MASARYK UNIVERSITY
Faculty of arts

Department of Psychology

Lenka Štěpánková

TESTING SPATIAL COGNITIVE ABILITY

Doctoral Dissertation

Dissertation supervisor: prof. PhDr. Tomáš Urbánek, Ph.D.
Brno 2019

Acknowledgements

First I would like to thank my supervisor prof. PhDr. Tomáš Urbánek, Ph.D. for his expertise, patience and support. Secondly, I would also like to thank all of the faculty members for their help, namely Mgr. Tatiana Malatincová, Ph.D. for the statistical consultations and PhDr. Katarína Millová, Ph.D. for a help with the recruitment of the participants, my IT consultant and programmer Ing. Jakub Bárta, who is responsible for the test battery and its design, Mr. Vít Kovalčík for the assistance with the calibration of monitors, my research assistants Mgr. Jan Horský, Mgr. Radim Chvaja and Kamila Vlčková and all the participants for their time.
Last but not least I would like to thank my family for their support and understanding.

This dissertation thesis was supported by the research infrastructure HUME Lab Experimental Humanities Laboratory, Faculty of Arts, Masaryk University.

Prohlašuji, že jsem disertační práci vypracovala samostatně
s využitím uvedených pramenů a literatury.

I declare that I have completed this dissertation independently,
and that all external sources of information have been properly cited.

....
Author's signature

Parts of this dissertation are based on previously published work of the author and a Master’s thesis of the author:

Krajčíková, L. (2017). (Ne)závislosť na poli: kognitívny štýl či schopnosť? *TESTFÓRUM: Časopis pro psychologickou diagnostiku, Praha: Českomoravská psychologická společnost, 10,* 4–13.

Krajčíková, L. (2014). Field independence as a spatial cognitive ability: Correlation study of verbal and spatial short term memory. Master’s Thesis, Masaryk University, Available online: https://is.muni.cz/th/dqjgb/Diplomova_prace_final.pdf?zpet = %2Fvyhledavani%2F %3Fsearch 3Dkraj %C4 %8D%C3%ADkov%C3%A1%20agenda:th%26start%3D1

Abstract

This study deals with the cognitive style of field dependence/independence described by Herman Witkin and colleagues in the 1950s. This cognitive style supposedly divides people into two groups, field dependent and field independent, according to the strategies they use in information processing and overall cognition. The strategies are supposed to be equally useful, with no clear advantage of either one. The presented work argues that the cognitive style of field dependence/independence is a part of spatial cognitive ability and tests created to measure this style are de facto tests of spatial ability and skills. The theoretical part of the thesis provides an overview of the literature about cognitive styles with an emphasis on field dependence/independence and tests created to measure this style. Spatial cognitive ability and its factors are also discussed. The end of the theoretical part is devoted to a study whose methodology was used as a control subtest, representing a conceptual replication of this study (Gilbert, Regier, Kay & Ivry, 2006). The empirical part examines the connection between tests created to measure field dependence/independence (Embedded Figures test and similar tests) and spatial cognitive ability and tries to find a connection between these tests and a test of mental rotations (standardised test of spatial ability). The results confirmed the hypotheses about a moderate correlation between the task based on the Embedded Figures test and mental rotations, which suggests a connection between field dependence/independence and spatial cognitive ability. The conceptual replication did not find the same results as the original study. This work's main application could be seen in the proposal to use the Embedded Figures test and tests based on it as tests for spatial ability and skills in entrance exams at universities and job interviews for the technical field.

Key words: spatial cognitive ability, field dependence/independence, cognitive style, mental rotations, universality of language categories, replication

Abstrakt

Predkladaná práca sa zaoberá kognitívnym štýlom závislosti/nezávislosti na poli popísanom v päťdesiatych rokoch Hermanom Witkinom a jeho kolegami. Tento kognitívny štýl predpokladá rozdelenie ľudí na závislých a nezávislých na poli na základe stratégií, ktoré používajú pri spracovávaní informácií a celkovo kognícií. Predpokladom definujúcim štýl je, že ani jedna z týchto stratégií by nemala byť výhodnejšia a užitočnejšia než druhá. U kognitívneho štýlu závislosti/nezávislosti na poli sa však zdá, že tento predpoklad neplatí. V uskutočnenej štúdii bolo snahou ukázať, že tento kognitívny štýl je v skutočnosti súčasťou priestorovej kognitívnej schopnosti a testy vytvorené na jeho meranie (hlavne Embedded Figures test a testy založené na podobnom princípe) môžu byť použité ako testy priestorovej schopnosti. V teoretickej časti je diskutovaná téma kognitívnych štýlov s dôrazom na závislosť/nezávislosť na poli, sú predstavené testy na meranie tohto štýlu, následne je spomenutá rozsiahla kritika štýlu závislosti/nezávislosti na poli. Hlavnou staťou teoretickej časti sú kapitoly venované priestorovej kognitívnej schopnosti a jej meraniu. Na záver teoretickej časti je popísaná metodológia konkrétnej štúdie (Gilbert, Regier, Kay & Ivry, 2006), ktorá bola použitá ako kontrolný subtest v empirickej časti a predstavuje konceptuálnu replikáciu originálnej štúdie. Empirická časť práce sa snaží objasniť vzťah medzi testom vytvoreným na základe Embedded Figures test a testom mentálnych rotácií, ktorý je štandardizovaným testom priestorovej schopnosti. Výsledky potvrdzujú hypotézu, ktorá predpokladá stredne silnú koreláciu medzi týmito testami, ktorá naznačuje, že závislosť/nezávislosť na poli je v skutočnosti súčasťou priestorovej kognitívnej schopnosti a testy založené na podobnom princípe ako Embedded Figures test, môžu byť použité ako testy priestorových schopností. Replikáciou štúdie Gilbert a kolegov (2006) neboli preukázané rovnaké výsledky, ako získala originálna štúdia. Za hlavnú aplikáciu výsledkov môže byť považovaný návrh na použite testov vytvorených na základe Embedded Figures test, ako testov priestorových schopností pri prijímacích skúškach na technicky zamerané vysoké školy a pri prijímacích pohovoroch do technicky orientovaných profesií.

Kľúčové slová: priestorová kognitívna schopnosť, závislosť/nezávislosť na poli, kognitívny štýl, mentálne rotácie, univerzálnosť jazykových kategórií, replikácia

Contens

INTRODUCTION

"*Indeed, many of those who have studied spatial abilities have […] reactions that range from amusement to annoyance.*"

Lohman, 1996, p. 3

This dissertation deals with cognitive styles. It offers a new point of view regarding the well-known and wildly debated cognitive style of field dependence/independence (further FD/FI). This style has been raising questions since the time it was described by Herman Witkin and colleagues in the 1950s. The authors divide individuals according to tests (such as the Embedded Figures test, further EFT) into two groups, a field dependent and a field independent group. This cognitive style is supposed to represent similar strategies for processing information. However, as it seems, this particular style is not, in fact, a style, but rather a cognitive ability. This thesis shows an alternative perspective of the cognitive "style" of field dependence/independence, and a potential new use of the Embedded Figures test as a test for spatial cognitive ability.

The theoretical part of the thesis provides a review of the literature on cognitive styles, describes field dependence/independence as well as tests created to measure this style, and provides an overview of the criticism of this style. As this thesis tries to find a link between FD/FI and spatial cognitive ability, spatial ability and its testing is also one of the main parts in the theoretical part of the thesis. The last chapter is dedicated to the Sapir-Whorf hypothesis and thoroughly describes Gilbert *et al.*'s (2006) study, which was replicated as a part of the empirical part of the thesis and used as a control subtest for the experiment.

The empirical part describes a test battery that consists of three subtests, a Hidden figures subtest, which is based on the original Embedded Figures test (which could not be used due to copyright reasons), Mental rotation subtest, which is a standardized test of spatial ability, and a Target position subtest, which is based on Gilbert's study and represents a control subtest. Results show a moderate correlation between participants' performance in the Mental rotation subtest and Hidden figures subtest, which suggest that the FD/FI cognitive style is a part of spatial cognitive ability. Likewise, tests based on EFT are tests of spatial cognitive ability. Also, the control subtest based on Gilbert's study showed no correlation with the rest of the battery.

The replication does not show similar results as the original study.

The application of the results is seen in the possibility of using the EFT–like tests as tests for spatial ability, for example, in entrance examinations for technical study fields, or in the job market.

The criticism of the FD/FI cognitive style is supported by the findings, thus suggesting to consider FD/FI a part of spatial cognitive ability rather than a style.

1 FIELD DEPENDENCE/INDEPENDENCE: COGNITIVE STYLE OR COGNITIVE ABILITY

1.1 Cognitive styles

To be able to define a cognitive style one should reach out to work of Kozhevnikov (2007), whose frequently cited article provides an overview of cognitive style research. The author sees cognitive style as *"[…] identifying individual differences in cognition that are stable, value-free and related to personality and social relationships*" (Kozhevnikov, 2007, p. 464).

Goodenough (1976) defines cognitive style as: "*[…] dimensions of individual differences involving the form of cognitive functioning, with expressions in a wide array of content areas including perceptual, intellectual, social, interpersonal, and personality-defensive processes."* (Goodenough, 1976, p. 1).

The authors overall agree on a cognitive style being a stable and consistent strategy in information processing that affects thinking, decision making, learning, and also social relations (Ausburn & Ausburn, 1978; Messick, 1976; Witkin, Moore, Goodenough & Cox, 1977). Kozhevnikov (2007) also admits that it is very difficult to measure cognitive styles because other cognitive functions or general intelligence often overshadow them. This is one of the reasons, together with increasing criticism, why the interest in cognitive style research weakened in the seventies.

Research focusing on the individual differences in information processing begun in the 1940s and 1950s with researchers concentrating on differences in participants' performance in various cognitive tasks. Sternberg & Grigorenko (1997) date the idea of individual differences to the early 20th century referring to Jung's psychological types. Hanfmann observed differences in strategies used by participants to group blocks and describes perceptual and conceptual style (Hanfmann, 1941 as cited in Kozhevnikov, 2007). Ash & Witkin (1948) were interested in the perception of an upright position, Klein (1951 as cited in Kozhevnikov, 2007)[1] described so-called levelers and sharpeners according to one's ability to detect and perceive differences in stimuli. Levelers are better at spotting similarities but are not as good at spotting differences. Sharpeners, on the other hand, are

1 Klein (1951 as cited in Kozhevnikov, 2007) does not refer to individual differences in perception as cognitive styles, but "perceptual attitudes".

better at perceiving and realizing differences in stimuli. Witkin (1950) focused on individual differences in perceiving and processing perceptual field, defining field dependence/field independence style (FD/FI; in the literature also called individual differentiation style; see *chapter 1.2* for more information). Nisbett (Nisbett & Masuda, 2003) inspired by Witkin's FD/FI style, describes holistic and analytic style while taking culture into consideration. The holistic individuals usually come from collectivist cultures, while the analytic individuals from individualist cultures. Kagan (1958 as cited in Sternberg & Grigorenko, 1997) differentiates between reflectivity and impulsivity in decision-making processes, while impulsive individuals tend to respond impulsively, reflective individuals tend to act after some reasoning and thinking. In the literature, this style is sometimes referred to as conceptual tempo (Sternberg & Grigorenko, 1997). More applied theories of styles define different approaches towards learning (learning styles; e.g., Kolb, 1974 as cited in Sternberg & Grigorenko, 1997) and teaching (teaching styles; e.g., Henson & Borthwick, 1984 as cited in Sternberg & Grigorenko, 1997).

The following table summarizes the most relevant cognitive styles, their authors, and a short description. The styles are divided into three groups, those aiming at cognition, those dealing mostly with personality and those focusing on activity and behavior (according to Sternberg & Grigorenko, 1997).

Tab. I: *List of most relevant and cited cognitive styles*

Style	Author	Definition
Cognitive approach		
Perceptual and Conceptual style	Hanfmann, 1941	Tendency to formulate hypotheses before approaching certain tasks
Field Dependence/Independence	Witkin, 1950	Perception of an upright position; Tendency to process perceptual stimuli regardless of the background, ability to work with figure regardless of the background
Levelers and Sharpeners	Klein, 1951	Tendency to perceive differences vs. similarities
Reflectivity and Impulsivity (Conceptual tempo)	Kagan, 1958	Tendency to react impulsively vs. thinking before approaching a certain task.
Personality approach		
The theory of psychological types	Jung, 1923	Three dimensions of personality types: extraversion/introversion; thinking/feeling, intuition/sensing
Behavioral approach		
Learning styles	e.g., Kolb 1974	How pupils are comfortable with various methods of teaching; pupils' approach towards learning
Teaching styles	e.g., Henson & Borthwick, 1984	Different approaches to teaching in the educational process

Source: Kozhevnikov, 2007; Sternberg & Grigorenko, 1997; modified by the author

Cognitive styles are considered to be equally useful strategies of information processing and adjustment to the environment. The 1970s brought a weakening of interest in cognitive styles which Kozhevnikov (2007) explains by the fact that many styles seemed to have one beneficial, advantageous strategy and one disadvantageous strategy. The general definition of cognitive style is based on the premise that cognitive styles are equal regarding the quality of information processing. One-dimensional style with two poles is represented by two different approaches of processing information with both being equally beneficial for the individual. Cognitive styles are simply different but equally beneficial approaches towards perceiving and processing information. In line with this statement, Berry, Poortinga, Segall and Dasen (2002, p. 137) also state that *"[...]cognitive styles refer more to "how" (stylistic) rather than "how much" (ability) aspects of a person's cognitive life."*

This does not seem to apply to the cognitive style of field dependence/field independence (Witkin, 1950; 1965; Witkin, Goodenough & Oltman, 1979). The next chapter focuses on this particular cognitive style, tries to explain problems with its testing, and provides an overview of its criticisms.

1.2 Field dependence/independence

Herman Witkin and his colleague Solomon Asch started the debate about the FD/FI cognitive style dividing people according to their ability to process visual stimuli and to perceive upright position some seventy years ago (Asch & Witkin, 1948). To this day, this style raises questions among researchers not only in cognitive psychology research but also in personality psychology or educational psychology research. Field dependence/independence is supposed to be representing a one-dimensional bipolar cognitive style capturing differentiation of articulated vs. global cognitive functioning represented in differences in intellectual and perceptual tasks (Witkin & Goodenough, 1977). Later on, FD/FI has been expanded on various other domains such as social interaction (Fitzgibbons, Goldberger & Eagle, 1965; Goodenough, 1976) and language processing (verbal skills; Khodadady, 2012). As the original Witkin's research suggests, field independent individuals (FI individuals) are better at processing visual information from the visual field in terms of processing figure and the background separately according to the momentary importance. Based on Witkin's results, FI individuals can operate with figure separately from the background and are better at perceiving an upright position regardless of their environment (Asch & Witkin, 1948). Field dependent individuals (FD individuals) are better at social situations and have more developed verbal skills (Khodadady, 2012) and also have a better involuntary memory of socially relevant material

(Fitzgibbons, Goldberger & Eagle, 1965; Goodenough, 1976). They are not as good as FI individuals at visual tasks and spatial tasks, and they are sensitive to laws of Gestalt (see Wagemans *et al.*, 2012b). According to Witkin (and his colleagues), FI and FD individuals also differ in defense mechanisms of the Ego (Witkin, Goodenough & Oltman, 1979). For a more elaborate list of differences between FI and FD individuals see Tab. II.

Tab. II: *The differences between Field independent and Field dependent individuals*

Area	Field independent	Field dependent
Memory	Greater capacity of working memory, especially central executive and visuospatial sketchpad (Miyake *et al.*, 2001; Rittschof, 2010)	Better memory for social interactions (Fitzgibbons, Goldberger, & Eagle, 1965; Messic & Damarin, 1964), better memory for socially relevant material (Goodenough, 1976) as opposed to focal tasks (Fitzgibbons, Goldberger & Eagle, 1965)
	Better at selective encoding (Richardson & Turner, 2000)	Better at selective combining and selective comparing processes (Richardson & Turner, 2000)
Attention	Better at most of the attention tests (Guisande, Páramo, Tinajero & Almeida, 2007)	Problems with maintaining attention among primary school children (Guisande *et al.*, 2007)
Personality	Better self-object differentiation and sense of separate identity (Witkin, Goodenough & Oltman, 1979)	Relate to others about opinions on themselves; greater social orientation (Goodenough.1976)
Suggestibility		More suggestible and prone to believe in paranormal phenomena (Hergovich, 2003)
		Greater conformity when working with FI individual in a laboratory setting (Solar, Devenport, & Bruehl, 1969); more competitive strategies in conflict resolution tasks in real life situations (Hearly, Harkins & Ray, 2010).
Ego defense mechanisms	Isolation, intellectualization and projection (Witkin, Goodenough & Oltman, 1979)	Denial, repression (Witkin, Goodenough & Oltman, 1979)
Academic performance	Overall better academic performance (Khodadady, 2012; Tinajero & Páramo, 1997; Zhang, 2004)	

Source: modified from Krajčíková, 2014

1.2.1 Figure-ground organization

One of the main interests of the Field dependence/independence cognitive style is the ability to perceive and manipulate with the figure regardless of the background. The main tool for testing the FD/FI cognitive style (Embedded Figures test) is based on this ability. To understand how figure-ground organization in human perception works, Wagemans and colleagues (2012a) provide an overview of figure–ground organization principles.

First of all, there are two points of view on whether a past experience has any effect on figure-ground organization. As Wagemans (*et al.*, 2012a) states, from a structuralist point of view (e.g., Wundt) a past experience is a strong (if not the only) determinant of human perception of figures and a background in a visual field. The Gestalt tradition is opposite in opinion (e.g., Koffka, Wertheimer). They claim that relying on past experience in visual perception would be disadvantageous because it would depend too much on memory (Wagemans *et al.*, 2012a). They argue certain principles that are used in figure-ground organization in human visual perception processes do not rely on the familiarity of the perceived.

The figure and the ground are the regions of a visual field that share the same border, but the figure usually "owns the border", the ground continues as the background (Wagemans *et al.*, 2012a). The figure–ground organization principles could be divided into those explaining the organization of static picture–based stimuli and those dealing with motion and organization of moving figure–ground stimuli. The first set of principles will be described here as it is relevant to the rest of the thesis. It seems that the strongest principles used in figure–ground organization are convexity and symmetry.

Convexity (Fig. 1) refers to the tendency to perceive the convex parts of the stimulus as figure while the concave parts are considered to be the background (Kanizsa & Gerbino, 1976 as cited in Wagemans *et al.*, 2012a). Several experiments showed that perception of the convex regions as the figure works if the convex regions are homogeneous in colour (the colour itself does not play a significant role), the colour of concave regions and its homogeneity does not matter.

Symmetry (Fig. 2) is a principle which assumes that symmetrical regions would be more likely perceived as figures than asymmetrical regions. The principle of global symmetry is a little bit weaker than the convexity principle, but research suggests that it plays a role in figure–ground perception (Wagemans *et al.*, 2012a).

Fig. 1: *The example of convexity in figure–ground organization.*
Source: Wagemans *et al.,* (2012 a) with permission.

Fig. 2: *The principle of symmetry*
Source: Wagemans *et al.*, (2012 a) with permission.

Fig. 3: *Principle of a lower region*
Source: Vecera, Vogel, & Woodman (2002) with permission.

Fig. 4: *Top-Bottom principle. In this particular example, the familiarity also plays a role as the ocean waves are usually familiar to the observer*
Source: Vecera, Vogel, & Woodman (2002) with permission.

Lower region is perceived as a figure after a horizontal border divides a rectangular object into two parts. The region above the border is usually perceived as the ground. This principle is the strongest with the border consisting solely of straight vertical and horizontal lines (Fig. 3).

