zmogljivosti vizualnega delovnega spomina:

\[ C = \frac{IP*(H - FA)}{1 - FA} \]

Da bi zagotovili, da ocena zmogljivosti vizualnega delovnega spomina ne bo kontaminirana s prispevkom besednega delovnega spomina, so v teh poskusih uporabili tri vrste postopkov:

a) dražljaji ki jih je težko poimenovati

b) kratka predstavitev dražljajev
   Ker je glasovno kodiranje počasnejše od vizualnega kodiranja, zelo kratka predstavitev vizualnega građiva zmanjša prispevek besednega delovnega spomina.

c) besedna obremenitev
   Ta postopek je znan tudi kot slušna ali govorna potlačitev in se nanaša na brezpomenski ponavljavači se govor, ki zmoti slušno kodiranje in nima skoraj nobenih drugih učinkov. Ker ne škoduje učinkovitosti pri nalogah, ki potrebujejo vizualno kodiranje (npr. Murray, 1967; Klapp, Marshburn, & Lester, 1983; Richardson, 1984; Baddeley, 1986), ta postopek omogoča testiranje vizualnega delovnega spomina z daljšim trajanjem prikazovanja. (Frick, 1985).

Običajno se uporabita dva od teh treh postopkov zaradi zmanjševanje kontaminacije z besednim delovnim spominom.

Čeprav obstajajo razprave o naravi vizualno informacijske hrambe in vzdrževanja, vsi ugledni raziskovalci soglašajo, da je vizualni delovni spomin omejen na ne več kot 4 - 5 postavk.

Nekatera od njihovih študij vsebujejo nekaj kontrolnih poskusov (npr. Vogel et al., 2001; Wheeler & Treisman, 2002), da bi ugotovili, če je majhna opažena zmogljivost posledica omejitev vizualnega delovnega spomina, ali pa je rezultat nekaterih drugih procesov.

Ker lahko prispevek besednega delovnega spomina povzroči precenjenost realne zmogljivosti vizualnega delovnega spomina, nakopičene odločitvene napake lahko vodijo do njegove podcenjenosti. Posebno pri nalogah odkrivanja sprememb z velikimi množicami, morajo udeleženci primerjati vse spominske reprezentacije.
začetnega prikaza z vsemi spominskimi reprezentacijami testnega prikaza. Če prikaz vsebuje 12 postavk, potem morajo sprejeti 12 odločitev „enako-različno“ in vsaka taka odločitev je podvržena določeni stopnji napake. Da bi preverili če je slaba storitev v nalogah odkrivanja sprememb artefakt teh napak ali pa posledica omejitev vizualnega delovnega spomina, je bilo narejenih več kontrolnih poskusov. V teh poskusih so storitev na klasičnih nalogah odkrivanja sprememb, primerjali s storitvijo v pogojih, ko je moral a sprejeta zgolj ena odločitev. V enem takih eksperimentalnih pogojih, je bila postavka, ki bi se lahko spremenila, označena v testnem prikazu, udeleženci pa so bili seznanjeni, da do spremembe ni prišlo na nobeni drugi postavki razen na označeni. Storitev je bila pod tem pogojem enaka kot v pogoju brez znakov, toda zaradi domneve, da je sama prisotnost neoznačenih postavk lahko škodila storitvi s sprožitvijo samodejnih odločitvenih procesov, je bil narejen še en kontrolni poskus. Da bi raziskali to možnost, so v poskusnem pogoju, po retencijskem intervalu, prikazali samo eno postavko. Ker popolna odstranitev nesprenjenih postavk moti pomnilniško storitev (Klemer, 1963; Vogel et al., 2001), so bili očrtani kvadrati (rezerviran prostor) postavljeni na položaje nerelevantnih dražljajev. Toda, tudi v tem pogoju, v katerem so morali udeleženci sprejeti le eno odločitev „enako-različno“, storitev ni bila nič boljša, kot v klasičnih nalogah odkrivanja sprememb z celotnimi testnimi prikazi. Zato so menili, da vizualni delovni spomin za objekte z eno lastnostjo, izmerjen v nalogi odkrivanja sprememb, ni podcenjen zaradi nakopičenih odločitvenih napak (Vogel et al., 2001; Wheeler & Treisman, 2002).

Ob nakopičenih odločitvenih napakah, so bile omejitve vkodiranja, drug možen vzrok podcenjenosti zmogljivosti vizualnega delovnega spomina v nalogah odkrivanja sprememb. Ker je bila v skoraj vseh poskusih začetna predstavitev dražljajev tahistoskopska (da bi preprečili besedno vkodiranje dražljajev), je možno, da so tako kratki predstavitveni intervali (navadno trajajo od 100 do 150 ms) prekratki za ustrezno vkodiranje vseh dražljajev, predvsem pri velikih množicah. Možnost, da so take omejitve vzrok podcenjenjevanja v ocenjevanju zmogljivost vizualnega delovnega spomina je bila ovržena po dveh poskusih. V prvem so bile uporabljene vizualne iskalne naloge. Prikazan je bil obarvani kvadrat, po 500-ms časovnem zamiku pa je bila za 100 ms prikazana preglednica s 4, 8 ali 12 obarvanimi kvadrati. Udeleženci so morali odgovoriti, ali preglednica vsebuje ciljno barvo ali ne. Točnost odgovorov je bila blizu 100 %, celo z največjimi množicami, tako, da so menili da je
100 ms dovolj dolg interval, da udeleženci lahko prepoznajo barve vseh kvadratov (Vogel et al., 2001). V drugem poskusu je bil začetni predstavitveni interval podaljšan na 500 ms, storitev udeleženca pa so primerjali s pogojem, ko je začetni prikaz trajal 100 ms, medtem ko so bile vse ostale spremenljivke ohranjene nespremenjene. Čeprav je bila udeleženčeva storitev približno enaka v obeh pogojih, so menili, da je 100 ms interval dovolj dolg za ustrezen vkodiranje do 12 postavk (Vogel et al., 2001).

Tako rezultati vseh teh poskusov kažejo, da znaša zmogljivost vizualnega delovnega spomina za eno-lastnostne objekte približno 4-5 postavk (ali manj, odvisno od vrste gradiva, ki si ga je potrebno zapomniti). Prav tako je bilo ugotovljeno, da ocena ni niti precenjena zaradi vpliva besednega delovnega spomina, niti podcenjena zaradi nakopičenih odločitvenih napak, omejitev vkodiranja ali česarkoli drugega.

To prepričanje smo preverjali v pričujoči raziskavi. V dosedanjih raziskavah so spremnjeni različne vrste spremenljivk (npr. nastavljena velikost, vrste testnih prikazov, vrsto in zapletenost vizualnega gradiha, trajanje meddražljajskega intervala itn.), vendar so bili vsi poskusi zasnovani na enaki predlogi:

a) kratka simultana ekspozicija dražljajev
b) retencijski interval med katerim lahko pride do spremembe
c) preizkus.

Ker so vsa relevantna raziskovanja obnašanja temeljila klasični paradigmi detekcije sprememb, je bil cilj našega poskusa preizkusiti spomin za preproste eno-lastnostne predmete z drugačnim postopkom.

Kodirne omejitve so bile v žarišču sedanje raziskave. Čeprav rezultati več kontrolnih poskusov (ki so bili že opisani) tako kažejojo, so bili ti dokazi vendarle nezadostni za sklep, da ocene zmogljivosti vizualnega delovnega spomina niso bile podcenjene zaradi kodirnih omejitev. Nikoli se nisop ukvarjale s zelo pomembnim spornim vprašanjem. Predvsem je pozornost ključna za kodiranje vizualnega gradiha. Če ni usmerjena k določenemu dražljaju, potem obstaja zelo majhna verjetnost, da bo ta dražljaj vkodiran v spomin (e.g. Coltheart, 1980; Irwin, 1991; Rensink, O'Regan, & Clark, 1997). Drugo, pozornost ne more biti hkrati usmerjena na več kot 4 predmete (naprimer Pylyshyn & Storm, 1988; Yantis, 1992) in ker si več kot štiri objekte ne moremo zapomniti, je to lahko bodisi izredno naključje ali pa pomeni, da udeleženci ne morejo ohraniti več kot štiri objekte v svojem vizualnem delovnem spominu,
preprosto zato ker ti predmeti niso bili nikoli vkodirani tu, ker so bili sočasno podani v vseh teh poskusih. Osnovni cilj naše raziskave je bil odgovoriti na to vprašanje.