Also, regions that are wider at the bottom and narrower at the top are usually perceived as a figure compared to those narrower at the bottom and wider at the top. The **Top-bottom polarity** (Fig. 4) principle is explained by Wagemans (*et al.*, 2012a) as an effect of gravity where objects on earth tend to be wider at the bottom for more stability.

There are other principles dealing mostly with moving stimuli (such as **Articulating motion** and **Advancing region motion**), which are not of interest here. Also, principles dealing more precisely with region grouping (such as **Region groupings and gradient cuts** and **Edge-region grouping**) will not be closely discussed here. For more examples and a research overview see the work of Wagemans (*et al.*, 2012a). This chapter briefly introduced the figure–ground organization principles, which are important for understanding the FD/FI cognitive style. Next, the development of this style is discussed.

1.3 Ontogenetic development of FD/FI

To address the development of styles during ontogenesis, Moskowitz, Dreyer & Kronsberg (1981) took an environmental approach and studied mother-child interaction among preschool children concerning FD/FI. The authors did not find any correlation between the behavior of the mothers towards their children and the development of cognitive style, thus the environmental influence seems to play a marginal role while biological factors could be the real influencer in the development of styles (Waber, 1977). Waber (1977) regards the neurobiological, endocrinological and genetic factors as much more influential than the environmental ones. The environmental and cultural effects on the development of style are described by Čeněk (2015) in his work (based on work of Berry *et al.*, 2002), where he explains three levels of cultural influence on FD/FI. The first level is represented by *a family*. Families that support individuality and independence in children tend to raise more field independent offspring, whereas more conform families with more control over their kids' activities tend to raise more field dependent children. The second level has to do with culture. More traditional and conform societies support field dependence, more open and liberal cultures support the development of field independence. The third level reflects the way of gathering supplies. Field independence is typical for hunter/gatherer kind of societies and field dependence for agrarian, more conform societies (Berry *et al.*, 2002; Goodenough & Witkin, 1977; Witkin & Berry, 1975).

1.4 Testing FD/FI: A historical approach

Testing FD/FI was first based solely on the perception of the upright (Asch & Witkin, 1948). The first tests used to measure FD/FI cognitive style were the Rod-and-Frame test (RFT), Rotating-Room test (RRT) and Body-Adjustment test (BAT).

The Rod-and-Frame Test (Fig. 5) consists of a glowing rod located inside a glowing square-shaped frame that is tilted in a particular way (e.g., 28° from the upright position). A subject is supposed to adjust the rod to its upright position regardless of the frame in 24 subsequent trials. Supposedly FI subjects are better at setting the rod upright compared to FD subjects, who seem to have problems to separately work with the rod (figure) and the frame (background).

The Body-Adjustment Test[2] (Fig. 6) is based on a similar premise about the participants' capability of perceiving the upright (Witkin, 1965; Witkin, Goodenough & Oltman, 1979).

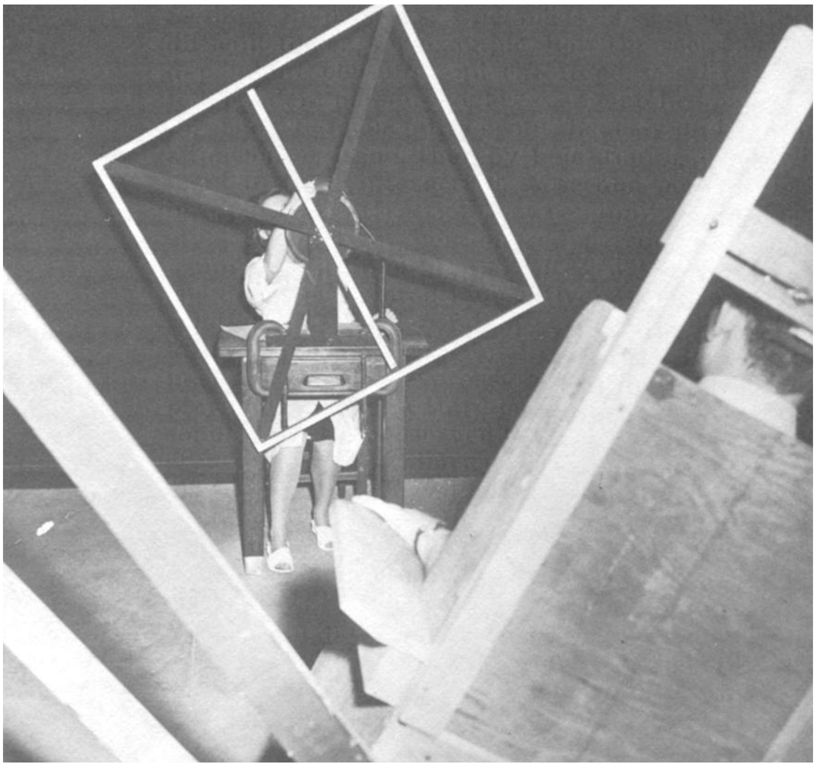


Fig. 5: *Rod-and-Frame test*
Source: Witkin, *et al*., (1977).

2 Originally called Tilting-Room-Tilting-Chair test (Čeněk, 2015).

In this test, a participant is sitting in a tilted chair and the task is to rotate the chair into an upright position regardless of the room, which is also tilted. By the same logic as in the previous test, the FI subjects perform better than FD subjects at setting the chair to the upright position regardless of the tilted room.

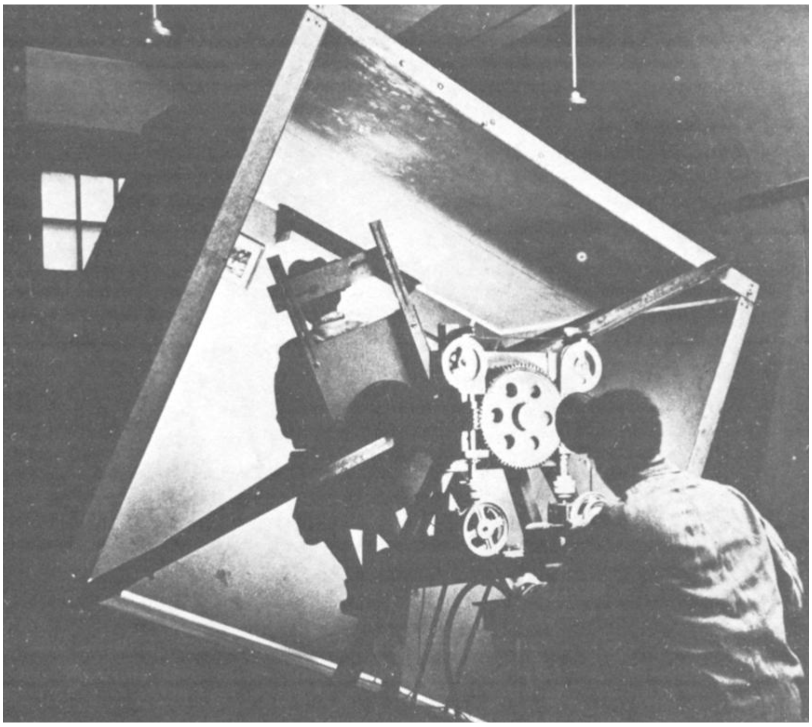


Fig. 6: *Body-Adjustment Test*
Source: Witkin *et al*., (1977).

Alternatively, a participant can operate with the room itself and get it to the upright position regardless of the tilted chair s/he is sitting on (Room-Rotation-Test).

The tests mentioned here (RFT, BAT, RRT) are all very demanding regarding space and equipment. The administration of only a single participant at a time together with the problematic technical setting that occupies a lot of laboratory space calls for a method that can overcome these obstacles[3].

3 The RFT was tested in semi–immersive 3D environments by Reger et al. (2003), but it did not reach reliability comparable to the classical 2D method. Interestingly, the authors of this study correlated RFT with test of spatial ability and came to the conclusion that RFT could be used to measure visuospatial skills.

1.4.1 Embedded Figures test

Even though the Embedded Figures test (EFT) does not examine the perception of the upright, it is very similar to RFT and BAT in its "[…] *essential perceptual structure"* (Witkin *et al.*, 1977, p. 20). This test is derived from Gottschaldt's (as cited in Witkin, 1950) research examining the influence of past experience on visual perception. The Embedded Figures test (for example of the materiel see Fig. 7) measures the ability to separately operate with figures regardless of the background, more specifically, finding the figure in a complicated complex background (see *chapter 1.2.1* for more information on figure-ground organization).

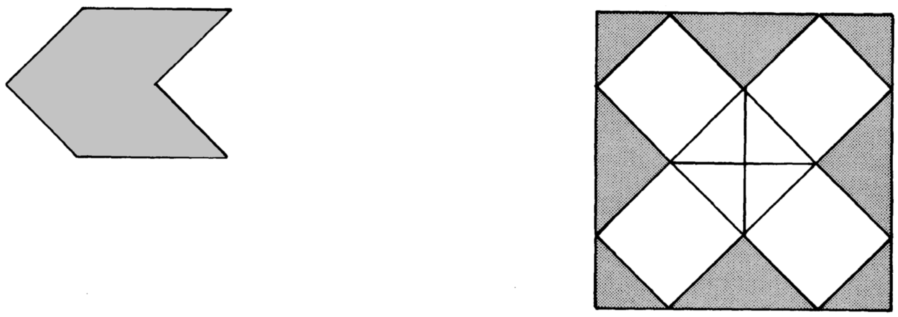


Fig. 7: *Example of a trial in EFT*
Source: Adapted from Witkin *et al.*, (1977).

Witkin's research showed a significant advantage of FI participants at EFT in comparison to FD participants (Witkin, 1950). The test was originally administered individually and consisted of 24 trials. In each trial, the participant was presented with a complex design for 15 seconds, then with a simple figure for 10 seconds and with the complex design again for a maximum of 5 minutes. If the participant was not able to find the previously presented simple figure in the complex design within 5 minutes, the trial was considered failed. The participant could ask an administrator to show the simple figure again, but the complex design was then hidden so that the participant could not see the simple figure and the complex design at the same time. Also, time measurement continued. The time needed to complete all 24 trials was measured together with a number of trials that were solved correctly.

The norms for the original test were established based on 120 undergrad students. Witkin (1950) initially suggested separate norms for male and female participants as the collected data suggested that women were more frequently field dependent while men seemed to be more often field independent. This claim, however, has not been supported by later research where many authors suggest the same norms for both genders because they did not find any significant difference between male and female participants (Cummings & Murray, 1987; Hao *et al.*, 2013; Jackson, Messick & Myers, 1964; Lis & Powers, 1979).

Later on, the group version of EFT was experimentally tested (e.g., Cummings & Murray, 1987; Panek, Funk & Nelson, 1980) and also a short version of the test was introduced. It turned out that the short version of GEFT (12 items) satisfyingly distinguishes between extreme groups of FI/FD participants with a strong correlation to the original 24 item version (Jackson, 1956), therefore the short test is a valid alternative to the original longer version.

1.5 Criticism of the FD/FI cognitive style

The concept of FD/FI as a cognitive style was widely criticized in the research community mainly because it seems that nobody really knows, what the tests such as EFT or RFT really measure. First of all Lis and Powers (1979) did find only 4–16% of a common variance of data obtained by EFT and RFT. Also, field independent individuals seem to perform better in various tasks, which is inconsistent with the cognitive style definition of different but equally beneficial approaches towards information processing. Miyake, Witzki and Emerson (2001) tested their participants using dual-task paradigms, using EFT and a task loaded on a visuospatial sketchpad and central executive of working memory. Results showed that working memory capacity is a strong predictor of an EFT performance, which could mean that field independence is in fact closely related to working memory capacity and it does not measure a cognitive style in Witkin's point of view (Miyake, Witzki & Emerson, 2001; Rittschof, 2010). Tinajero and Páramo (1997) measured academic achievement in a group of high school students and found a positive correlation between EFT performance and achievement in six school subjects. The authors suggest a relation between FI and general intelligence. Similar results were obtained by Zhang (2004), who found a positive correlation between EFT performance and performance in a university geometry course in favor of FI students. Guisande *et al.* (2007) and Fitzgibbons, Goldberger and Eagle (1965) found a connection between FI and attention, Guisande's (2007) results show that the more field independent the participant was, the better would his/her performance be in an attention task (2007). Fitzgibbons and colleagues (1965) found that FI were significantly better in a focal task (digit symbol test), whilst FD participants were better at an incidental learning of material with social relevancy (such as recalling the content of a conversation between the experimenter and a colleague in the next room) but were significantly worse in a focal task.

1.6 The connection between FD/FI and spatial cognitive ability

Following previously mentioned research it seems that FD/FI itself is a problematic concept. The question is whether FD/FI is really a cognitive style or whether it represents a form of cognitive ability as suggested for example by MacLeod, Jackson and Palmer (1986) and others (Miyake, Witzki & Emerson, 2001). MacLeod and colleagues tested 60 students using GEFT and RFT and correlated their performance with the spatial cognitive ability performance in spatial rotations and block design subtest of WAIS and verbal performance (Lorge-Thorndike verbal battery). The verbal test served as a control subtest. In fact, the authors did not find any difference in verbal score between FI and FD participants. The results revealed that field independent participants' performance was significantly better in spatial ability tests than the performance of FD participants. The authors suggest that spatial ability and the concept of FD/FI are not two separate constructs. Goldstein and Chance (1965) found out that training in EFT and similar tasks significantly narrows the difference in performance between genders. These findings support the notion of FD/FI not being a cognitive style because training should have an effect on performances in cognitive ability tasks, not cognitive styles (Wright *et al.*, 2008). Ludwig and Lachnit (2004) obtained the same results, and their findings even suggest a transfer of a training effect to never trained, new tasks. This suggests that FD/FI is not a cognitive style, but a cognitive ability. Massa, Mayer & Bohon (2005) even used GEFT as a measurement of spatial cognitive ability based on their notion that GEFT "[…] *is thought to tap one aspect of spatial ability*" (p. 101).

The idea of FD/FI being a spatial cognitive ability rather than style was tested in the author's previous work (Krajčíková, 2014). The methodology was very similar to MacLeod, Jackson and Palmer (1986). We used spatial and verbal memory tasks to test differences between FI and FD participants. Participants were supposed to decide whether a target changed its location on a computer screen (first, three dots appear on a computer screen, then a mask and three dots again while this time a target was marked with a red circle, the two other dots were distractors. The participants did not know which dot would be the target, so they had to remember all three of them). We got the same results as MacLeod, Jackson and Palmer (1986) with FI participants being better in spatial memory tasks than FD participants. There was no difference between the performance of FI and FD participants in the verbal test (Auditory – Verbal Learning Test; *Czech name: Paměťový test učení*; Preiss, 1999). Even though we did not use the original EFT but a version of hidden figures test based strictly on Witkin's original EFT we were able to confirm MacLeod's results (MacLeod, Jackson & Palmer, 1986). This approach towards FD/FI style and Witkin's EFT could bring new use for EFT in psychological diagnostics. If EFT really is a test of spatial cognitive ability, it could be used to test this ability or its factors (see c*hapters 2.4* and *2.5.4*). The next chapter provides an overview of spatial cognitive ability, its testing and connection to FD/FI.

2 SPATIAL COGNITIVE ABILITY

This chapter is a review of known facts about spatial cognitive ability. It also provides an overview of tests created to measure this ability and its factors. Attention is mainly paid to a connection between spatial cognitive ability and EFT as a measure of FD/FI cognitive style. This part of the thesis can be considered the most crucial part, and it is closely linked to the empirical part of the thesis. It starts with basic foundations of the visual perception of space, provides definitions, explains tests, shows examples of tasks, and describes factors playing a role in spatial ability.

2.1 Visual space perception

Space perception is based on the ability to transform a two-dimensional picture reflected on an eye retina to a three–dimensional representation. To be able to perceive spatial information, the perceiver uses three types of cues (monocular, binocular and oculomotor; Eysenck & Keane, 2008; Šikl, 2012).

Monocular cues: These cues are often used by painters to create an illusion of a perspective. They are providing the information about the depth of a perceived scene even if the perceiver is observing the scene with only one eye (hence the term *monocular*).

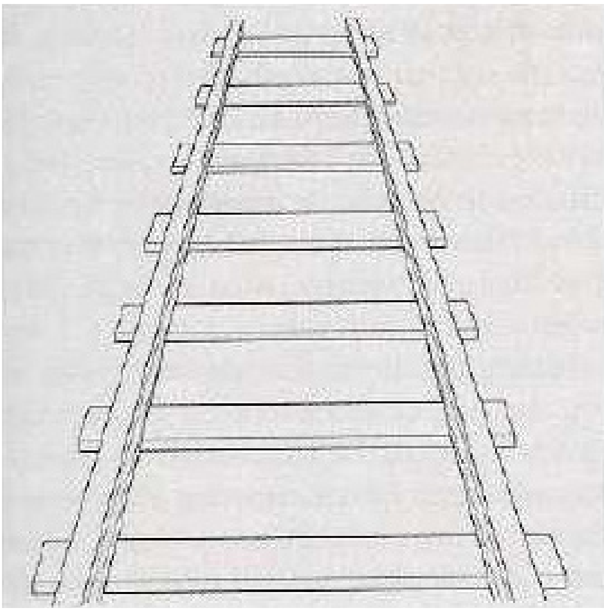


Fig. 8: *Example of a linear perspective*
Source: Depth Perception. (n.d.). [Picture]. Available from: https://bit.ly/2M9Ogsh, Downloaded 12. 7. 2018; 13.00 p.m.).

- **Linear perspective:** Parallel lines running into the distance from the observer seem to get closer to each other with increasing distance. A good example of this effect can be found when observing train tracks disappear at the horizon. The rails seem to be closer to each other with increasing distance (Fig. 8).

- **Interposition:** In the case of two objects overlapping, the closer object usually partially covers the further objects, so that the further object is not fully in sight. Objects that are partially covered by another object are perceived as more distant. Example of this cue is the Kanisza square (Fig. 9).

- **Texture gradient:** The density of a perceived texture is changing with distance. The bigger the distance, the higher the perceived density. We can experience this monocular cue again when observing train tracks. The railroad ties seem to be closer to each other with increasing distance, similar to pavement tiles which seem to be closer to each other (the density of the tiles seemingly increases with distance). This effect is caused by less focused observation of details with increasing distance (Šikl, 2012).

- P**revious experience** also plays a role in the perception of space. The **familiarity of the size** of an object affects the judgment of the distance of the object from the observer. The **aerial perspective** ensures a perception of objects that are a little blurry as more distant from an observer. Also, the **perception of the shades** on the object and the shade the object throws influence the perceived size and distance of the object. **Motion parallax** helps the observer to judge the distance of moving objects (Eysenck & Keane, 2008).

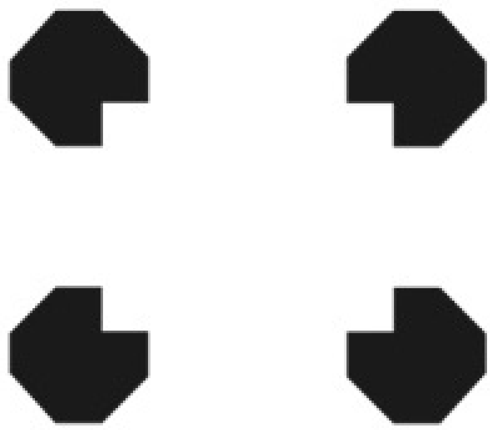


Fig. 9: *Kanisza square*
Source: (The original version of Kanisza square. (2011). [Picture].
Available from: https://bit.ly/2RsekVs, Downloaded 12.7. 2018; 11.30 a.m.).

Binocular and oculomotor cues: It is necessary to be able to perceive with both eyes at the same time to make judgments about the observer's environment based on binocular cues.