Da bi to naredili, smo pripravili eksperimentalni načrt v katerem smo manipulirali z vrsto začetne predstavitve dražljajev. Udeleženčevmo storitev v klasični nalogi odkrivanja sprememb smo primerjali s storitvijo preoblikovane naloge odkrivanja sprememb z zaporedno začetno predstavitvijo dražljajev.
7.2 Metoda

V eksperimentu z 2 x 2 mešanim načrtom z vrsto začetne podaje dražljajev (sočasen / zaporeden) kot medsubjektnim faktorjem in vrsto preizkusnega prikaza (polni / delen) kot znotrajsubjektnim faktorjem, je sodelovalo šestnintrideset udeležencev. Vsak udeleženec je opravil 60 poskusov v katerih so bili kvadrati, določeni le z barvo, uporabljeni kot dražljaji.

Vsi relevantni dejavniki poleg vrste začetne predstavitve in vrste testnega prikaza, so bili konstantni:

a) nastavljene velikosti so niso spreminjale – vsi poskusi so vsebovali 8 kvadratov

b) celoten interval začetne predstavitve je bil vedno 500 ms

c) barve in lokacije stimulansov so bile za vsak poskus izbrane naključno, vendar tako, da so bile naloge vedno enako zahtevne v vseh pogojih, istih 60 poskusov je bilo uporabljenih v vsakem eksperimentalnem pogoju (le njihov vrstni red se je naključno spreminjal z udeleženci)

d) besedni postopek obremenitve je bil enak v vseh pogojih

Zagotovljeno je bilo torej, da so bile razlike v storitvi udeležencev med eksperimentalnimi pogojmi, lahko pripisane le vplivu neodvisnih spremenljivk.

Glede vrste začetne predstavitve dražljajev v pogoju sočasne predstavitve je bil prikazan celoten prikaz hkrati, v trajanju 500 ms, medtem ko so se v pogoju zaporedne predstavitve prikazi razvijali od leve proti desni strani v petih 100 ms predstavitvenih intervalih. V prvem od njih je bil prikazan le prvi stolpec; v drugem intervalu, sta bila vidna prvi in drugi stolpec, in tako naprej, vse do zadnjega 100 ms intervala, v katerem je bila razvidna celotna slika (ista kot pri celotnem pogoju sočasne predstavitve). Tako je bil celoten predstavitveni čas enak v obeh pogojih – 500 ms.

Glede vrste testnega prikaza, je bilo v pogoju polnega testnega prikaza podanih vseh osem kvadratov, medtem ko je bil v pogoju delnega testnega prikaza podan le en pobarvan kvadrat v testni ureditvi (v 50 % poskusov je barva tega kvadrata enaka kot barva kvadrata, kije bil prvotno prikazan na tej lokaciji). Drugi kvadrati so bili zamenjani z rezerviranim mestom – črnim obrisom kvadrata enake velikosti, ki je bil postavljen natančno na mesto začetno prikazanih kvadratov.

Da bi zmanjšali prispevek besednega delovnega spomina, smo poleg kratke predstavitve dražljajev, uporabili še postopek besedne obremenitvene naloge: udeleženci so morali glasno ponoviti frazo Coca-cola med obdobji začetnega prikaza, spominskega odloga in testnega prikaza.

Raziskovalna metoda je prikazana na Slikah 16 in 17.
7.3 Rezultati in razprava

Udeleženčeva storitev je bila merjena z odstotkom pravilnih odgovorov, s Pashlerjevo (1988) oceno zmogljivost in z reakcijskim časom. Mešana dvosmerna ANOVA je bila opravljena posebej za vsako od treh odvisnih spremenljivk z vrsto začetne predstavitve kot medsubjektnim faktorjem in vrsto testnega prikaza kot znotraj subjektnim faktorjem.