- **Stereopsis (binocular)**: There is a slight difference in object representation on each retina between both eyes. This difference serves as a cue in the perception of the 3D environment.

- **Convergence**: The closer the object, the more the eyes have to turn towards each other.

- **Accommodation**: The lenses are more narrowed during the perception of distant objects compared to close objects when it is more convex (Eysenck & Keane, 2008).

These cues are used in perceiving the environment visually and are the base for spatial cognitive ability. For thorough explanations and more examples of each cue see *Eysenck & Keane, (2008)* or *Šikl, (2012)*.

2.2 What is spatial cognitive ability

To define *ability* in the most general way, we should turn to the Oxford Dictionary. The first definition refers to *ability* as "*possession of the means or skill to do something* ", it is also a „*talent, skill, or proficiency in a particular area*", or doing something "[…] *as well as one can*" (*Oxford dictionaries* [online], 2018). Surprisingly, the articles dealing with testing abilities, measuring them or modeling hierarchies of abilities do not provide very satisfying definitions. What stands out is a characteristic of ability in Khatib and Hosseinpur's article (2011) referring to a difference between a style and an ability, where *"*[…] *abilities are* [considered to be] *unipolar. Having more of an ability is usually considered beneficial*" (Khatib & Hosseinpur, 2011, p. 640).

According to this definition, a certain ability, for example, a spatial ability, is a continuum where having more of the ability means being able to perform better in specific tasks related to the ability. Also, abilities could be trained. They are affected by practice (Wright *et al.*, 2008). Wright and colleagues (2008) examined the training of spatial ability via specific tasks and the generalization of that training and practice to other tasks. In the case of spatial ability, training is more effective in individuals with a lower level of spatial ability before the training compared to individuals with higher level of spatial ability before the training (David, 2012).

Gardner (1999) states that dimensions of intelligence (in other words factors, or abilities) are content-dependent, while for example styles (cognitive, learning, etc.) are content-independent.

Massa, Mayer and Bohon, (2005) talk about two different ways of how to perceive ability. First, Carroll (1993 as cited in Massa, Mayer & Bohon, 2005) sees ability as something that is "*a property of a person,* [it is] *in the person's head*", while Cronbach (2002 as cited in Massa, Mayer & Bohon, 2005) adds a context to the equation and sees ability as a combination of context and a person's characteristics, more like a "person-in-situation". Massa, Mayer & Bohon (2005) conclude that performance in cognitive ability tests actually depends on the interaction between a person's characteristics and the situation together with an interpretation of the situation.

A more specific definition of spatial cognitive ability is found in Anastasi: "*Spatial ability is an individual's skill in perceiving fixed geometric/spatial relations and in applying mental transformations such as rotation or reconfiguration to existing spatial relations.*"(Anastasi, 1976 as cited in MacLeod, Jackson & Palmer, 1986, p. 141) alternatively, in Lohman (1996, p. 3): *"Spatial ability may be defined as the ability to generate, retain, retrieve, and transform well–structured visual images".*

Gardner's definition (1999) of spatial intelligence states that spatial intelligence is a set of abilities that ensure a relatively accurate perception of the visual world, enable to transform and modify the primary perceived information, and make visual images from visual experiences . Thanks to these abilities we can construct objects in our mind and manipulate them. As Gardner (1999) claims, spatial abilities are a set of factors, where if one performs well in tasks testing one of the factors, s/he will perform well in the general spatial ability tasks as well (this does not apply to other abilities such as verbal or musical ability), and the practice and training will be transferable from one factor to the other factors and spatial ability as a whole. Gardner (1999) also explains why the word "spatial" is used rather than "visual". Even though spatial intelligence is closely linked and in the case of healthy individuals more or less dependent on visual perception, it is not exclusively dependent on it. Similar to the verbal ability, which is not solely dependent on auditory perception, spatial ability is not solely dependent on visual perception. Blind people also have a certain level of spatial ability (tested by tasks comparable to those testing spatial ability visually but with a tactile administration; Gardner, 1999, p. 197), but operate on different principles compared to healthy individuals.

As Gardner (1999) suggests, spatial ability is most likely linked to intelligence, thus the goal of the next chapter is to place spatial ability into models of human intelligence.

2.3 Spatial cognitive ability as a part of human intelligence models

It is crucial to explain, why spatial cognitive ability deserves any attention at all. According to Yeo and colleagues (2016), specialization for verbal or spatial cognitive ability is independent of the general cognitive ability, and at the same time, spatial ability is a factor playing a role in human intelligence. "S*patial intelligence*", in other words "*visuospatial intelligence*" is, according to Román and colleagues (2014), part of so-called *first order factors,* together with fluid and verbal-crystallized intelligence, related yet separate from general intelligence (*g* factor). Lohman (1993) claims that "*factorial studies of spatial ability routinely show that spatial tests are good measures of g*" (p. 11). It is also tested as a part of intelligence test batteries (e.g., Wechsler Abbreviated Scale of Intelligence, Yeo *et al.*, 2016; Kyodai Japanese IQ test, Ando, Ono & Wright, 2001). Spatial ability is included in several intelligence models[4], both in hierarchical (e.g., Vernon, 1950) and in multi-factor models of intelligence (e.g., Gardner; 1999). In the CHC hierarchical model of intelligence (Cattell-Horn-Carroll model of human cognitive abilities), human intelligence is divided into three strata of intelligence factors (Schneider & McGrew, 2013; McGrew, 2009). The first one is the *General* stratum represented by the *g* factor. The middle stratum is represented by 16 *Broad* factors (e.g., *Visual processing*, *Cognitive processing speed, Reading and writing,* etc.). The most specific stratum, the *Narrow* factors, specifies what each of the broad factors is represented by. In our case, spatial ability would be part of the *Visual processing* "Broad" factor where the "Narrow" factors related to spatial ability would be *visualization, spatial relations (speeded rotation in Schneider & McGrew's terminology, 2012), spatial scanning, perceptual alternations and imagery* (with the rest of the factors plaing an important role in spatial ability; Ichise, 2018; McGrew, 2009). For a closer description of the CHC factors see, for example, Carroll, (1993) or McGrew, (2009) and a following Tab. III for detailed look on visual processing factor of CHC model.

Also, in Gardner's multi−factor model of intelligence *"spatial intelligence gives one the ability to manipulate and create mental images in order to solve problems"* (Brualdi, 1996, p. 1).

Thurston (1938 as cited in MacLeod, Jackson & Palmer, 1986) also includes spatial ability in his multi-factor theory of intelligence. Spatial ability could be found on the background

4 The difference between a model and a theory provide Conway & Kovacs (2015). The authors define theory as a set of ideas, which are supposed to explain something and are testable (could be falsified), whereas a model is an implementation of the theory, which requires a certain level of assumption for it to fit the theory.

Tab. III: *Detailed look into the visual processing factor of the CHC model with all its subfactors, edited by the author from McGrew, (2009, p. 5). See McGrew's work for a list of all the factors and subfactor of the CHC model.*

Visual processing	**Visual processing**	**Broad visual perception**	
(Gv) (CHC terminology)	(Gv) (Cattell–Horn terminology)	(Gv) (Carroll terminology)	"The ability to generate, store, retrieve, and transform visual images and sensations. Gv abilities are typically measured by tasks (figural or geometric stimuli) that require the perception and transformation of visual shapes, forms, or images and/or tasks that require maintaining spatial orientation with regard to objects that may change or move through space. •Visualization (Vz), Spatial relations (Sr) */ or speeded rotation (in Schneider & McGrew, 2012)*, Closure speed (Cs), Flexibility of closure (Cf), Visual memory (Mv), Spatial scanning (Ss), Serial perceptual integration (Pi), Length estimation (Le), Perceptual illusions (Il), Perceptual alternations (Pn), Imagery (Im)" (McGrew, 2009. p 5).[5]

Source: Adapted from McGrew (2009).

of Thurston's *spatial visualization,* which would correspond to spatial orientation and mental rotation factors by Hegarty and Waller (2004), *perceptual speed* which would link with Miyake, Witzki and Emerson's (2001) factor of perceptual speed and with Ekstrom, Dermen and Harman's (1976) flexibility of closure.

The visuospatial component is a direct part of the working memory model (Ando, Ono & Wright, 2001; Baddeley & Hitch, 1974). Other authors also recognize the spatial component as a part of the working memory. According to Constantinidis & Wang (2004), spatial working memory is manifested (among other brain regions; for a detailed description see Constantinidis & Wang, 2004) in the prefrontal and parietal association cortices.

5 For detailed definitions of all the factors see Schneider & McGrew (2012). For our purposes the most important factor is Visualization defined as: "*The ability to perceive complex patterns and mentally simulate how they might look when transformed (e.g., rotated, changed in size, partially obscured, and so forth).* [This] *is central to Gc, this is the core ability of Gv*" (Schneider & McGrew, 2012, p. 9).

2.4 Individual differences in spatial cognitive ability

A lot has been said about gender differences in spatial cognitive abilities throughout the years of research in the area. For that reason, we decided to address this issue briefly in this chapter, which is focused on gender differences in performance at specific spatial tasks.

Chraif's (2013) study shows a significant difference in spatial ability tasks among undergraduate student with an obvious advantage on the male's side. Males had significantly more correct responses in the mental rotation task compared to females. Górska (2005) very interestingly describes her findings where she claims that female students enter technical universities with a lower level of spatial abilities compared to their male classmates. As we know from David (2012), training is very beneficial to individuals with lower levels of spatial skills, so we could speculate that female students would benefit more from training their spatial skills prior or after entering the university and therefore narrow the gender gap. This could be useful information for teachers and mentors, who want to encourage female high school students to enroll in technical fields of study at universities. The question still stands, why female students have a lower level of spatial abilities or skills compared to their male counterparts. One explanation lies in the cultural influence on girls and boys at a young age. Parents' attitude toward gender stereotypes could play a certain role in forming attitudes towards specific subjects at school and specific professions (such as researcher, mathematician, etc.) for young children. According to Meyer, Cimpian and Leslie (2015), the gender gap in performance could be explained by so-called "field-specific ability beliefs", which means that those fields that are believed to be based on sheer talent and also are believed to be more difficult for women (it is stereotyped that women are not as talented as men in them) would be less attractive to women, they would *a priori* refuse to participate in them or practice them. Smetackova (2015) also provides examples of gender schemata in math that are applied at primary schools and high schools among boys and girls influencing their beliefs about their competencies doing math. Massa, Mayer and Bohon (2005) tested the influence of gender role (measured by Bem sex role inventory) on spatial ability tests. They found out that female participants with more masculine gender roles performed better at the test when described as measuring spatial ability. Female participants with more feminine gender roles performed better at the same test when described as measuring empathy. They did not find significant differences between male participants with feminine and masculine gender roles. The authors conclude that a participant's gender role beliefs, and the way a test or a task is described as measuring feminine or masculine skills could play a significant role in a participant's performance at various cognitive tests (Massa, Mayer & Bohon, 2005). Participants supposedly try harder at tasks that they interpret as in line with their gender role. Feminine participants would achieve better results at tasks measuring skills believed to be linked with female gender. Masculine participants would perform better at tasks believed to be more in line with male gender. These beliefs almost certainly have their roots in culture.

The cultural explanation could be partly true, but is there not more behind these gender differences? Yeo and colleagues (2016) examined the brain of 244 healthy young adults using MRI and found specific brain regions in men related to cognitive abilities, predicting spatial but not verbal ability. They suggest (in line with other authors) that cortical thickness could be a predictor of spatial ability. The authors conclude that levels of gender−specific cognitive skills could be affected by sex hormones and environment as well as differences in neurodevelopmental trajectories between sexes. Toivainen, Papageorgiou, Tosto and Kovas (2017) state that gender differences in general cognitive ability are negligible while it is still believed that females are better in verbal tasks whereas males being better in spatial tasks. More recent literature shows that it does not have to be the case. Females perform better in natural language abilities such as writing and reading while males outperform females in word analogies (Geary, 2014) which, together with the fact that differences at spatial tasks are the most robust in mental rotations, could suggest that gender differences are task−specific, not ability−specific (Toivainen *et al.*, 2017). The authors also claim that research on gender differences in verbal and spatial skills does not provide stable evidence. Many reasons are debated from cultural influence to the hormonal effect of testosterone on spatial skills. Interestingly, Berenbaum, Bryk, and Beltz (2012) found that women who were exposed to a high level of prenatal androgens in utero (congenital adrenal hyperplasia) perform better at mental rotation tasks compared to controls. Higher levels of prenatal testosterone can also negatively influence vocabulary in infants according to Lutchmaya, Baron−Cohen and Raggatt (2001). It is important to say that these differences could be specific to a particular developmental phase and the differences might disappear over time (Toivainen *et al.*, 2017). After what has been said about task-specific and developmental-stage-specific differences between genders in cognitive abilities, we could speculate that these differences based on biological, developmental or neuroanatomical foundations might play a role in forming social attitudes and parental and teachers' expectations towards children. These may cause later avoidance of certain tasks and activities in both genders with later impact on chosen fields of study (Shi, 2018).

2.5 Defining factors of spatial cognitive ability

There are several opinions supported by research studies on how many and which factors play a role in spatial cognitive ability (e.g., Borich & Bauman, 1972; Hegarty & Waller, 2004. Zacks *et al.*, 2000). According to the authors, two factors are playing a crucial role in the spatial ability debate, the **spatial orientation factor** (also perspective taking factor or egocentric transformation) and the **mental rotation** factor (also spatial visualization factor or object−based transformation; Kozhevnikov & Hegarty, 2001). The best definition of the factors is provided by Hegarty and Waller (2004).

*"The **spatial visualization factor** has been conceptualized as the ability to make object−based spatial transformations in which the positions of objects are moved with respect to an environmental frame of reference, but one's egocentric reference frame does not change. In contrast, the **spatial orientation factor** has been interpreted as the ability to make egocentric spatial transformations in which one's egocentric reference frame changes with respect to the environment, but the relation between object-based and environmental frames of reference does not change*" (Hegarty & Waller, 2004, p. 176).

Miyake *et al.*, (2001) divide spatial ability into spatial visualization factor, spatial relations and the **perceptual speed** factor and claim that these three factors correlate but are distinguishable. After examining tests used by Miyake and his colleagues to test these three factors in participants, we concluded that both spatial visualization and spatial relations are subfactors of the mental rotation factor. It is not clear, whether this applies to the perceptual speed factor, so we will discuss measuring this factor and tests used to examine this at the end of this chapter.

In Hegarty and Waller's (2004) studies, the authors tested undergraduate students using several methods to test the spatial orientation factor (the perspective taking factor) and the mental rotation factor (spatial visualization factor). The methods are described in the chapter dealing with testing spatial cognitive ability (*see below*). They tested two models on CFA, one testing the hypothesis that spatial orientation and mental rotation tests are both saturated by one spatial factor. The second hypothesis tested the possibility that the spatial orientation factor and the mental rotation factor are two different factors and they are loaded by different tests of spatial ability. In both studies (operating with different tests), the two-factor model had a better fit with the data. The two spatial factors highly correlate ($r = .80$), so it is safe to say they share considerable variance (Hegarty & Waller, 2004).

Borich and Bauman (1972) found evidence for the convergent validity between tests measuring spatial visualization and spatial orientation but did not obtain satisfactory evidence of discriminant validity. The evidence suggests that there are two factors in spatial ability, but these factors highly correlate with each other.

Lohman (1996) provides an overview of spatial ability factors while recognizing five factors: Visualization, Speed Rotation, Closure Speed, Closure Flexibility, and Perceptual Speed. After examining definitions of each factor, we decided to use Hegarty and Waller's (2004) terminology regarding spatial ability factors. In Lohman's case, Visualization and Speed Rotation would fall under the Mental Rotation factor, Closure Speed, Closure Flexibility and Perceptual Speed fall under the Perceptual Speed factor (Ekstrom, Dermen & Harman, 1976; Miyake *et al.*, 2001). A spatial orientation factor is not discussed in Lohman's work (1996). For better understanding see Attachment 4.

Even though the debate about factors in spatial ability is still open, tests to measure spatial ability could be divided into two groups: those testing spatial orientation and those testing mental rotation (visualization) with consideration of a possible third factor, perceptual speed (Ekstrom, Dermen & Harman, 1976; Miyake *et al.*, 2001). In this work, the mental rotation factor is used to measure spatial ability, so the chapter is more focused on tests used to measure this particular factor. The mental rotation test (Shepard & Metzler, 1971) is described thoroughly. The method used in this work is described in the *Method* section of the empirical part of the thesis.

2.6 Testing spatial cognitive ability – an overview

2.6.1 Testing the mental rotation factor

The first part of this chapter consists of a list of tests intended to test the mental rotation factor of spatial cognitive ability together with an explanation of the trials and, in some cases, examples of the material.

First of all, a Card Rotation Test (Ekstrom, Dermen & Harman, 1976) may be used to test the **mental rotation factor**. In this test, the participant is presented with a polygon and has to decide between five options which alternative is the rotated version of the target polygon (the distractors are mirror images of the target) while time and accuracy are measured.

The second test worth mentioning is the Flags Test (Fig. 10; Thurstone & Thurstone, 1941 as cited in Hegarty & Waller, 2004), where participants judge which of the six options is a planar rotation of a flag seen before.

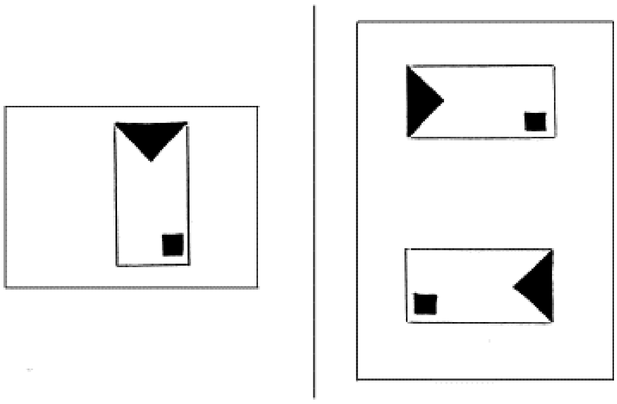


Fig. 10: *Example of a trail in the Flag Test*
Source: Davidoff, & Warrington (2001) with permission.

In the Cube Comparisons Tests (Ekstrom, Dermen & Harman, 1976) two drawings of cubes with letters and numbers on their sides are presented. The task is to determine, whether the two cubes are the same rotated cube, or whether they are two different cubes (Fig. 11). Interestingly, Cube Comparison and Card Rotation Tests are both listed under Spatial orientation factor in Ekstrom, Dermen and Harman (2001) while in this work we will follow (Hegarty & Waller, 2004), are listed under Rotation factor.

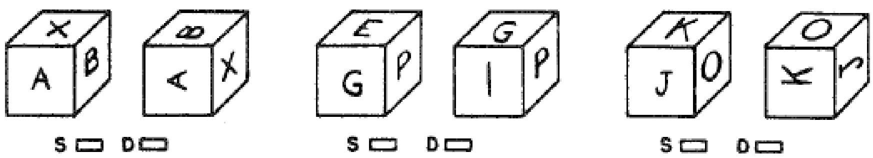


Fig. 11: *Cube Comparison test*
Source: Ekstrom, Dermen & Harman, (1976) with permission. Copyright © 1976 Educational Testing Service. www.ets.org. Figures from the *Manual for Kit of Factor-Referenced Cognitive Tests* are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this publication is provided by the author and no endorsement of any kind by Educational Testing Service should be inferred.

The Guilford-Zimmerman Test for Spatial visualization requires the participant to choose out of five options which one the position of an alarm clock has after rotation (Guilford & Zimmerman, 1956 as cited in Borich & Bauman, 1972). Neumann, Fitzgerald, Furedy and Boyle (2007) created a 2D test of mental rotations, where participants judge pairs of histograms, whether they are the same (rotated) or a mirror image.