Analiz so pokazale, da glavni učinek vrste začetne predstavitve ni bil pomemben. Tudi interakcija med vrsto začetne predstavitve in vrsto testnega prikaza ni bila pomembna. Ta ugotovitev ne potrjuje domneve, po kateri so bile ocene zmogljivosti vizualnega delovnega spomina dobljene v prejšnjih raziskavah podcenjene zaradi vkodirnih omejitev. Prav nasprotno, ker se je zaporeni začetni prikaz (v katerem je bilo vsak dražljaj zagotovljeno, da bo nedvomno vkodiran v vizualni delovni spomin) izšel brez prednosti v primerjavi s sočasnimi začetnimi prikazi dražljajev, ta rezultat kaže, da je v klasičnih nalogah odkrivanja sprememb vizualno građivo ustrezno vkodiran, tako kot so tobirato Vogel in sod. (2001) skupaj z drugimi raziskovalci.


a) nastavljena velikost je enaka osmim postavkam

b) zmogljivost vizualnega delovnega spomina je enaka štirimi postavkam.

V teh simulacijah se je kriterij odgovarjanja spreminjal s pomočjo navodil udeležencem:

a) v prvi simulaciji so udeleženci prejeli navodilo, da naj težijo k visoki stopnji zadetkov, ne glede na delež zmotnih preplahov – da bi se izognili pogreškom
Jim je bilo rečeno da naj bodo pozorni in si zapomnijo le dve od osmih prikazanih postavk (rečeno jim je bilo naj odgovorijo “sprememba prisotna” v vseh poskusih v katerim so bili negotovi, tudi zaradi zagotovitve, da bi se izognili pogreškom);

b) v drugi simulaciji, so udeleženci prejeli navodilo, da dosežejo visoko stopnjo zmotnih preplahov, ne glede na delež zadetkov – da bi se izognili zmotnim preplahom jim je bilo rečeno, naj bodo pozorni in si zapomnijo le dve od osmih prikazanih postavk (rečeno jim je bilo, da v vseh poskusih v katerima so bili negotovi odgovorijo “sprememba odsotna”, tudi kot zagotovilo, da bi se izognili zmotnim preplahom);

c) v zadnji simulaciji, niso udeleženci prejeli nobenega navodila glede kriterija odgovarjanja.

Stopnje zadetkov, pogerškov, zmotnih preplahov in pravilnih zavrnitev, skupaj z odstopom pravilnih odgovorov in Pashlerjeve (1988) ocene zmogljivosti, so predstavljali iznos simulacij, in ti iznosi kažejo, da Pashlerjev (1988) postopek ocene zmogljivosti vizualnega delovnega spomina:

a) teži k precenjevanju resnične zmogljivosti vizualnega delovnega spomina ter

b) je odvisen od kriterija odgovarjanja.

Če imajo udeleženci veliko zadetkov in veliko zmotnih preplahov (blag kriterij: H=1; FA= 0,75), potem je zmogljivost spomina ocenjena z Pashlerjevim (1988) postopkom višja (C = 8 postavk), v primerjavi s situacijo v kateri imajo udeleženci manj zadetkov in manj zmotnih preplahov (strog kriterij: H = 0,25; C = 2 postavke), kljub temu, da je H – FA konstanten (H – FA = 0,25) in kljub dejstvu, da je v obeh simulacijah resnična zmogljivost vizualnega delovnega spomina dejansko znašala le 2 postavki. V realistični simulaciji (brez navodila) je bilo pokazano, da znaša Pashlerjeva (1988) ocena zmogljivosti vizualnega delovnega spomina 5,33 postavk, medtem ko je znaša resnična zmogljivost 4 postavke.