The Vandenberg Mental Rotations Test (Vandenberg & Kuse, 1978) uses three-dimensional figures based on Shepard & Metzler's (1971) test of Mental Rotations (*see below*). Participants decide between four figures and have to pick two that represent the rotated target.

In the Paper Folding Test (Ekstrom, Dermen & Harman, 1976) a piece of folded paper (penetrated at a particular spot, see Fig. 13) is presented and the participant has to decide how the piece of paper would look like unfolded. The Surface Development Test is somewhat similar to the Paper Folding Test. The participant is supposed to choose which 3D object would be created by folding a piece of paper (Fig. 12; Ekstrom, Dermen & Harman, 1976; Wang *et al.*, 2014).

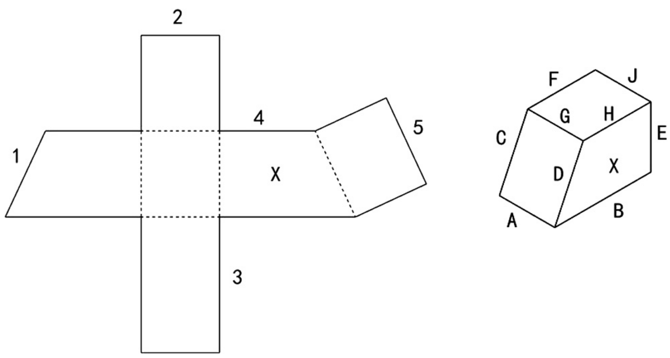


Fig. 12: *Example of the Surface Development test*
Source: Wang *et al.,* (2014) with permission.

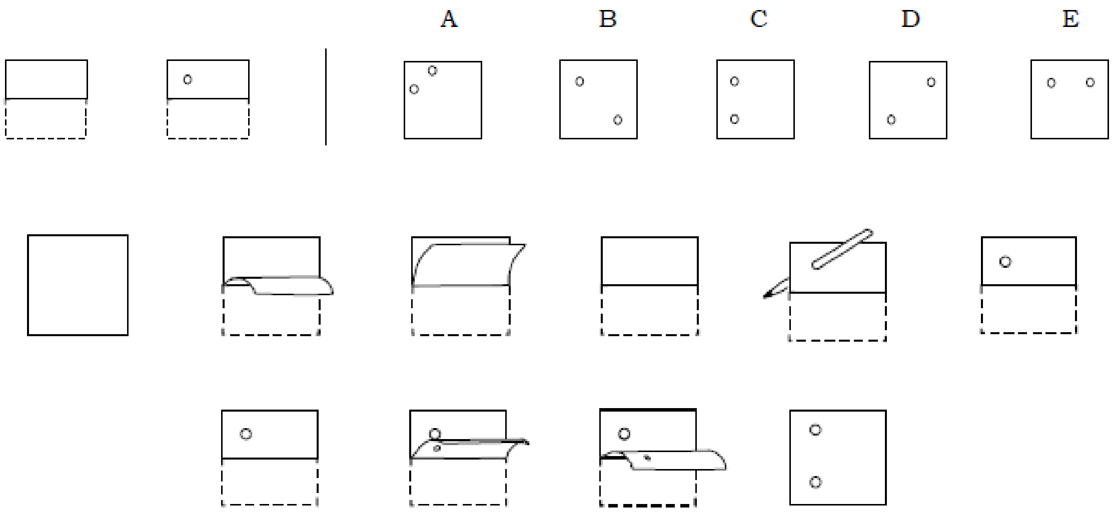


Fig. 13: *Paper Folding test*
Source: Ekstrom, Dermen & Harman, (1976) with permission. Copyright © 1976 Educational Testing Service. www.ets.org. Figures from the *Manual for Kit of Factor-Referenced Cognitive Tests* are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this publication is provided by the author and no endorsement of any kind by Educational Testing Service should be inferred.

2.6.2 Origins of Mental rotation tests

"*Mental rotation is an important paradigm for spatial ability*" (Xue *et al.*, 2017, in abstract).

The pillar of the mental rotation tests lies in the classical article published in *Science* in 1971 by Shepard and Metzler (Shepard & Metzler, 1971). As it is usually the case in the *Science* journal, it is a short study of only three pages, but it has started a dispute about measuring the ability to mentally imagine objects and work with them. That was a rather new approach at the era of behaviorism, which was still a respected mindset about mental processes in psychology at that time, and which claimed that mental processes were solely dependent on verbal representations, in other words, language. Despite this common notion among psychologists, Shepard and Metzler (1971) showed, that their task did not require verbal representations to complete it correctly (even though there was a slight shade of doubt by Marks, 1999; Pylyshyn, 2003). The task required a simple decision about displayed three-dimensional objects. Participants are presented with the study object and a test object (objects were presented side by side, both at the same time), which is either the same as the study one only rotated (rotation by depth or rotation in the picture plane), or it is a mirror image of the study object (also rotated). The participant has to decide as fast as possible if the test object is the same as, or the mirror image of the study object (Fig. 14). The results are based on eight adults responding to 1600 pairs of objects, half of which were the "same" and half were the mirror images.

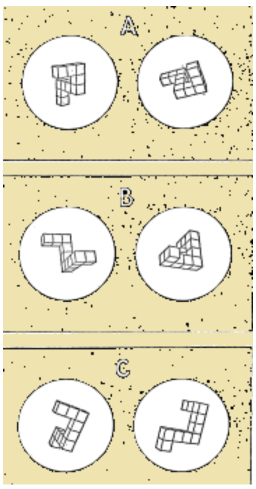


Fig. 14: *Example of mental rotation tasks*
Source: Adapted from Shepard & Metzler (1971).
A: Same objects differing by a rotation in the plane of the page
B: Same objects differing by a rotation in depth
C: Mirror – image objects, rotated in depth

The objects all consist of three – dimensional cubes put together. There are always three right angles in the object. The study objects have been rotated to a different degree. The results show a linear relation between the mean reaction time for the "same" objects and a degree of a rotation, which had been interpreted to be due to the longer time for evaluating objects that are rotated more because it takes more time for the participants to rotate the objects in their mind. The reaction time for the "different" object also shows similar patterns and is overall longer than for the "same" objects. The interpretation of the results was confirmed by subjective statements of participants, who claimed that the more the object was rotated, the more time they needed to answer. This study suggests that the comparison of two rotated visual stimuli (objects) requires mental rotation of these objects (Shepard & Metzler, 1971; Cooper, 1976).

This study, like any other, was not left without criticism and additional comments. Hochberg and Gellman (1977) claim that so-called "landmarks" of the object are crucial for mental rotation tasks. Landmarks refer to some characteristics of the object that help participants to make the "same/different" decision in mental rotation tasks. Mental rotations of the objects with salient landmarks show lower slopes in the time/angle linear function than those with less distinguishable landmarks (Hochberg & Gellman, 1997). On the other hand, Just and Carpenter (1976) used eye-tracking to find out how mental rotations are processed within human cognition. They assume that duration of eye fixation on a certain spot while completing the task refers directly to process duration for an individual. Even though the authors obtained results comparable to Shepard & Metzler (1971), they propose the notion that the duration of eye fixations can explain underlying processes in mental rotation tasks. They claim that the more the object is rotated, the longer it takes for the participants to compare the objects, therefore, longer eye fixations are needed. The authors describe three stages that are required to complete the mental rotation task: initial search part (where the participant searches for comparable parts of the object; similar to "landmark" idea of Hochberg & Gellman, 1977), transformation phase (the rotation itself), and the final comparison process. The more the object is rotated, the longer are the eye fixations. This is how Just and Carpenter (1976) explain the relation between the degree of rotation and the time needed to answer in the Shepard and Metzler's (1971) experiment.

To confirm the data obtained by Shepard and Metzler (1971), Shepard's student Lynn Cooper used a variety of stimuli for mental rotation tasks such as random polygons (Fig. 15), or alphanumeric symbols (Fig. 16).

Fig. 15: *Original stimulus and its rotated variation (rotation by 90º)*
Source: Adapted from Cooper, (1975).

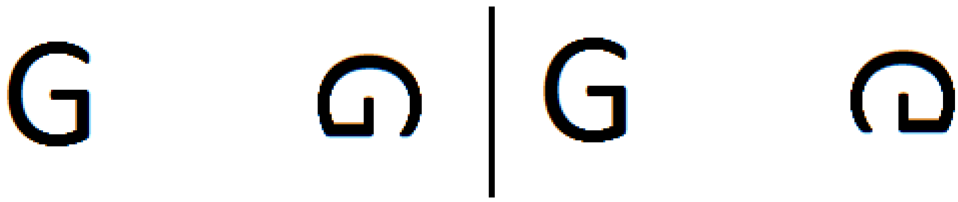


Fig. 16: *Alphanumeric stimuli*
Source: Adapted from Cooper & Shepard (1973).

Not only Cooper's (1975; 1976) experiments show more or less the same pattern[6] in results as Shepard and Metzler (1971), but the author also argues that her results suggest *"... an analogue nature of mental rotations"* (Cooper, 1976, p. 301). This claim was largely criticized and tested experimentally. For example, Xue *et al.* (2017) worked with eye-tracking (similar to Just & Carpenter in 1976) and tried to show that the nature of mental rotations is discrete instead of analog (holistic). They support Just and Carpenter's (1976) notion of rotations being processed piece by piece, using so-called *piecemeal strategy* (Xue *et al.*, 2017; Just & Carpenter, 1976; Pylyshyn, 1979). Xue and colleagues' research support this claim. They suggest a five-stage strategy in performing mental rotations contrary to the three suggested by Just and Carpenter (1976; Xue *et al.*, 2017).

6 In experiments with alphanumeric stimuli results showed that familiarity of an object prior to the testing (such as familiarity with an alphabet or numbers and their visual form) can play a role in final results. The further the object is rotated from its known form, the longer the reaction time. It means that the reaction time is prolonging to approx. 180° rotation and shortens by getting closer to 360° (Cooper & Shepard, 1973; Shepard & Metzler, 1988). Hock and Tromley (1978) explain these results and claim that participants determined the upright position of the letter/number prior to the rotation and rotated the object through the minimum angle to reach the upright position, which is consistent with Cooper and Shepard's (1973) explanation that the rotation is not required, if the object is tilted only very little from its upright position. In those cases the object is still perceived as upright even though it is rotated from its upright position (Hock & Tromley 1978; Cooper & Shepard, 1973).

Shepard and Metzler (1988)[7] used the polygons from Cooper's (1975) study and the three−dimensional blocks from Shepard & Metzler's (1971) study to show the difference between two – and three−dimensional stimuli. They were also interested in the difference between the one stimulus task, where the participants are presented with the standard two−dimensional stimulus and afterwards with test objects one by one and they then have to decide whether the test objects are the same or different than the previously presented standard stimulus (Cooper, 1975; 1976) and the two stimuli task, where the participants are presented with two objects at the same time and they have to decide whether the objects are the same or mirror images (Shepard & Metzler, 1971). The results were consistent with previous findings of Shepard and Metzler (1971).

The number of participants in studies of mental rotations in the seventies is interesting. They are surprisingly low; some studies only have four (Hock & Tromley, 1978), others eight participants (Shepard & Metzler, 1971). The question stands if the results would be any different if there had been more participants. These studies were replicated many times with similar (or at least comparable) results.

2.6.3 Testing the spatial orientation factor

Now, after explaining the spatial ability factors and introducing the first one, the mental rotation factor, the orientation factor is presented in more detail.

Spatial orientation ability can be measured by the Object Perspective Test (Kozhevnikov & Hegarty, 2001), where the participant is presented with a layout of objects and has to imagine standing at the layout with specific tasks based on his/her position (such as: Imagine standing in front of the house, facing the STOP sign. Now point at the tree; Fig. 17).

In the version of the Money Standardized Test of Direction Sense (Money & Alexander 1966 as cited in Zacks *et al.*, 2000) the participant follows a root on a map and has to decide whether each of the 36 marked turns is a left or right turn without turning the map itself.

7 Interestingly, the authors are a different Shepard and a different Metzler than the authors of the famous article from 1971. This coincidence is quite amusing.

Fig. 17: *Example of the Object Perspective Test*
Source: Kozhevnikov & Hegarty (2001) with permission.

One trial of the Pictures Test (Hegarty & Waller, 2004) consists of four pictures. The first one represents three people (A, B, C) each taking a picture/photograph of their surrounding from a different perspective. The second, third and fourth picture is a "photograph" representing the surrounding taken supposedly by one of the people on the first picture. The question is which person (A, B, C, or neither) took each of the photographs (Hegarty & Waller, 2004).

In Guilford-Zimmerman's Test for Spatial Orientation, the participant sees a picture of the bow of a boat s/he supposedly stands on. The task is to choose the correct direction of the boat after it changes course (Guilford & Zimmerman, 1956 as cited in Borich & Bauman, 1972).

The last, but not least, Water Level Task was used by Piaget to test the development of children's spatial skills (Piaget & Inhelder, 1948; 1956 as cited in Vasta & Liben, 1996). It is a test on the boundary between the mental rotation factor and the spatial orientation factor. The task is to draw a line reflecting the water level in a tilted container. The test requires a certain level of realization that water levels are always horizontal, independent of the inclination of the container.

2.6.3.1 Virtual reality in testing and training of spatial cognitive ability and diagnostics

Virtual reality (VR) is another way of testing and training spatial abilities that has emerged in recent decades and has been developing rapidly. Even though the VR tasks are mostly focused on the spatial orientation factor, spatial memory (Tascón, Castillo, León & Cimadevilla, 2018) and most commonly spatial behavior (Hardiess, Mallot & Meilinger, 2015), there are a few that focus on spatial visualization (e.g. Martín-Gutiérrez, Contero & Alcañiz, 2015; training of spatial visualization among engineering students, *see below*). As it is not a focus of this work to examine the possibilities of virtual reality in testing spatial abilities, only a brief overview of options about using VR is provided here. It is more practical and resourceful to base the test battery to examine the mental rotation factor and the perceptual speed factor on traditional tests such as mental rotations (Shepard & Metzler, 1971), as VR can be very expensive and usually requires a specialist to run the experiments (Hardiess, Mallot & Meilinger, 2015). Using VR would be a waste of resources for us, but it might be very useful for others.

Tascón and colleagues (2018) examined age and gender differences in spatial orientation and spatial memory tasks in VR while using a walking design (participants can move around in their environment in VR) and a non–walking design (participants explore the VR environment from a stable position). They found out that spatial orientation and spatial memory abilities decline with age. Men performed better compared to women, and the non–walking design requires more cognitive effort than the walking one. Martín–Gutiérrez, Contero & Alcañiz (2015) used augmented reality (AR) to train and test spatial skills in engineering students. They found out that VR and AR have a positive effect on learning engineering graphic content. Based on their extensive literature review Cogné and colleagues (2017) are suggesting that VR tools are better at assessing navigation strategies in patients with spatial orientation impairment (such as patients with schizophrenia, dementia or brain damage) compared to paper–pencil tasks and could be used in the neurorehabilitation of these patients.

Both, VR and AR have a promising future in the industry. Eschen and colleagues (2018) suggest using AR and VR in the aviation industry, especially during inspection and maintenance processes. Grajewski *et al.* (2013) propose using VR to plan and design manufacturing workspaces with the emphasis on ergonomics.

As it is not the main focus of this work, VR and AR are not discussed any further.

2.6.4 Testing perceptual speed

The **perceptual speed** factor (Miyake *et al.*, 2001) represents an ability to make a relatively simple perceptual judgment. Tests aiming at this factor measure individual differences in speed and efficiency. According to the authors, it is usually tested by the **Identical Pictures Test** (Fig. 18), where participants are required to identify a target between five options (Ekstrom, Dermen & Harman, 1976). The **Hidden Patterns Test** (Fig. 18) is also a test that loads highly on this factor. Participants are required to find a simple target pattern in a complex pattern. There are five options and participants have to choose only those complex patterns that contain the simple target pattern (Ekstrom, Dermen & Harman, 1976)[8].

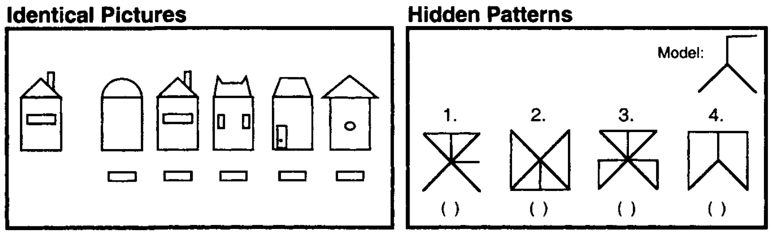


Fig. 18: *Example of Identical Picture Test and Hidden Patterns Test*
Source: Ekstrom, Dermen & Harman, (1976) with permission. Copyright©1976 Educational Testing Service. www.ets.org. Figures from the *Manual for Kit of Factor-Referenced Cognitive Tests* are reprinted by permission of Educational Testing Service, the copyright owner. All other information contained within this publication is provided by the author and no endorsement of any kind by Educational Testing Service should be inferred.

We find an interesting overlap here between the concept of hidden figures as a test of spatial cognitive ability and Witkin's concept of FD/FI cognitive style. As Ekstrom and colleagues (1976) state, flexibility of closure (here Perceptual speed, part of spatial cognitive ability) seems related to field independence, but the authors conclude, that there is not enough evidence for this claim. They advise readers to use the original Witkin's test (Embedded Figures test) to measure field independence. But is this really necessary? The concept of simple figures (patterns, objects) hidden in a complex scene (design) is a premise for both tests (EFT and Hidden Pattern test), and the authors themselves (both Witkin and Ekstrom) admit, their work derives from Gottschaldt Figures Test (Ekstrom, Dermen & Harman, 1976; Witkin, 1950). The EFT works on a premise, that the simple figure needs to be found in a complex design. Interestingly, the Hidden Patterns Test (according to Miyake *et al.*, 2001 also a test of perceptual speed as a part of spatial cognitive ability) works on a very similar principle, where participants are required to find a simple figure in a set of complex designs and mark those that have the simple figure embedded in them (see Fig. 18). The tests are based on a very similar principle, and there is no evidence they measure separate constructs. The main purpose of this work is to show a possible unnecessity of the FD/FI concept and to show that tests used to measure the level of field independence could be used as tests

8 Even though Ekstrom and colleagues (1976) point out in their manual that the tests mentioned above load on a factor called Flexibility of Closure, the authors admit that Flexibility of Closure is somewhat similar to perceptual speed (Tab. 4). Also Miyake and colleagues claim that tests of closure flexibility load on the perceptual speed factor (Lohman, 1988 as cited in Miyake *et al.,* 2001).

of spatial cognitive ability and its factors. The question is if the field independence test (EFT) is a test measuring the flexibility of closure (perceptual speed), which is according to Miyake *et al.* (2001) a part of spatial ability, maybe even part of the mental rotation factor of spatial ability. And that is what we are trying to find out here. Following Tab. IV presents an overview of relevant tests of spatial ability.

Tab. IV: *Overview of tests mentioned in the text used to test various factors within spatial cognitive ability. The aim is not to provide an overview of all existing tests, but of those that are relevant to this work. The question still stands, whether* **perceptual speed***, introduced in Miyake et al. (2001) is a lower-order factor within the mental rotation factor (together with spatial relations and spatial visualization) or whether it is a higher-order factor together with the mental rotation factor and spatial orientation factor. For more tests on cognitive abilities see Ekstrom, Dermen & Harman, (1976).*

Factor	Test	Author
Spatial orientation		
	Water Level Task	Piaget & Inhelder 1948; 1956
	Guilford–Zimmerman Test for Spatial Orientation	Guilford & Zimmerman, 1956
	Money Standardized Test of Direction Sense	Money & Alexander 1966. modified by Zacks *et al.*, 2000
	Object Perspective Test	Kozhevnikov & Hegarty, 2001
	Pictures Test	Hegarty & Waller, 2004
Mental rotation		
Spatial relations factor by Miyake *et al.*, 2001	Flag Test	Thurstone & Thurstone, 1941
	Card Rotation Test	Ekstrom, Dermen & Harman, 1976
Spatial Visualization factor by Miyake *et al.*, 2001	Space Relation Test	Bennet, Seashore, & Wesman, 1972 as cited in Miyake *et al.*, 2001
	Paper Folding Test	Ekstrom, Dermen & Harman, 1976
	Surface Development Test	Ekstrom, Dermen & Harman, 1976
	Cube Comparisons Tests	Ekstrom, Dermen & Harman, 1976
	The Vandenberg Mental Rotations Test	Vandenberg & Kuse, 1978
	Guilford–Zimmerman Test for Spatial Visualization	Guilford & Zimmerman, 1956
	Mental Rotation Test	Shepard & Metzler, 1971; Cooper, 1975
Perceptual speed		
	Identical Pictures Test	Ekstrom, Dermen & Harman, 1976
	Hidden Pattern Test	Ekstrom, Dermen & Harman, 1976

Source: Hegarty & Waller (2004); Miyake *et al.*, (2001).

3 QUESTIONABLE UNIVERSALITY OF COLOUR CATEGORIES

At first glance, this chapter on colour categorization and language may come as a surprise. The reason, why this chapter on linguistic relativity is a part of the thesis, which otherwise focuses on spatial cognitive ability, is a study by Gilbert *et al.* (2006). They examined the Sapir–Whorf hypothesis with a specific method which, according to us, could be partly a test of spatial ability. However, first, the Sapir–Whorf hypothesis needs to be introduced, and it is necessary to work from colour perception, colour categorization and testing the Sapir–Whorf hypothesis all the way to Gilbert's study, to be able to understand, why it is relevant to this work.

3.1 Visual perception of colours

To be able to grasp the research described in the following chapter in its full complexity, it is necessary to start with a brief description of colour perception in humans.

The importance of perceiving colours lies in two processes, **detection,** and **discrimination** (Eysenck & Keane, 2008). The colour of the object helps to detect an object (figure) in a complex background. This was mentioned in the chapter on figure-ground organization. Discrimination is used when objects or groups of objects are distinguished from each other.

The process starts in the human eye. The retina has two types of receptors. Rods focus on perceiving motion and focused vision in low lighting. Cones aim at perceiving colours. Cones are less represented on the retina than rods (approx. 1:18). Several authors described the process of colour perception. Two of the most accepted theories are described here, together with the attempt of combining their findings.

The Young–Helmholtz theory of colour perception suggests three types of receptors on the retina, each sensitive to a different light-wavelength. The S receptors (S for short, as in short wavelength) reacts to blue colour, the M receptors (M for middle) process the information from the green spectrum of colours, and L receptors (L for long) process red colour. The theory assumes that most of the colours are perceived by the combined work of all three, or at least two, receptors at the same time. Further research supports the theory of three separate visual receptors but does not assume that the receptors are this specific in processing only a certain wavelength, and therefore processing only a certain colour.

Hering's opponent–process theory assumes three pathways are playing a role in human colour vision. Each of the pathways secures the processing of two opponent colour categories (black – white; blue – yellow; green – red). This theory is supported by the fact that individuals do not report colours such as blue – yellow or red – green. Those colours seem to be opposite (Abramov & Gordon, 1994 as cited in Eysenck & Keane, 2008).

Even though these two theories provide different ideas on colour perception, there is a chance that they are not completely irreconcilable. **Hurvich** (1981 as cited in Eysenck, 2008) attempts to combine these theories into one complex explanation of human colour vision. He suggests that the detection of light and its wavelength is processed on the retina in line with the **Young–Helmholtz theory,** and further processing in the neural level is in accordance with the **opponent–process theory**[9] (Eysenck & Keane, 2008).

3.2 Sapir–Whorf hypothesis

Historical roots of the Sapir–Whorf hypothesis (S–W hypothesis) originate in work of Sapir's teacher Franz Boas in the 19th century (Kay & Kempton, 1984). Boas and his student Sapir studied unwritten non–European languages and showed (opposed to the popular view of 19th century anthropology) that these languages could be as rich as modern European languages. Whorf, Sapir's student, joined his teacher in the research on various languages and on the way to the linguistic relativity hypothesis (Kay & Kempton, 1984), which was formed based on their work and publications (as rumor has it, neither Sapir nor Whorf formulated the hypothesis themselves). The main claim of the hypothesis, both in its weaker and stronger version, is that structural differences in languages are parallel to non–linguistic cognitive differences not otherwise specified. In other words, the structure of a language influences (weak view on S–W hypothesis), or determines (strong view on S–W hypothesis) the world view of the native speaker of a certain language (Kay & Kempton, 1984). This hypothesis suggests relativity in an observer's cognitive perception of the world, and this relativity is more or less dependent on language structure[10].

9 For those who are interested in a neuroscientific take of the matter of colour perception, see Eysenck & Keane (2008).

10 A very good empirical example of a strong version of the S–W hypothesis is described in a short story by Ted Chiang *(Story of your life; 1998)*, where a group of aliens visits Earth to warn human kind of the upcoming apocalypse. Louise, the scientist and linguist, is learning their language which does not have a linear understanding of time. The more she learns the language, the more she internalizes the alien way of thinking, therefore, she is able to mentally time travel to the past and future. The story was turned into a motion picture named *Arrival* in 2016. starring Amy Adams.

The question that emerged from formulating the hypothesis was as follows: What is the best way to test linguistic relativity and its influence on cognition? Colour was apparently the best way to empirically test language relativity for multiple reasons. First of all, many colours are visible to the human eye, but there are not as many names for colours. Also, languages tend to differ in naming the colours. Children learn colour categorization relatively slowly. Colour thus seemed like a rich area suitable for research on the relation between language and thought (Roberson & Hanley, 2010). Colour is not the only domain providing possibilities for research on linguistic relativity. Casasanto (2008) reviews different names for time periods in various languages and different language expressions about time. Also, Gordon (2004) presents research on the S–W hypothesis from the Amazonian tribe of the Pirahã. They did not develop a counting system apart from words representing "one" and "two". According to Gordon (2004), members of the Pirahã tribe are strongly affected by their lack of a numerical system in specific numerical tasks, which Gordon considers proof for a strong S–W hypothesis (with criticism from Casasanto, 2005).

At first, studies seemed to support the S–W hypothesis, but empirical evidence started to shift and provided more evidence against the relativity hypothesis in the late seventies (Kay & Kempton 1984; Kay & McDaniel, 1978).

When one wants to present classical work opposing the relativistic view towards colour categories, one has to describe Berlin & Kay's research from 1969 (in Kay & McDaniel, 1978; Hardin, 2013). Their original research involved 20 languages, later more than 70 additional languages were examined. The original research methodology was based on choosing a colour chip from an array of stimuli based on the colour categories of a subject's language, which that subject had provided during an interview before the experiment. Subjects were supposed to choose colour chips which were the best examples of certain colour categories. The second task was to choose all the chips that belonged to a particular category. This way, the authors got colour category boundaries. The authors conclude that colour categorization is not random but has some universal tendencies among the languages examined (Berlin & Kay, 1969 in Kay & McDaniel, 1978). They identified eleven universal colour categories: black, white, red, orange, yellow, brown, green, blue, purple, pink, and grey.

Interestingly, they found out that if a language has only two colour categories, they are black and white. If there is a third category, it is red. The fourth category is either yellow or green (but not both). Five categories are represented by black, white, red, yellow, and green. If the language has six categories, blue comes into the picture. Seven categories would contain brown, and the eighth category is purple, pink, orange, grey or a combination

(Berlin & Kay, 1969 in Kay & McDaniel, 1978). There is some criticism of this methodology by Kay & McDaniel (1978) based on the fact that Berlin did not measure to what degree a certain colour belonged to a particular category, how close it is to the category norm (some hue of red would be considered more of a representative of *the red category* than some other hue of red)[11].

Even though there is some criticism of Berlin & Kay's classical work, Kay & McDaniel (1978) also claim in their article that colours have universal categories based on the biology of the human vision. They say universal basic colour categories can be derived from an ability of the human eye to see a certain wavelength, those categories are red, yellow, green, and blue. The authors later add two additional categories, black, and white. They claim that all the basic colour categories in all languages are one form of these fundamental universal colour categories (Kay & McDaniel, 1978).

Interesting insight consistent with Kay & McDaniel's view (1978) is presented by Lindsey & Brown (2002). They examined languages that do not linguistically distinguish between blue and green (known as "grue" languages). They investigated languages that use different words for blue and green, and languages that do not distinguish between these two short-wave stimuli. The authors claim that some linguistic categories (or a lack of them in certain languages) might be a direct consequence of a different sensitivity to short-wave light among people because of different exposure to a certain sort of wavelength in specific parts of the earth. Lindsey and Brown (2002) claim that people who live in parts of the earth where they are exposed to ultraviolet – B light to a high degree are prone to abnormally high optical density in the lens (yellow pigments accumulating in the lens as it ages, which causes blue light to seem greenish). Their analysis of 203 languages showed that "grue" language cultures live close to the equator (high UV−B exposure) and blue/green language cultures live in higher altitudes (relatively low UV−B values), of course with differences of some sort between individuals within a culture. The authors say that these results show that linguistic relativity is not needed to explain why certain cultures have different colour naming systems than others. Language categories seem to be formed based on perception specifications caused by the physiology of the eye (Lindsey & Brown, 2002).

11 For further reading on research examining the fuzzy set theory of colour categories that deal with the notion of unspecific and inexplicit boundaries between colour categories, see Kay & McDaniele (1978).

Kay & Kempton (1984), too, conducted an experiment with participants speaking two different languages, a "grue", and a "non–grue" language: English, with the semantic category green distinctive from the category blue ("non–grue" language), and Tarahumara (a language spoken in northern Mexico) with only one *siyóname* category for green and blue ("grue" language). The subjectively perceived difference between various blue and green hues was measured. According to the S–W hypothesis, English speaking subjects would perceive green and blue as more distinct, while Tarahumara speakers would not perceive this difference. The English speaking participants really tended to perceive green and blue as more distant from each other than two hues of blue or green. The results were inconclusive among the Tarahumara speakers. These results support the S–W hypothesis in a way that if there is a language colour category in a certain language, then a native speaker seems to perceive the distance between the colours of the two categories as bigger than native speakers of a language that does not have that categorization (Kay & Kempton, 1984). A limit of this study, which could weaken the interpretation of the results, is the number of participants in each language group (there were 5 English speakers and 4 Tarahumara speakers). The authors themselves argue that the results could be an effect of participants more or less consciously selecting colours as distant from each other because they would name them differently (blue vs. green), even if participants would not cognitively perceive a bigger difference between blue and green compared to two green hues or two blue hues (*the name strategy*). In a second experiment, Kay & Kempton (1984) wanted to eliminate the effect of the name strategy, so they did not ask about subjectively perceived differences between colours but presented three colour chips to the participants and asked them which of the three was the most different. The colour presented in the middle was always on the edge of the green/blue colour boundary, either on the blue side or the green side. The S–W effect from the first experiment disappeared among English speaking participants. They did not perceive two blues as closer compared to the green, nor did they perceive two greens as closer compared to blue. This result suggests that the cognitively perceived distance between colours from different language categories is no bigger between languages having and not having these categories. There could be an effect of knowing a name of a colour category in rating colours and subjectively perceived differences between them, which might be interpreted in conformity with the weak S–W hypothesis. This is consistent with Wolff and Holmes' (2011) conclusion, which says that language might be a powerful tool that somehow affects thinking (for example, its effect on mathematical tasks; Gordon, 2004), but they do not support linguistic determinism.

3.3 Testing of the Sapir–Whorf hypothesis by Gilbert *et al.* (2006)

In 2006. Gilbert and his colleagues tested the S–W hypothesis using a very interesting methodology (Gilbert *et al.*, 2006; Rieger & Kay, 2009). Part of a study presented in the empirical part of this thesis was inspired by Gilbert's article, so this methodology will be described here for better understanding. First, participants rated colour chips as either blue or green. This way, the colour boundary between green and blue was set for further experiment. After this, participants were presented with a circle made of 12 colour chips, eleven of them were of the same colour. A chip of a different colour represented a target. The participants' task was to determine as quickly and as accurately as possible if the target was on the left or the right side of the circle (Fig. 19).

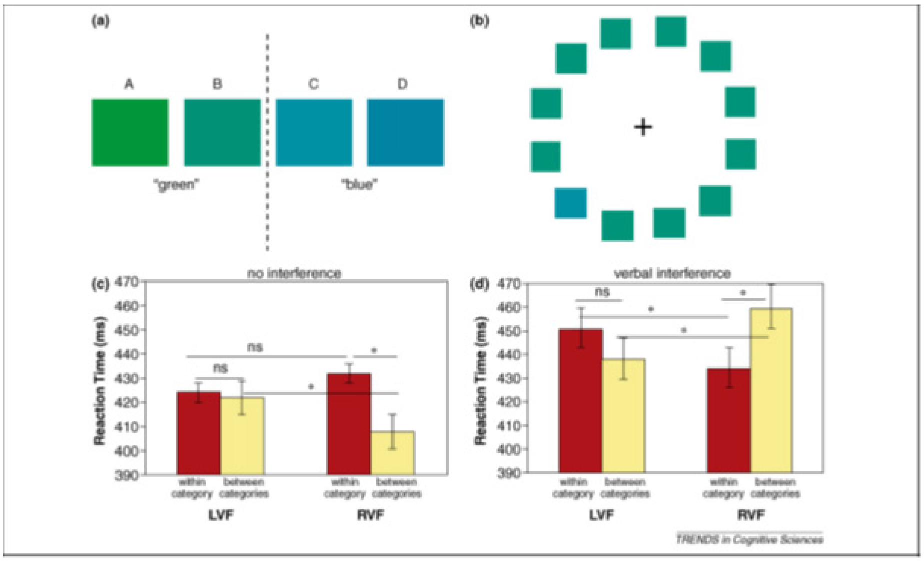


Fig. 19: *Example of the Gilbert at al. methodology, and results for tasks with and without verbal interference.*
Source: Gilbert *et al.*, (2006) with permission.

All the pairs of colours (A, B, C, D) from Fig. 19 were used to test whether the participants' reaction time would be shorter for a colour that crosses the colour boundary (according to the S–W hypothesis, subjects are supposed to perceive differences between colours from different semantic categories easier and faster than between those from the same colour category; Gilbert *et al.*, 2006). The results show that the authors' hypothesis was correct, but only for the right visual field. They interpreted these findings as a result of the brain regions for language (Broca's and Wernicke's area) being located in the left brain hemisphere. The left hemisphere has a direct connection to the right visual field, and that is why the results were only conclusive for the right visual field. In the second part of the study, the participants had to complete a verbal interference task while working on a focal task, and the effect disappeared. The authors interpreted this result as supportive of their original interpretation. If the brain regions for language are occupied by some other task, they do not seem to affect the colour categorization and recognition task. Gilbert *et al.* (2006) present this study as evidence in favor of the S–W hypothesis, but only for the right visual field.

Winawer *et al.* (2007) report similar results for Russian and English speaking participants, although taking a different approach. The authors had their participants decide in several trials which one of two colour chips matched the standard one. They found out that for the Russian speaking participants, there was a difference between the trials including colours crossing the colour boundary (Russian distinguishes between *dark blue* and *light blue*; *siniy* and *goluboy,* respectively), and the trials that included colours from the same colour category (either all of the colours were *siniy,* or all of them were *goluboy*). This difference was not found among English speaking participants (in English, dark and light blue is considered a single colour category: *blue*). Also, verbal interference did play a role in Russian speakers. In trials with verbal interference (repeating a set of numbers out loud), the difference disappeared. Non–verbal interference (spatial one, where participants had to remember a certain object in their mind during the trial) did not have any effect. Neither interference had any influence on the English speaking participants' performance in any conditions. The authors' results are in accordance with Gilbert's findings, even though this study did not focus on the difference between the right and left visual field. The conclusion of this study could be that the distinction between colours that do not belong to the same linguistic category is easier than that between colours that are from the same linguistic category; but again, *the name strategy* could play a role (Kay & Kempton, 1984).

Drivonikou in cooperation with Kay, Rieger, Ivry, Gilbert and others (2007) replicated the effect found by Gilbert *et al.* (2006), where the participants were able to detect a target faster in the right visual field if the target was from a different colour category than the distractors. On the other hand, they found the same (but weaker) effect for the left visual field while Gilbert's study (2006) showed no or very little effect on the left visual field.

Gilbert, Regier, Kay & Ivry (2008) interestingly tried the same method of left – right decision about the target position but with different stimulus material. They used black silhouettes of cats and dogs on a white background and found the same results. In the cross–category trials (dog among cats/cat among dogs), participants showed faster reaction times in the left – right decision about the target position in the right visual field, while the effect disappeared in verbal interference trials.

Franklin *et al.* (2008) used eye tracking to find similar results as in the studies described above, but in children, who have already learned colour categories (faster differentiation of colours from different categories in the right visual field), while children only learning colour categorization at the time of the experiment, the authors found a reversed effect (faster differentiation between colours from different categories compared to colours from the same category but for the left visual field). The authors (Franklin *et al.*, 2008) conclude that their results suggest developmental changes in the localization of categorical perception of colours, and they consider their findings a proof of language's influence on perception (weak S−W hypothesis).

The main concern here is Gilbert's approach towards this problem. The question we asked in the studies presented is whether his methodology really tests the S−W hypothesis in colour categorization. The tasks (left – right judgment) are, among other things, also based on spatial orientation and seem to be closely linked to the perceptual speed factor of spatial cognitive ability (*see below*).

3.3.1 Left – right judgment tasks

Since Gilbert and colleagues used left – right judgment for their experiment, this chapter tries to clear what left – right judgment task is really able to measure. It is not easy to find a clear explanation of what the left – right judgment would be a measurement of. It seems that a big part of research uses so-called left – right judgment tasks (LRJT) to elicit motor imagery in patients with chronic pain (Alazmi, Gadsby, Heneghan & Punt, 2018; Breckenridge *et al.*, 2017), identifying dysfunction in body maps and rehabilitation (Zimney *et al.*, 2018), or even test mental rotation (Cooper & Shepard, 1975). It needs to be said here that these LRJTs are not a typical left – right decision tasks comparable to Gilbert's task. The usual goal of LRJT is recreating a motor imagery. Its method usually consists of decisions whether displayed body parts have a left or a right orientation. In the mental rotation study, the authors (Cooper & Shepard, 1975) presented pictures of hands (left/right) in various rotated positions and asked their participants, whether they see the left or the right hand. These studies are not based on the decision of the position of a target (left – right decision). However, we have to say that there could be a certain link to the Body−Adjustment Test, and overall the tasks testing the upright position because they also derive from a body map and body position imagery.

A very interesting study suggests that left – right decision could be linked to spatial information, especially in blind individuals (Tinti *et al.*, 2018). Based on the idea that blind individuals can make spatial representations from auditory information from the environment, the authors had blind participants (together with a control group of healthy yet blindfolded individuals) examine a tactile 3D map and afterwards provided sound information about the position of the observer from either the left hand−side, right hand−side or from the front. They found out that auditory information can trigger a certain perspective taken from the origin of the sound in blind as well as in healthy individuals (Tinti *et al.*, 2018).