Da bi ugotovili, če so v tej raziskavi, udeleženci spremenili svoj kriterij odgovarjanja med polnim in delnim testnim pogojem, stabilni stopnja zadetkov in stopnja zmotnih preplahov vneseni v dve posebni mešani dvostranski ANOVA-i, z vrsto začetnega prikaza in vrsto testnega prikaza kot faktorji. Analize so pokazale, da sta bili stopnja zadetkov in stopnja zmotnih pšreplahov pomembno višji v pogoju delnega testnega prikaza, v primerjavi s pogojem polnega testnega prikaza. Tako udeleženci niso uporabili istega kriterija odgovarjanja, in to je razlog zakaj so bile

Vendar ta rešitev vzroka nasprotujočih si rezultatov glede na vpliv vrste testnega prikaza zastavlja še eno uganko: Zakaj so udeleženci uporabili različne kriterije odgovarjanja v pogojh polnega in delnega testnega prikaza? Da bi odgovorili na to vprašanje, so bili opravljeni nizi podrobnih simulacij, v katerih so bili vsi možni izidi upoštevani ločeno za pogoja polnega in delnega testnega prikaza (glej Sliko 23):

a) izid 1: poln testni prikaz + sprememba odsotna + vse postavke prikazane na testnem prikazu;
b) izid 2: poln testni prikaz + sprememba prisotna + ena od postavk, ki je bila shranjena v spominu se je spremenila;
c) izid 3: poln testni prikaz + sprememba prisotna + ena od postavk, ki ni bila shranjena v spominu se je spremenila;
d) izid 4: delni testni prikaz + sprememba odsotna + ena od postavk, ki je bila shranjena v spominu je bila prikazana na testnem prikazu;
e) izid 5: delni testni prikaz + sprememba odsotna + ena od postavk, ki ni bila shranjena v spominu je prikazana na testnem prikazu;
f) izid 6: delni testni prikaz + sprememba prisotna + ena od postavk, ki je shranjena v spominu se je spremenila;
g) izid 7: delni testni prikaz + sprememba prisotna + ena od postavk, ki ni bila shranjena v spominu se je spremenila;

Iznose teh simulacij smo nato primerjali s podatki te raziskave in ugotovili, da je bil vzorec rezultatov podoben. V obeh, v iznosu simulacij in v rezultatih te raziskave, je razvidno da:

1. so odstotki pravilnih odgovorov približno enaki za poln in delen testni prikaz;
2. je $H - FA$ približno enak v pogoju polnega in delnega testnega prikaza in znaša približno 0,50;
3. je stopnja zadetkov višja v pogoju delnega testnega prikaza;
4. je stopnja zmotnega preplaha višja v pogoju delnega testnega prikaza;
5. je zmogljivost vizualnega delovnega spomina, ocenjena s Pashlerjevo (1988) metodo višja v pogoju delnega testnega prikaza.
Zato je s pomočjo teh simulacij pojasnjeno zakaj udeleženci spreminjajo svoje kriterije odgovarjanja med različnimi vrstami testnih prikazov, čeprav hranijo 4 objekte v pogojih polnega in delnega testnega prikaza (ena od domnev v vseh simulacijah je bila, da štiri postavke lahko shranimo in ohranimo v spominu, v obeh pogojih, pri polnem in pri delnem testnem prikazu), čeprav sta odstotek pravilnih odgovorov in \( H – FA \) ohranjena konstantna med tema pogojema. V teh simulacijah smo tudi pokazali, da Pashlerjeva (1988) formula teži k precenjevanju resnične zmogljivosti vizualnega delovnega spomina.


Boljši postopek za oceno zmogljivosti vizualnega delovnega spomina temelji prej na odstotku pravilnih odgovorov, kot na stopnji zadetkov in zmotnih preplahov. Alvarez in Cavanagh (2004) sta v svojem eksperimentu uporabila klasično paradigmo odkrivanja sprememb in spreminjala nastavljene velikosti od 1 do 15 objektov (i = 2). Da bi ocenili zmogljivost vizualnega delovnega spomina sta poiskala 75% pražna številko za vsakega udeleženca ter ga delila z dva, da bi dobila oceno zmogljivosti. V razlagi sta navedla, da če udeleženci shranijo polovico prikazanih objektov, na primer pri nastavljeni velikosti 8, potem bi dobili 75 % pravilnih odgovorov, ker bi opazili 50 % vseh sprememb in bi v vseh ostalih poskusih odgovorili z “sprememba odsotna”. Tako bi udeleženci dosegli 100 % pravilnih odgovorov v vseh poskusih v katerih ne pride do spremembe in 50 % pravilnih odgovorov v vseh poskusih, ki vsebujejo spremembo. Vendar ima tudi ta postopek nekaj slabosti:

1. ker je bil oblikovan posebej za tiste specifične ekspermente, zahteva kompleksne eksperimentalne načrte in se ne more uporabiti eksperimentih v katerih ni nastavljena velikost spremenljiva;
2. ne upošteva ugibanja;
3. ta postopek v svoji izviri obliki ni mogel biti uporabljen v eksperimentih z uporabo nalog delnega testnega prikaza, ker je
bistvena logika odločanja popolnoma drugačna (npr. v poskusih delnega testnega prikaza, udeleženci ne odgovorijo “sprememba odsotna” v vseh poskusih ki ne vsebujejo spremembe, niti “sprememba prisotna” v 50 % vseh poskusov, ki vsebujejo spremembo).


\[
C_v = \frac{IP \cdot (PC - 50)}{50}
\]

kjer je:

- \(C_v\) zmogljivost vizualnega delovnega spomina ocenjena z odstotkom pravilnih odgovorov
- \(PC\) odstotek pravilnih odgovorov
- \(IP\) nastavljena velikosti (število začetno prikazanih postavk)

Teoretično smo pokazali, da ta formula velja za različne nastavljene velikosti in za obe (polni in delni) vrsti testnega prikaza.


1. medtem ko Pashlerjevv (1988) postopek precenjuje zmogljivost vizualnega delovnega spomina, je ta postopek ne;
2. medtem ko Alvarezov in Cavanaghov postopek zahteva kompleksne eksperimentalne načrte (ki vključujejo sprememljive nastavljene velikosti), se ta postopek lahko uporabi v bolj gospodarnih eksperimentih, še vedno pa se lahko uporabi v eksperimentih v katerih...
se spreminja nastavljena velikost (v teh eksperimentih lahko izračunamo srednjo zmogljivost za različne nastavljene velikosti in to bi bila bolj zanesljiva ocena zmogljivosti);

3. medtem ko Pashlerjeva formula daje včasih nesmiselne ocene, jih naša formula ne daje;

4. medtem ko je Pashlerjev postopek pod močnim vplivom kriterija odgovarjanja, naš postopek ni;

5. medtem ko Alvarezov in Cavanaghov postopek ni veljaven ne za pogoj delnega in ne za pogoj polnega testnega prikaza, naš postopek je.


Novi postopek ocenjevanja zmogljivosti vizualnega delovnega spomina je veljaven za dražljaje, opredeljene z eno značilnostjo, medtem ko je za večznačilnostne dražljaje položaj veliko bolj zapleten. Toda, ker je raziskovanje omejitev zmogljivosti vizualnega delovnega spomina zelo pomembno za razumevanje višjih kognitivnih funkcij, lahko kmalu pričakujemo nastanek novega postopka ocenjevanja, ki bi lahko veljal za večznačilne objekte.
7.4 Sklep

Zaporen prikaz dražljajev ni imel nobenih pozitivnih učinkov na storitev udeležencev v primerjavi z sočasno predstavitvijo dražljajev. Zato v klasičnih nalogah odkrivanja sprememb zmogljivost vizualnega delovnega spomina ni podcenjena zaradi neustreznega vkodiranja dražljajev.


Končno smo oblikovali nov postopek ocenjevanja zmogljivosti vizualnega delovnega spomina. Ta postopek ocenjuje zmogljivost spomina na osnovi odstotka pravilnih odgovorov in ne iz stopnje zadetkov in lažnih preplahov:

\[ C_{\%} = \frac{IP^* (PC - 50)}{50} \]

Ko smo podatke iz naše raziskave vnesli v to formulo, smo odkrili, da znaša zmogljivost vizualnega delovnega spomina približno štiri postavke.

Nov postopek smo primerjali z drugimi običajnimi postopki ocene zmogljivosti vizualnega delovnega spomina in pokazali, da je boljši od ostalih, ker ne
precenjuje resnične zmogljivosti, ni pod vplivom kriterijev odgovaranja in se lahko uporablja v različnih eksperimentalnih načrtih.