Imai (2017) talks about languages that do not have *left – right – behind – in front* spatial descriptions in their repertoire and therefore do not use these terms to describe the spatial position of an object or a direction towards an object (e.g., Tenejapa language; Mexico). She describes an experiment conducted by Brown and Levinson (1992 as cited in Imai, 2017) where the authors used a version of mental rotations to test whether the participants speaking the Tenejapa language would be able to detect the difference between pictures and their mirror images. Interestingly, Tenejapa speakers did not report any difference between objects and their mirror images compared to the Dutch speaking participants. The authors together with Imai (2017) explain this result with the lack of a left – right language category in Tenejapa, therefore, their lack of attention to the left – right orientation as an important feature of a perceived visual information. Here, we see that Imai and others (e.g., Brown & Levinson, 1992 as cited in Imai, 2017) also understand left – right orientation as a part of spatial orientation skills and abilities that, together with measuring the reaction time, could be a part of the spatial ability factor *Perceptual speed* (see Tab. V).

Even though the study described by Imai (2017) is not based on completely the same design as Gilbert's study, we can sense that there is a certain spatial information included in left – right decisions. The question of why Gilbert and colleagues used this particular way of testing their hypothesis about language categories remains unanswered.

3.3.2 Motivation to enclose Gilbert's study into the method and its link to the rest of the thesis

The first reason why Gilbert's method was included in this work is the aim to replicate his study with some slight changes that could improve the statistical strength of the results. The main question is whether the results will be the same with more participants and with a slightly different way of choosing the colours. As we did not use the same participants and exactly the same conditions as Gilbert (exact replication), nor did we use the exact conditions with a different participant pool (approximate replication), we would consider our replication a conceptual one (more on replications in Appelbaum *et al.*, 2018; Mackey, 2012). That means, we try to test the same concept, expect results that can be compared with the original study, but we use a different set of stimuli and different participants.

The second reason for using Gilbert's method as a part of this work is its slight resemblance of tasks used to test spatial ability, specifically the perceptual speed factor of spatial ability. Miyake and colleagues (2001) define perceptual speed as the ability to make relatively simple judgments and they introduce so called Identical Picture test, where the task is to spot the target between distractors (see Fig. 18), which is basically what Gilbert asked his participants to do in his study. Ekstrom, Dermen and Harman (1976) call the perceptual speed factor *flexibility of closure* and use two tests to measure it – the Identical Picture test and Hidden Pattern test, the latter is very similar to Witkin's EFT. So, it seems that Gilbert's design, where participants have to complete perceptual tasks of spatial nature (*left – right judgment about target position*) as quickly as possible with the greatest efficiency is in fact a task saturating the perceptual speed factor of spatial ability, while the perceptual speed factor has a clear link to EFT by Witkin (1950), as we can see in Ekstrom, Dermen & Harman, 1976). The remaining questions are still pending. What factors are saturated by which tests, and are all these tests loaded onto the general spatial ability factor?

4 EMPIRICAL PART
4.1 Methods

The studies conducted as a part of this thesis are based on three subtests. First, there is a Witkin's EFT – based Hidden figures subtest, second a Mental rotation subtest, and third a subtest based on Gilbert's study described above (chapter 3.3). All three parts are described and explained in this chapter.

4.1.1 Hidden figures subtest

It has been written a great deal about Witkin's Embedded Figures test, which served as a model for the subtest of Hidden figures in this work. The original Witkin's EFT (GEFT) is not available for research purposes due to copyright reasons. The company owning the copyright provides a group electronic version of the test in its 24 item form, but after some communication, it has become clear that the results consist only of the number of correct and incorrect responses for each participant. Reaction times for individual trials are not available. This, together with a rather high price, are the reasons why we have not used the original Witkin's EFT. The test used here is very similar to what we did in previous work (Krajčíková, 2014), where we created trials ourselves. Dr. Šašinka is a co−author of the pictures used in the Hidden figures subtest in this thesis. The pictures were created strictly according to Witkin's instructions. The task is to find a simple figure in a complex colourful design. The simple figure is always embodied in the complex design, it is of the same size and the same orientation. There are no trials, where the simple figure would not be embodied in the complex design. This information is clear to the participant upfront. The example of the material can be found in the Attachment 1.

We chose a 12 item shortened version of the Hidden figures subtest because the short version of EFT shows the same psychometrical characteristics as the original 24 trial version. Also, we did not only plan to work with the number of correct answers for each participant, but also with the reaction time for each trial. In the end, however, we had to eliminate one figure of the final 12 item test because of its resemblance to one of the original Witkin's figures. Thus we were left with eleven trials, each consisting of a simple figure and a complex design (Fig. 20).

The participant was instructed to find a simple figure in a complex design[12]. The trial consisted of a simple figure presented on a computer screen together with a button "Start task". As soon as the participant clicked on the button, we started the measurement of the reaction time. After the participant had clicked on the "Start task" button, s/he was presented with a simple figure and could move (by clicking on a button) to the complex design where s/he was instructed to find the simple figure (presented right before) using a polyline, which could have been created using a mouse. The participants learned how to create a polyline on a test trial. The participant had a chance to switch back to the simple figure by clicking on a button "Back to the simple figure", which appeared above the complex design. In this case, the participant could see the simple figure again and then go back to the complex design using a button "Back to the complex design" above the simple figure. The reaction time measurement was still running. The reaction time measurement stopped as soon as the participant clicked to the complex design (the sign that s/he started to create the polyline). We also measured accuracy together with the reaction time for each trial. The instructions about switching back and forth between the simple figure and the complex design together with the fact that the measuring of the reaction time would still be running was explicitly communicated to the participants in written instructions beforehand. Also, they knew that if they were unable to find the figure in the design within five minutes, they would be automatically moved to the next trial. Witkin's original instructions gave a maximum of three minutes for each trial, but we decided to change this to 5 minutes as we expected some trials to be more difficult than the original Witkin's ones. Since we did not intend to compare the results with Witkin's norms, this should not be an issue.

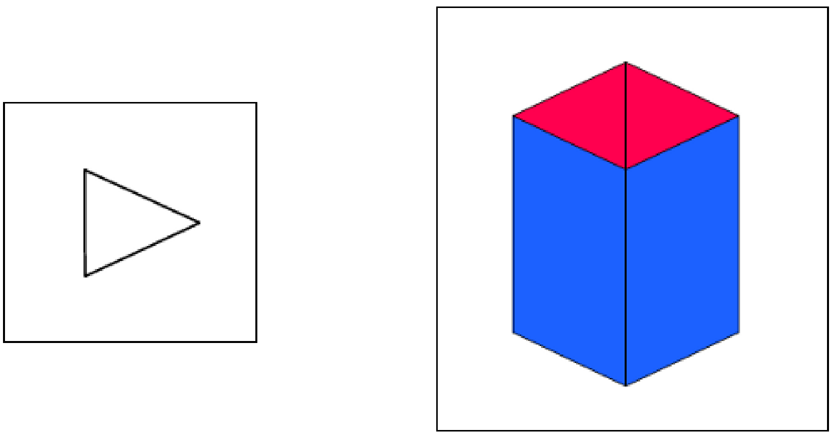


Fig. 20: *A test trial of a hidden figures subtest – a simple figure and a complex design.*
Source: Krajčíková, (2014).

12 All of the instructions for all the subtests were in Czech language. Some instructions are translated to English here for explanatory reasons.

4.1.2 Mental rotation subtest

The mental rotation test was adapted from Woodcock–Johnson's test of Cognitive Abilities (IIIrd edition by Woodcock, McGrew, & Mather, 2001; Woodcock, McGrew, Mather & Schrank, 2003), specifically its visualization subtest, which consists of two parts, one part operating with 2D objects (spatial relations) and another part operating with 3D objects (mental rotations). The test was administered on a computer screen as a part of a battery together with the Hidden figure subtest and a Target position subtest. Each trial consisted of a model object and five options. The task for the participant was to click on two options that represented a rotated model object. The trial was considered successful only if both the options were selected correctly. Once the participant had made a decision and clicked on an option using a left mouse button, the decision could not be changed. Participants knew this beforehand from the instruction page. They were also instructed to complete a training item to try out the functionality and to get familiar with the task. The training trial was not part of the analysis. There were overall 24 trials. The first three test trials were simple detection tasks, where the participant was instructed to click on a picture of an object that is similar to the model object (geometrical 2D shapes such as square, triangle, rectangle). These trials resemble the Identical Picture Test (Ekstrom, Dermen & Harman, 1976) that measures the perceptual speed factor of spatial ability. Otherwise, the rest of the items resemble Shepard & Metzler's stimuli material (1971) and Vandenberg & Kuse's version (1978) of picking two options out of four (in our case five).

For copyright reasons, we could not present an actual example of a stimulus from the original Woodcock–Johnson test of Cognitive Abilities here, as we signed the contract with the authors working on a Czech validation of the test, which says that the material would be used solely for academic purposes and would not be published (not even the example of it) in any author's or third party's work.

Woodcock and Johnson's Test of Cognitive Abilities IIIrd Edition (W–J test) is an intelligence, ability, and achievement test based on the C–H–C theory of cognitive abilities (Taub & McGrew, 2014). The C–H–C– theory is discussed in chapter *2.3* where it says that one of its main factors and many of its sub–factors are directly linked to spatial cognitive ability.

4.1.3 Target (left – right) position subtest

The subtest based on Gilbert's study (Gilbert *et al.*, 2006) required its own pilot study. As it was established before, this subtest is a conceptual replication of Gilbert's study. It is typical for replications to try to test a certain methodology on different, and, most importantly, more participants compared to the original study. That was our goal, too. Not only did we focus on a greater number of participants but also on a different way of choosing the stimulus material.

The task was to spot a target between eleven distractors. The target was a square, which had a different colour than the rest of the squares (the distractors). The squares were arranged into a circle with a fixation cross in the middle (Fig. 21). The task was to respond as quickly as possible (and of course correctly), on which side of the circle of squares the target is located. It was a left – right judgment task, where participants responded pressing a corresponding key on a keyboard ("Q" for left and "P" for right). Our goal was to find out if the judgment would be faster for target–distractor pairs if the target colour and the distractor colour is from a different colour category (in Gilbert's case green/blue boundary).

The subtest consisted of 144 trials. Each trial contained a target and eleven distractors, which means that in each trial, one colour was the target and one colour was the distractor. That made 12 colour pairs. Also, we wanted the target to randomly appear on all 12 positions around the circle (12 × 12 = 144 trials). The sequence of colour pairs and the position of the target within the circle were randomized for each participant.

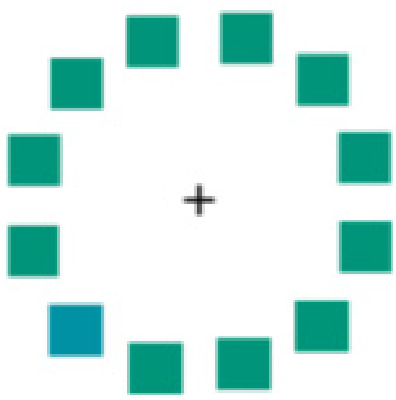


Fig. 21: *Example of a trial in the Target position subtest. The correct response is "left" (pressing the "Q" key on the keyboard) with the target being at the lower left–hand side (the approximate position of the number 7 on a clock)*
Source: Gilbert *et al.,* (2005) with permission.

As of Gilbert and his colleagues, they only had eleven participants taking part in their study. We planned on increasing this number. The problem was that with Gilbert's eleven participants, he was able to find out quickly and easily, where the colour boundary lies for each participant. Gilbert let the participants draw a line between a green and a blue category. As we wanted to increase the number of participants, we chose a different way.

4.1.3.1 Choosing the colour categories: Pilot study I.

For the replication of Gilbert's study, we needed to find those colour categories that would make most of the participants agree on the boundaries between the categories.

If the majority of the participants in the pilot study agreed on the categories for the chosen colours, and therefore the colour categories could be claimed to have clear boundaries (where one category ends, the other one starts for most of participants), these categories would be used for all the participants in the main study. The colours chosen for the pilot study are presented in the Attachment 2 together with their RGB colour code. The RGB distance between each two colours next to each other were standardized. The colour brightness (value, refers to brightness or darkness of colour) and saturation (chroma, refers to "purity" of colour; how "dull" or "sharp" the colour is) were standardized.

Method

The pilot study was conducted for two days in May 2016 at the HUME lab II at the Masaryk University, Brno. The laboratory room is on the underground level with no windows, so there was a standard lighting with no natural light changing during the day, which may have caused changes in the visual perception of the colours presented on the computer screen. The computer screen had been calibrated by a professional using the colour calibration device Spyder 3 (Elite version) by Datacolour company. The participant sat in front of the computer screen in a standard distance from the screen. The chair was always adjusted so that the participant was facing the screen perpendicularly.

Psychology students participated in the pilot study (N = 34, 10 male). The study consisted of an introduction and instructions in Czech, one test trial to introduce the task to the participants, 44 trials, and of a couple of questions about demographic characteristics of the participant. Each participant was given information about the study in written form and signed an informed consent before the study. The test took approximately 10 minutes. The participants were tested individually in the presence of a research assistant.

Each trial started with a fixation cross, then a square of a specific colour hue appeared together with six colour categories written under it. The categories were: brown, yellow, red, green, blue, and purple (written in Czech). After the participant had chosen one of the categories and clicked on the word representing the category with the left mouse button, another trial started. After all 44 trials, we thanked the participant for his/her help and s/he then left.

Results

The results showed the most assigned category for each colour. Gilbert's study used green/blue categories.

It is necessary to point out here that when we test how fast and how correctly participants can discriminate between colours with an emphasis on certain colour pairs (those crossing the colour category boundary) being distinguishable more quickly than others (those from the same colour category), we have to think about the fact that the colours we use should have at least approximately the same distance between each other in colour space.

First, we transferred the RGB code of each colour to the so-called CIE−L*ab colour system (i.e., Munsell colour system) using easyrbg.com[13] (the same as Gilbert *et al.*, 2006). Then we calculated a stimulus distance between two colours that our participants had assigned to a certain colour category with more than 94% agreement. We found out that the blue/green boundary is too far from each other in colour space, which would mean that two blue hues and two green hues would be much closer to each other than the closest blue and green hues. That could radically affect our results in favor of the colours from the same category. This is why we had to choose colours that were agreed on to at least 94 % and that are approximately the same distance from each other in colour space (Tab. V). We calculated the colour distance in CIE−L*ab colour space using the CIE76 formula and an online calculator[14].

The table 5 shows the chosen colour hues, their RGB code, CIE−L*ab coordinates and the final distance between each pair of the colour hues.

13 EasyRGB. (n.d.). Available from: https://www.easyrgb.com/en/.
14 Delta−E Calculator. (n.d.). Available from: http://colourmine.org/delta−e−calculator.

Tab. V: *Characteristics of colours chosen for the Target position subtest.*

r	g	b	Colour hue	Agree.%	L*	a*	b*	Distance
102	149	164		100 **blue**	58.974	– 11.914	– 12.921	11.8992
118	141	170		94.1 **blue**	57.85	– 1.189	– 17.951	7.9244
136	136	166		97.1 **purple**	57.651	6.447	– 15.842	8.1414
152	132	156		100 **purple**	57.635	12.107	– 9.99	

Obviously, the distance between the first two colours, two blue hues, is larger than the distance between the other colour "neighbours". This occurred in Gilbert's study as well *(distance between the green hues was 12.3, distance between green and blue hue at the border of the categories was 12.5, and the distance between two green hues was 15.7;* Gilbert *et al.*, 2006, p. 493, footnote*).* The authors, therefore, analyzed the data separately for the green and the blue category and compared if the data within the blue category was different from the data within the green category. They did not find any significant differences and did not consider this a problem in their further analysis. We will proceed in a similar way.

After this procedure, we had the colours chosen. We were able to continue with the subtest itself.

4.2 Procedure

The data collection took place in the HUME Lab II. laboratory at the Faculty of Arts in Masaryk University in Brno during two weeks in April and May 2018. Initially, the testing was supposed to take place during the summer semester only, but as we found out shortly after the advertising of the testing among students, we had not found enough participants.

The first wave of participants were recruited by email sent by the author and her colleagues mostly to students from courses taught by the author, or the author's colleagues in the past in the Psychology Department of the Faculty of Arts. Also, recruitment happened by word of mouth among the author's friends and acquaintances. The reasons why only 50 participants volunteered in this period is unknown. We assume it was the fact that the testing took around 45–60 minutes in a laboratory (it was not possible to let participants do it at home, as we were using computers with calibrated monitors, and a room with a standard lighting), and participation was not paid or rewarded in any other way (except a great deal of "thank – you" from the author and her research assistants). As this is a common problem, we decided to launch a second data collection at the beginning of the winter semester 2018 (two weeks in September and beginning of October 2018) with the help of the HUME Lab infrastructure, which provided us with a list of possible participants (all in line with GDPR). The second wave brought another 63 participants, therefore at the end we could work with a complete 113 participants.

The data collection took place thanks to three research assistants (RA) working separately, one at the time, who were provided by HUME Lab and were trained to do the job as a RA. After receiving an email about the testing explaining the data collection and the privacy issues and voluntary participation[15], the potential participants could make an appointment at the lab on a Doodle document created by the RAs. The list of participants and their email addresses were only available to the RAs, not even the author had access to the personal demographic details about the participants. The participants who signed up for a session received an email two days prior to the scheduled meeting about where and when to be present. Participants had an email address of the RA responsible for the testing in case they wanted to cancel their session.

After the participant arrived, s/he was introduced to the lab by the RA and handed the "Information about the study" document and the "Informed consent form" in writing. They were provided with the author's and the ethical board's contact information (email and telephone number) for possible further inquiry or complaints. After signing the document, the participant was asked to sit in front of a computer and was instructed to read all the instructions provided. The instructions for each subtest were embedded directly in the test, so they appeared on the computer screen. All the instructions were in Czech (as the author is not a Czech native speaker, the instructions were corrected by a professional).

15 This research was approved of by the Ethical Board of Masaryk University on the 21st of August in 2016; "Etická komise pro výzkum Masarykovy university").

The laboratory setting was the same as during the pilot study described above. The room with no windows and standard lighting provided standard conditions.

During the first stage of planning the data collection, we intended to test two participants at the same time as we had two computers available. The computers are the same (especially the monitors are the same) and are next to each other. There are two light sources in the room, one above each of the monitors, which provide the same lighting for both monitors.

Later, we realized that we would not have enough participants to test them only in pairs, but that there would be participants tested individually. The most serious influence could have been a certain social influence (social facilitation). This issue is addressed in the *Limitation* section of the thesis.

As was said before, the testing took approximately 45–60 minutes, about which the participants were informed beforehand. The time varied because of the participants' differences in reaction time. The participants had very detailed instructions available together with a test trial for each of the subtests. The subtests' order was as follows: Target position subtest, Mental rotation subtest, Hidden figure subtest. The order was standardized across participants.

After the testing, participants were thanked by the RA and left.

4.3 Research questions and hypotheses

RQ1: How strongly does the standardised test of spatial ability (Mental rotations subtest) correlate with the EFT–like test (Hidden figures subtest)?

RQ2: How strongly does the Target position subtest (based on Gilbert's method) correlate with the rest of the tests (Mental rotation subtest and Hidden figures subtest)?

RQ3: Can the EFT–like subtest (Hidden figures subtest) be considered a part of spatial cognitive ability based on its correlation with the Mental rotation subtest?

RQ4: Are there any gender differences in performance in each of the subtests?

RQ4: Are the findings published by Gilbert and colleagues (2006) replicated after certain alterations of the methodology?

H1: The performance (reaction times) in the test used to measure spatial ability (Mental rotation subtest) correlates with the performance in the Hidden figures subtest. The correlation is moderate.

H2: The performance (reaction times) in the test used to measure spatial ability (Mental rotation subtest) correlates with the performance in the Target position subtest. The correlation is weak.

H3: The performance (number of correct trials) in the test used to measure spatial ability (Mental rotation subtest) correlates with the performance in the Hidden figures subtest. The correlation is moderate.

H4: The performance (number of correct trials) in the test used to measure spatial ability (Mental rotation subtest) correlates with the performance in the Target position subtest. The correlation is weak.

H5: The performance (reaction times) in the test Hidden figures subtest correlates with the performance in the Target position subtest. The correlation is weak.

H6: The performance (number of correct trials) in the Hidden figures subtest correlates with the performance in the Target position subtest. The correlation is weak.

H7: There is no difference in performance between male and female participants in any of the three subtests.

The replication of Gilbert *et al.,* 2006 study:

The results we hypothesize here are in accordance with Gilbert's (2006) results for the purpose of the argument. The replication should anticipate that the results we are trying to get are in line with the study we are replicating.

H8: The reaction times in trials with a target and distractors from different colour categories show shorter reaction times in detecting the target's position compared to the trials with a target and distractors from the same colour category, but only in the right visual field.

4.4 Participants

The participants were recruited by word of mouth from the author's social circles as well as from those of students of Masaryk University. This is reflected in the heterogeneity of certain demographic characteristics (such as age). One hundred thirteen individuals participated in the study. There were 72 female participants (63.71%) and 41 male participants (36.28%). At the beginning of the study there was a plan to compare the participants from the technical fields with those from the humanities, however, after finding that only 16 participants (14.15%) were from technical fields (mathematics, IT), we decided not to perform separate analyses on such a small sample. The rest of the participants were from study fields such as psychology (36.28%), various linguistic studies (22.12%), and other, mostly humanistic, fields (pedagogical studies, study of religion, etc.). The average age of the participants was 22.99 years (*Sd = 4.48; max = 46; min = 19*). Approximately one−third of the participants (39; 34.51%) had a sight impairment, which was corrected with glasses or contact lenses; participants were explicitly asked to use them throughout the whole testing session. There were no colour−blind participants.

There are certain changes in the data sets used in separate analyses, where we had to exclude certain participants. However, these steps are described separately in each analysis that is affected. The reason for this is that some analyses only concern one of the subtests and we did not want to exclude participants that failed to complete the other subtests but were successful in this particular one. The exclusion of participants is carefully examined and described in each separate analysis.

4.5 Preliminary data editing

There were some concerns about the data that had to be resolved before the main analyses.

4.5.1 Mental rotation subtest

The first problem that surfaced was the fact that the order of the trials in the Mental rotation subtest was from the easiest to the hardest, most difficult task. This could lead to complications if we want to use reaction times as a variable in the correlations. The problem is as follows: Since we use only the reaction times from successfully completed trials, a participant who does only the easy trials (for example, the first five trials) would have a faster average reaction time than a participant completing almost all, or all of the trials, including the difficult ones (which would take much more time than the easy ones). In this scenario, the first participant would seem as overall faster, even though s/he completed only a fraction of the trials compared to the latter participant, who is more successful in the number of completed trials but overall seems much slower.

To overcome this problem, we used a penalisation strategy for those trials that were either completed unsuccessfully (wrong answer), or not completed at all (missing value). The penalisation reaction time value was extracted from the dataset represented by the longest reaction time in a successful trial among all of the participants. This value (169 823 ms; almost three minutes) was then applied to the incorrect or missing trials of all participants. This way, we were able to get a variable called *Mean RT after penalisation* for all the participants, which we then used in the correlations. Table VI illustrates this approach. Firstly, the table shows the mean reaction times without penalisation compared to the number of correctly answered trials. It is obvious that faster participants had fewer correct answers. The table also shows the *Mean RT after penalisation* compared to the number of correct trials. This variable reflects not only the reaction times, but it also takes into consideration the number of correct trials. Therefore it reflects the overall success of the participant. The Tab. VI represents an example of described approach and does not show all data sets.

Tab. VI: *Mean reaction times from the correct trials, and number of the correct trials together with the new variable Mean RT after penalisation for the Mental rotation subtest. The colour scale from red to green represent the RTs from the highest to the lowest and the no. of correct answers from less to more.*

Id	**Mean RT, correct trials only (ms)**	**No. of correct trials**	**Mean RT after penalisation (ms)**
N06	7839.3	24	40236.1
N07	17485.0	24	47952.6
N08	17301.9	28	27470.0
N09	24198.4	22	63031.6
N10	12459.5	20	64914.0
N11	14305.4	25	40225.0
N12	28248.2	24	56563.2
N13	11648.1	26	32738.1
N14	16084.5	28	26333.7
N15	10598.4	27	26520.9
N16	12966.7	26	33880.8
N17	31542.1	29	36151.4
N18	12664.0	22	54573.1
N19	12096.4	19	69929.5
N20	21759.5	30	21759.5
N21	17428.4	26	37747.7

4.5.2 Hidden figures subtest

The Hidden figures subtest does not have a clear order of trials from the easiest to the hardest, but there are still some trials that are more difficult than others. Therefore the same approach as described above was applied to the data gained from this subtest. We found the longest reaction time of all correctly solved trials, and used this value as penalisation for wrong trials as well as skipped trials (RT = 104 816 ms; 1.74 min.). This approach is represented in Tab. VII.

Tab. VII: *The table shows the mean RT calculated from the correct trials only, and Mean RT after penalisation compared to the number of correct trials. The colour scheme is the same as in previous Tab. VI.*

Id	Mean RT, correct trials only (ms)	No. of correct trials	Mean RT after penalisation (ms)
N006	7507.1	8	34045.9
N007	16740.1	8	40760.8
N008	18982.3	4	73603.7
N009	8852.8	5	61196.4
N010	18815.6	7	50088.5
N011	24950.1	7	53992.3
N012	12534.1	7	46091.2
N013	21507.6	8	44228.1
N014	9295.3	8	35346.4
N015	18842.4	5	65737.1
N016	6813.0	4	69178.5
N017	43370.6	7	65714.4
N018	41411.8	8	58703.8
N019	6311.0	4	68996.0
N020	22170.4	9	37196.9
N021	9976.2	6	53085.2

The second complication regarding the Hidden figures subtest was the fact that not all of the participants completed a minimum of two tasks (this is necessary to be able to calculate a mean RT). We decided not to include those participants in the analysis. Based on this decision, 14 participants were excluded from the correlations between the Hidden figure subtest and the Mental rotation subtest (final N = 99). Six of those 14 had only one correct answer in the Hidden figures subtest, and eight did not complete any trial correctly. Out of those 14, 12 were women (85 %), only one was from the natural science field (the rest from the humanities), and the average age was 21 years (*Sd* = *1.88*; max = 25; min = 19).

These participants' data were excluded from both the Hidden Figure subtest and the Mental rotation subtest before the main analyses were performed.

4.5.3 Target position subtest

This subtest was partially analysed separately, as it is a conceptual replication of Gilberts's study (Gilbert *et al.*, 2006). Therefore, we did not exclude the participants that were excluded from the previous two subtests for the analyses that did not include these subtests. For correlation between the Target position subtest and the other subtests (Hidden figure subtest and Mental rotation subtest), we excluded the same 14 participants as described above.

In the Target position subtest, we decided to exclude participants when more than 10% of the trials were incorrect (more than 15 wrong answers). The most common reason for excluding a participant was an incorrect use of the response keys. It was obvious which participants used wrong keys because typically half of the trials were answered correctly. This suggests that those participants used one key as instructed (either Q or P), but made the mistake of using a different key for the other half of the trials. Seven participants were excluded from the analysis of the Target position subtest (5 male participants; 71.4%), average age 22.1 years (*Sd = 1.45.* max = 25; min = 20; final N = 106).

The mean reaction times were calculated based on the correct trials. As there were only 0.88% wrong answers (135) and most of the participants had only one or two wrong answers, we decided not to use the *penalisation for wrong answers* in this case.

4.6 Results

4.6.1 Correlations between Mental rotation subtest and Hidden figures subtest

First, the data had to be examined regarding normality. The variable *Mean RT after penalisation* was used as a main variable for the correlations, the number of correct answers was also correlated.

It is important to carefully define which participants are excluded from the dataset. The participants that were excluded from the Hidden Figures subtest *(see 4.5.2)* were excluded from the whole dataset used in this analysis. The final number of participants in this analysis thus was 99 (60 female; 60.6 %) with an average age of 23.27 years *(Sd = 4.67)*.

Normality tests have shown that the variable *Number of correct answers* in both the Hidden figures subtest and the Mental rotation subtest are not normally distributed. The Kolmogorov–Smirnov test showed significant differences between the data and a normal distribution (for the Hidden figures subtest $D(99) = 132$; $p < .001$; for the Mental rotation subtest $D(99) = .102$; $p = .013$; $p < .05$). The Shapiro–Wilk test showed the same results (Tab. VIII). These results were supported by histograms (Hidden figures subtest: skewness = .186; kurtosis = –1.037; Mental rotation subtest: skewness = –.550; kurtosis = –.050).

Therefore, when we used the number of correct answers as a correlated variable, we used non–parametric methods (Spearman's rho as a correlation coefficient).

The variable *Mean RT after penalisation* of the Mental rotation subtest is the only variable that did not show significant results in the Kolmogorov–Smirnov test ($D(99) = .081$; $p = .112$; $p > .05$). The Shapiro–Wilk test revealed significant differences between the data and a normal distribution ($W = .967$; $p = .013$; $p < .05$), which suggested using non–parametric methods. The *Mean RT after penalisation* variable for the Hidden figures test showed significant results in the Kolmogorov–Smirnov test ($D(99) = .094$; $p = .031$; $p < .05$) together with the Shapiro–Wilk test ($W = .966$; $p = .013$; $p < .05$). The results were supported by histograms (Hidden figures subtest: skewness = –.293; kurtosis = –.878; Mental rotation subtest: skewness = .614; kurtosis = .066).

The results of normality tests suggested using non–parametric methods for both variables (*Mean RT after penalisation, Number of correct answers,* see Tab. VIII).

Tab. VIII: *Normality tests for the data from Hidden figures subtest and Mental rotation subtest.*

	Kolmogorov–Smirnov[a]			**Shapiro–Wilk**		
	Statistic	**df.**	**Sig.**	**Statistic**	**df.**	**Sig.**
Mean RT after penalisation_ Rotation subtest	.081	99	.112	.967	99	.013
Mean RT after penalisation_Hidden figures sub.	.094	99	.031	.966	99	.013
No. of correct answers_Mental Rotation sub.	.102	99	.013	.961	99	.005
No. of correct answers_Hidden figures sub.	.132	99	.000	.944	99	.000

a. Lilliefors Significance Correction

The correlation test between the **number of correct answers** of the Hidden figures subtest and the Mental rotation subtest showed significant results of moderate positive correlation ($r_s = .336$; $p < .001$). Using bootstrapping (simple sampling method, 1000 samples), the confidence interval was defined between .142 and .519.

The results showed a moderate correlation between **Mean RT after penalisation** of the Hidden figures subtest and the Mental rotation subtest ($r_s = .382$; $p < .001$), with a confidence interval between .206 and .544. These correlations support the hypothesis which assumes a relation between these two subtests.

Out of curiosity, we tried to correlate the mean reaction times calculated only from the correct trials (no penalisation for incorrect or missed trials). The Kolmogorov−Smirnov test (together with the Shapiro−Wilk test) suggested a non−normal distribution of both variables (Tab. IX), so we used Spearman's correlation coefficient rho and found non−significant results ($r_s = -.039$; $p > .05$) with a confidence interval between = −.241 and .156. These results suggested no relation between these two subtests. This, however, is very unlikely considering the correlation between numbers of correct answers of the subtests, which was significant and moderately strong. This additional analysis suggested that our method of creating the new variable *Mean RT after penalisation* was justifiable and reasonable and that the correlation using this variable reflects the real relation between the data of the two subtests. The fact that the correlation coefficient calculated from the number of correct answers was close to the correlation coefficient calculated from the *Mean RT after penalisation* suggests that the *Mean RT after penalisation* is a variable reflecting RTs together with the number of correct answers and represents the overall success of a participant in the task.

Tab. IX: *Normality tests for the data from Hidden figures subtest and Mental rotation subtest calculated from the correct trials only.*

	Kolmogorov−Smirnova			Shapiro−Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
Mean RT_correct answers_ Mental rotation sub.	.118	99	.002	.949	99	.001
Mean RT_correct answers_ Hidden Fig. sub.	.110	99	.005	.910	99	.000

a. Lilliefors Significance Correction

Finally, we tried to do the correlations with the sample of N = 92. after we got rid of not only the participants that had problematic data in the Hidden figures subtest and the Mental rotation subtest (N = 14), but also those that had problematic data in the Target position subtest (N = 7). The correlation between the number of correct answers in the Mental rotation subtest and the Hidden figure subtest was significant ($r_s = .327$. $p < .01$), with confidence intervals between .129 and .496 (bootstrap sample N = 1000). The strength of the correlation did not change noticeably compared to the sample of N = 99 ($r_s = .336$. $p < .001$).

The correlation between the *Mean RT after penalisation* in the Mental rotations and the Hidden figures tests (N = 92; r_s = .376; p < .001, confidence interval between .182 and .533. bootstrap sample N = 1000) is also very similar to the one calculated based on the sample of N = 99 (r_s = .382; p < .001; with a confidence interval between .206 and .544).

4.6.2 Gender differences

The results did not show any significant gender differences in any of the variables (except one, see Tab. X). Levene's test shows non–significant results for all tested variables (except *Mean reaction time after penalisation* for Mental rotation subtest), which suggests equality of variances in the samples. We used t–test to compare genders regarding the number of correct answers and mean reaction times (both after penalisation and before). The number of correct answers in the Mental rotation subtest did not differ between genders (t = .146; df = 97; p > .05), which also applies to the number of correct answers in the Hidden figures subtest (t = –.770. df = 97; p > .05). The *Mean RT after penalisation* showed the same results in Mental rotations (t = .547. df = 97; p > .05) and in Hidden figures (t = .288. df = 97; p > .05). The only significant result in the gender differences analysis has been found in the variable *mean reaction time* calculated from correct answers (no penalisation) for the Hidden figures subtest (t = –.211. df = 97. p = .037), suggesting that female participants are overall faster than males. The tendency for male participants being faster than female participants (but non–significant) was found in the *mean reaction time* of correct answers for the Mental rotation subtest (t = 1.96. df = 97. p = .053). The groups divided by gender were not of the same size (m = 39; f = 60), which could also have played a role in the results. Also, the mean reaction time calculated only from correct answers is not the variable we used for our main analyses.

Out of curiosity, we did an analysis with respect to gender differences in the Target position subtest, where we found no significant differences between male and female participants (t–test for independent samples; t = . 231. df = 104. p = .818). Again, this subtest does not have the same number of male and female participants (m = 70. f = 36). Thus, we can conclude, that overall there are no gender differences in our sample.

Tab. X: *Descriptive statistics for gander differences in different subtests.*

	Group Statistics				
	sex	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Mean RT after penalization_Hidden figures sub.	1	60	61848.6	16510.8	2131.5
	2	39	60829.7	18269.7	2925.5
Mean RT after penalization_ Mental rotation sub.	1	60	43547.1	14226.5	1836.6
	2	39	41935.8	14489.7	2320.2
No. of correct answers_ Mental rotation sub.	1	60	24.9	3.1	.404
	2	39	24.8	3.3	.533
No. of correct answers_Hidden Fig. sub.	1	60	5.3	2.1	.268
	2	39	5.6	2.3	.363
Mean RT_correct answers_ Mental rotation sub.	1	60	17375.6	6595.8	851.5
	2	39	14699.2	6703.7	1073.5
Mean RT_correct answers_ Hidden Fig. sub.	1	60	14859.7	7866.7	1015.6
	2	39	18759.9	10480.8	1678.3

4.6.3 The reasoning behind creating a new variable (Mean RT after penalisation)

On first sight, it may seem to be too complicated to create a new variable *Mean RT after penalization* (see *chapter 4.5*). However, it turned out to be necessary to create this variable after the performance of some analyses that showed complications with respect to the method used in calculating the mean reaction time from correct answers only.

Our first analysis consisted of the correlation between correct answers in two subtests (Mental rotation subtest and Hidden figure subtest; the analyses were done with N = 99). As described before, the result was significant because therefore suggests a moderate relationship between those two subtests ($r_s = .336$; $p < .001$). However, a problem emerged after calculating the correlation between the mean RT based solely on correct answers. The results did not show any significance, thus suggesting no relation between mean RT of correct answers in the Mental rotation subtest and the Hidden figures subtest. This was implausible and surprising, given the fact that the number of correct answers of the two subtests correlated on a moderate level. We therefore analysed the correlations between the number of correct answers and mean reaction time calculated based only on correct answers separately for both subtests. The correlation between the number of correct answers and mean RT of correct answers in the Mental rotation subtest was significant with moderate to strong relation ($r_s = .597$. $p < .001$. confidence intervals between .434 and .737). The Spearmen's correlation is positive, which means the more correct answers, the longer the mean reaction time. This relation could be explained by the fact that the trials in the Mental rotation subtest are organized from the easiest to the most difficult trial. This, on the other hand, meant, the more trials were completed, the more time it took. If a participant completed only four out of 30 trials s/he had a much faster overall mean RT than those who completed for example 20 trials. The problem occurred when we did the same correlation for the Hidden figures subtest. The correlation between the number of correct answers and the mean RT calculated only from correct answers was significant ($r_s = .249$. $p = .013$), but with lower confidence interval almost crossing 0 (lower .060. upper .421. bootstrap sample = 1000). Pearson's correlation was not significant at all ($r = .177$. $p > .05$). This means that the mean RT does not depend much on the number of correct answers in the Hidden figures subtest (contrary to the Mental rotation subtest, where the mean RT depends on the number of correct answers). This suggest a weak relation between the number of correct answers and the overall mean RT (calculated from correct answers only) in the Hidden figures subtest, which could cause the non–existing relation between the mean RT from correct answers in Hidden figure subtest and Mental rotation subtest ($rs = -.027$. $p > .05$. confidence intervals between = –.221 and .183).

After creating the variable *Mean RT after penalisation* for both the Hidden figures subtest and the Mental rotation subtest, we found out that this new variable correlated strongly with the number of correct answers in each subtest (partly because it was created based on the number of correct answers and it should reflect the number of correctly solved trials). This variable reflects overall success in the subtest. The correlation between the number of correct answers and the new variable *Mean RT after penalisation* in the Mental rotation subtest is $r_s = -.932$. $p < .001$. with confidence intervals between –.959 and –.887. which suggest the more correct answers a participant has, the lower is the *Mean RT after penalisation*.

The same applies to the Hidden figures subtest, where the correlation between the number of correct answers and the Mean RT after penalisation is $r_s = -.966$. $p < .001$. with confidence intervals between –.977 and –.943 (bootstrap sample N = 1000). The more correct answers a participant has, the lower is the *Mean RT after penalisation* . All these findings suggest that the new variable can be seen as a reflection of the overall success in the task that reflects not only reaction times but also the number of correctly answered trials.

4.6.4 Correlations between the Target position subtest and the rest of the test battery

The correlations between all three subtests were calculated after all the problematic participants from all the subtests were excluded, therefore final N = 92. The correlations between the Mental rotation subtest and the Hidden figure subtest were thoroughly described in chapter *4.6.1* The main focus of this section is the Target position subtest. To correlate the results of the Target position subtest with the rest of the subtests we used the mean reaction time of only the correct answers from the whole Target position subtest for each participant. We did not use the method of penalizing participants in case of wrong answers, because after excluding participants that had failed in more than 10% of the answers, the rest of the participants had only a few wrong answers (max. 12 wrong answers in only one participant; the rest of the participants showed a maximum of three wrong answers).

We hypothesized that the Target position subtest was connected to the rest of the battery, however, this was not confirmed. The Target position subtest does not correlate with any of the variables we worked with (number of correct answers in both Mental rotation subtest and Hidden figure subtest; *Mean RT after penalisation* in both subtests). We used non−parametric Spearman's rho as correlation coefficient (both the Kolmogorov−Smirnov test and the Shapiro−Wilk test were significant for the mean RT in the Target position subtest, $p < .001$; skewness = 5.213. kurtosis = 34.436). The results are presented in Tab. XII.

Tab. XI: *The correlation between Target position subtest (Mean RT_Target position sub.) and the rest of the battery.*

Spearman's rho				Mean RT_Target position sub.	Mean RT after penalization_Hidden figures sub.	Mean RT after penalization_ Mental rotation sub.	No. of correct answers Mental rot. Sub.	No. of correct answers_ Hidden Fig. sub.
Mean RT_Target position sub.	**Correlation Coefficient**			1	0.146	−0.085	0.172	−0.136
	Sig. (2−tailed)			.	0.163	0.421	0.102	0.197
	N			92	92	92	92	92
	Bootstrap[c]	**Bias**		0	−0.003	−0.003	0.001	0.003
		Std. Error		0	0.107	0.104	0.103	0.11
		95% Confidence Interval	**Lower**	1	−0.079	−0.28	−0.041	−0.349
			Upper	1	0.355	0.14	0.367	0.103

After trying to use Pearson's correlation coefficient, we found out that there is a borderline significant correlation between the mean reaction time in the Target position subtest and the Hidden figures subtest (both the number of correct answers $r = -.205$. $p < .05$ and the *Mean RT after penalization* $r = .204$. $p < .05$). This correlation is weak and was calculated using a parametric test, which proofed not suitable in this case. We could conclude that there probably is a relation between the Target position subtest and the Hidden figures subtest, but it could be explained by any underlying condition, such as general intelligence, or overall psychomotor pace.

Surprisingly, a significant correlation was found between the mean reaction time of the Target position subtest and the mean reaction time calculated only from the correct answers of the Mental rotation task ($r_s = .314$. $p < .05$). This analysis was conducted out of curiosity and is really difficult to interpret. Again, it could be caused by underlying factors.

The overall result here suggests no relation (or a very limited one) of the Target position subtest and the rest of the battery. The hypothesis that the results of this subtest would also correlate with the tests of spatial cognitive ability (Mental rotation subtest and Hidden figure subtest, whose correlation did show significant results, see *chapter 4.6.1*) was not confirmed.

4.7 Replication of Gilbert *et al.* (2006)

The Target position subtest is based on a study by Gilbert *et al.* (2006); the study's method is described earlier (see *chapter 3.3*). The analyses carried out were supposed to find out whether there is a difference between the mean reaction time in the trials in which the participants distinguished between the colours of the same linguistic colour category and those of a different colour category. There were four colours (Tab. V, Fig. 22), therefore 12 colour pairs of target and distractor.

Each of the 12 conditions is represented by a double−digit code. The first number of the code represents the target in a particular condition and the second number represents the distractor. For example, the condition with the code "12" would be a one with a colour number 1 as a target and colour number 2 as a distractor, and so forth (21; 23; 32; 34; 43; 13; 31; 14; 41; 24; 42).

The conditions that are relevant for our cause are the conditions 12 and 21, 23 and 32, 34 and 43. What interests us the most is the differences between the conditions that contain

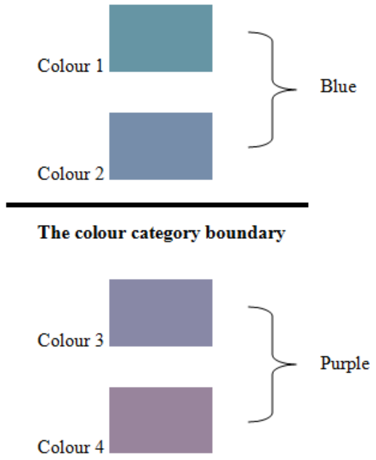


Fig. 22: *The representation of the colour categories (blue and purple) and the linguistic colour boundary.*

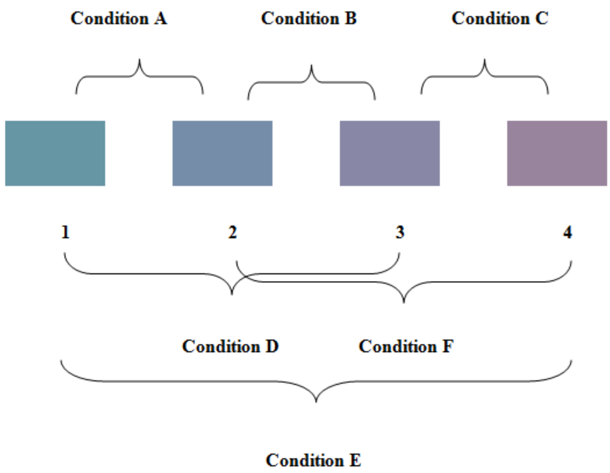


Fig. 23: *The conditions of Target position subtest.*

the colours from the same colour category (12 and 21, 34 and 43) and the conditions that contain colours crossing the colour category boundary (23 and 32). Based on Gilbert's (*et al.*, 2006) example, we combined condition 12 with condition 21; condition 23 with condition 32; and condition 34 with condition 43. We call conditions 12 and 21 Condition A, conditions 23 and 32 Condition B, and conditions 34 and 43 Condition C (for representation of conditions see Fig. 23).

This approach makes sense, as the colours within one condition (A, B, C) are the same, the difference lies only in the target/distractor division. We decided to test this assumption. We wanted to find out if there is any difference between two versions of the same condition (for example, if there is a difference between the trials where colour 1 is a target and colour 2 a distractor, and those with colour 2 as a target and colour 1 as a distractor). These analyses were all non–parametric because each of the variables (mean reaction time per condition) was not distributed normally (Kolmogorov–Smirnov test and Shapiro–Wilk test, see Attachment 3). The analysis we used to compare the two different versions of a single condition was a Wilcoxon signed–rank test (test for two measurements of a single sample).

The results are confusing. The condition B and the condition D show no difference between the two versions (for B the versions are 23 and 32; for D they are 13 and 31). Condition B (this is the condition where the linguistic colour category boundary is placed), $Z = -1.042$. $p = .298$. the condition D ($Z = -1.951$. $p = .051$).

The condition E (colours 1 and 4) showed significant level of difference on the $p < .05$ ($Z = -2.216$. $p = .027$).

The rest of the conditions have showed significant differences between versions of the conditions. The condition A (colours 1 and 2) $Z = -7.576$. $p < .001$, condition C (colours 3 and 4) $Z = -8.937$. $p < .001$, condition F (colours 2 and 4) $Z = -5.795$. $p < .001$.

The overall conclusion is that conditions B and D could be taken together. The mean reaction time could then be calculated for the whole condition altogether (not separately for two versions with target–distractor variations). The rest of the conditions showed a significant difference between the two versions. As we were most interested in the differences between condition A – B and B – C (because the linguistic boundary had been placed between colours 2 and 3 and this is the condition B where there is no significant difference between its two versions), we performed the analyses calculating the mean reaction times for condition B from the two versions of the condition together. After all, we assume this was the method used in the original study by Gilbert *et al.* (2006), even though it was not explicitly described in the respective article. To be sure of our results we performed the analyses also on the separate versions of all the conditions (we would not calculate the mean RT per condition but per version of a condition).

The first analysis performed was the factorial repeated–measures ANOVA (3 × 2), with a factor colour condition (A, B, C) and a target position (right/left). First, we tried to perform the analysis for the three colour conditions, therefore we put the two versions of the condition together.

Mauchly's test suggested a violation of the assumption of sphericity for colour, $X^2(2) = 42.108$, $p < .001$. and for the colour*side interaction $X2$ $(2) = 10.967$, $p < .05$. So for reporting the main effect of colour and the interaction effect of colour*side, we used degrees of freedom corrected by Greenhouse–Geisser estimates of sphericity.

The main effect of colour condition was significant at $p < .05$, $F(1.5, 157.5) = 5.239$. The main effect of side (left/right target position) was also significant $p < .05$, $F(1,105) = 4.139$. This suggests that there is a significant difference in reaction times between the left and right position of the target and also that there is a difference between colour conditions. The contrasts for the colour conditions suggests that the reaction time in condition A is significantly higher than in colour condition B $F(1, 105) = 24.325$; $p < .00$; the RT in conditions B and C do not differ significantly (for descriptive statistics see Tab. XIII and Tab. XIV).

The contrasts of a significant main effect of position (left/right) $F(1,105) = 4.139$, $p < .044$ suggests a difference between left and right position of the target with significantly faster reactions if the target was on the left side (contrary to the hypothesis and Gilbert's study).

The interaction effect between colour and side conditions was not significant, $F(1,190.8) = 2.28$, $p > .05$. suggesting that independent of the colour condition, the reaction time is faster if the target is on the left side. But when we look at the contrasts for the interaction effect (Tab. XV), we can see that the contrasts are significant for the colour condition A and B, even though the results are contrary to the hypothesis. In colour condition B, the participants had significantly faster reactions if the target was on the left side of their visual field, compared to the right side (Tab. XII, Fig. 24).

Tab. XII: *The mean RT for different conditions of Target position subtest. The mean RT's are listed separately for right and left target position trials for each condition. The condition B is listed first, because the linguistic colour boundary is placed between colours of this condition (2 and 3).*

	Mean RT	Std. Deviation	N
Con.B_Right	1490.7	1269.7	106
Con. B_Left	1353.3	1002.1	106
Con.A_Right	1614.4	1216.9	106
Con.A_Left	1593.0	1106.0	106
Con.C_Right	1550.7	831.4	106
Con.C_Left	1506.7	669.9	106

Tab. XIII: *The mean RT for condition A, B and C calculated for right and left target position trails together. Again, the condition B is listed first, because the linguistic colour boundary is placed between colours of this condition (2 and 3).*

Condition	Mean RT	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Con. B	1422.0	109.1	1205.6	1638.4
Con. A	1603.8	109.8	1386.0	1821.5
Con. C	1528.7	69.6	1390.6	1666.8

Tab. XIV: *The mean RT for all the trials (regardless the condition) separated by the right/left target position.*

Side	Mean RT	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
Right	1551.97	102.8	1348.1	1755.9
Left	1484.4	84.1	1317.7	1651.1

Tab. XV: *Tests of within – subject contrasts.*

Source	colour	side	Type III Sum of Squares	df	Mean Square	F	Sig.
colour	**con. A vs. con. B**		3500904	1	3500904	24.325	.000
	con. C vs. con. B		1205819	1	1205819	2.664	.106
Error(colour)	**con. A vs. con. B**		15111891	105	143922		
	con. C vs. con. B		47533252	105	452697		
side		**left vs. right**	484633	1	484633	4.139	.044
Error(side)		**left vs. right**	12293138	105	117077		
colour * side	**con. A vs. con. B**	**left vs. right**	1427380	1	1427380	4.334	.040
	con. C vs. con. B	**left vs. right**	925119	1	925119	3.464	.066
Error(colour*side)	**con. A vs. con. B**	**left vs. right**	34578789	105	329321		
	con. C vs. con. B	**left vs. right**	28042874	105	267074		

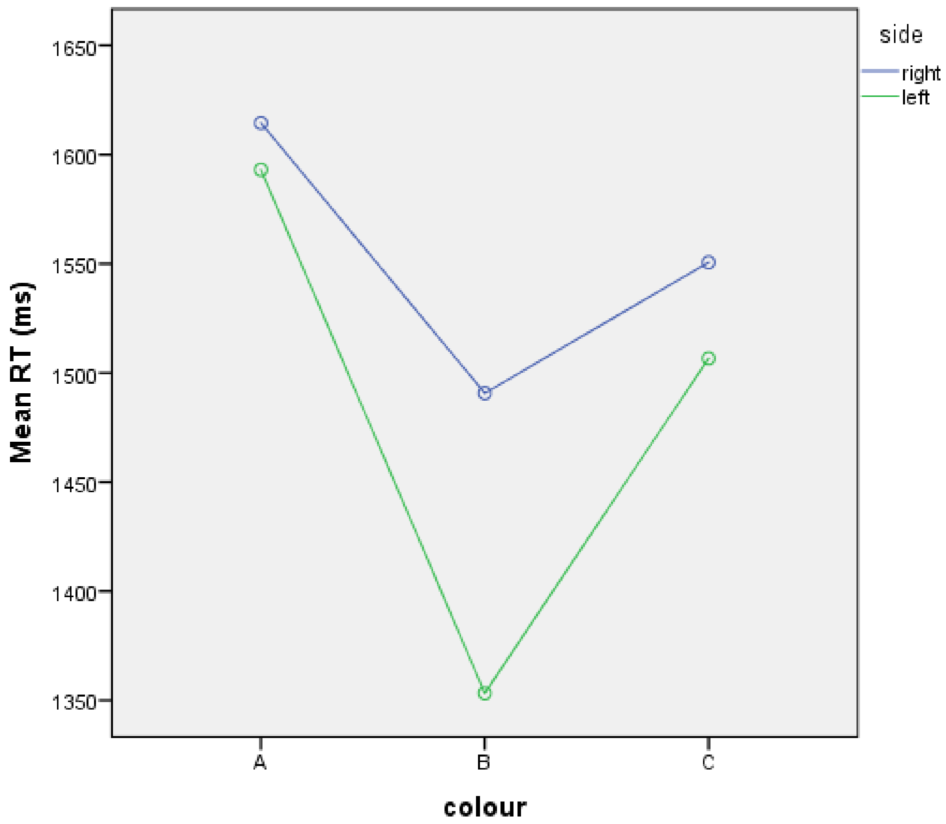


Fig. 24: *The mean RT difference between three colour conditions and positions of the target (right/left).*

4.7.1 The two versions of the conditions separated

The colour conditions that are the focus of interest here are the conditions A, B, and C. Condition B (the condition where the linguistic boundary was placed between blue and purple) is the condition, where there was no difference between mean reaction times in its two versions. The next analysis works with this condition as a whole, while the rest of the conditions are separated for each version of the target–distractor pairs.

Mauchly's test of sphericity is significant for the main effect of colour (X^2 (9) = 116.95, $p < .001$), and for the interaction effect (*X2* (9) = 101.27, $p < .001$), thus we used the Greenhouse–Geisser correction of degrees of freedom. In this case, only the main effect of colour is significant; F (2.5, 263.8) = 49.7, $p < .001$.

The main effect of side (target position) is not significant F (1, 105) = 2.43, p = .122), which also applies to the interaction effect of colour*side , F (4, 300.4) = .647, p = .578; see Fig. 25.

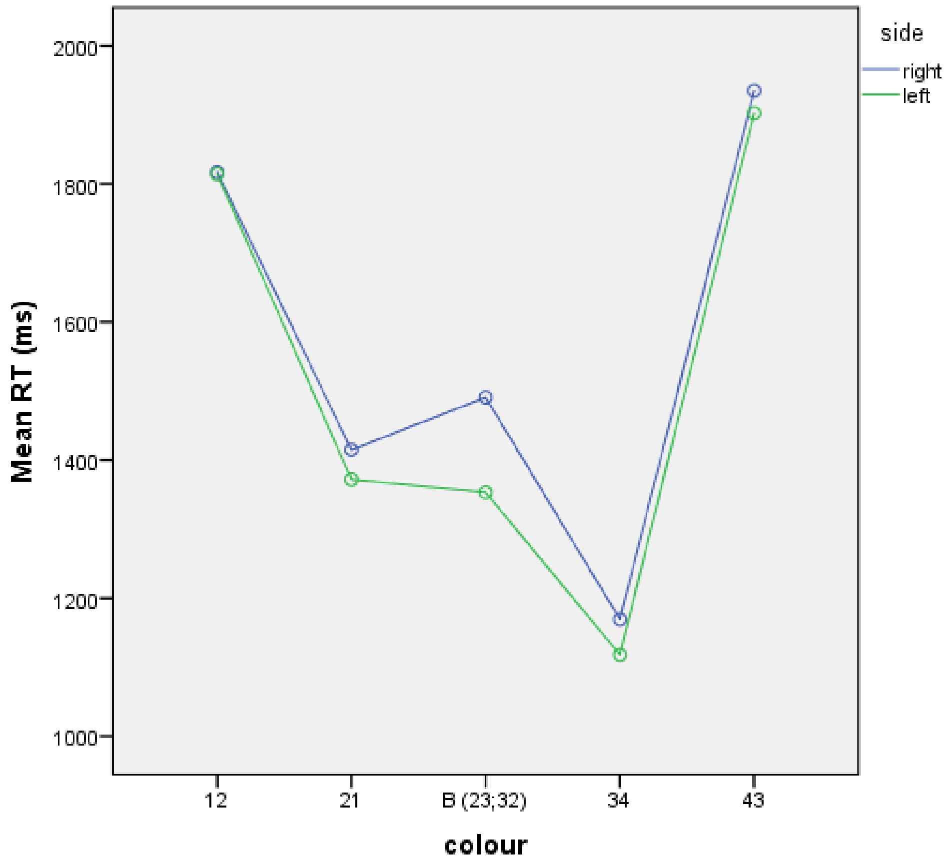


Fig. 25: *The mean RT difference between five colour conditions and positions of the target (right/left).*

We used the colour condition B as a reference condition for the test of within–subjects contrasts because it is the condition that contains colours that cross linguistic colour boundary. The contrasts suggest that there is a difference between condition B and the rest of the conditions (on a level of p < .001) with the exception of condition 21 (colour 2 as a target and colour 1 as a distractor), $F(1, 105) = .350$, $p = .556$ (Tab. XVI. Condition 21 , therefore, is not different from condition B (the one with the linguistic colour boundary) in the reaction time needed to spot the target position, even though participants were faster in condition 21 compared to condition B. Even more interesting is that whilst in the rest of the conditions (12 and 43), the participants were significantly slower in spotting the target position, in condition 34 they were significantly faster (see Tab. XVII and Tab. XVIII for descriptive statistics). This finding, together with the fact that in condition B the target was spotted faster if it was on the LEFT side of a visual field (Wilcoxson signed–rank test, $Z = -4.094$. $p < .001$), which opposes both our hypothesis and the findings of Gilbert *et al.* (2006), is the most surprising finding; which raises many questions and calls for interpretation.

Tab. XVI: *The tests of within – subjects contrasts between the con. B and the rest of the conditions.*

Source	colour	side	Type III Sum of Squares	df	Mean Square	F	Sig.
colour	**con.12 vs con B**		16434349	1	16434349	76.10	.000
	con 21 vs. con B		86428	1	86428	.35	.556
	con 34 vs. con B		8203752	1	8203752	16.19	.000
	con 43 vs. con B		26144978	1	26144978	48.97	.000
Error (colour)	**con 12 vs. con B**		22675304	105	215955		
	con 21 vs. con B		25951311	105	247155		
	con 34 vs. con B		53194091	105	506610		
	con 43 vs. con B		56060846	105	533912		
side		**left vs. right**	302222	1	302222	2.43	.122
Error(side)		**left vs. right**	13079563	105	124567		
colour * side	**con.12 vs con B**	**left vs. right**	19209767	1	1920976	3.16	.078
	con 21 vs. con B	**left vs. right**	949535	1	949535	1.18	.280
	con 34 vs. con B	**left vs. right**	784879	1	784879	3.14	.079
	con 43 vs. con B	**left vs. right**	1166278	1	1166278	1.95	.166
Error(colour*side)	**con.12 vs con B**	**left vs. right**	63762842	105	607265		
	con 21 vs. con B	**left vs. right**	84486012	105	804628		
	con 34 vs. con B	**left vs. right**	26239741	105	249902		
	con 43 vs. con B	**left vs. right**	62855721	105	598625		

Tab. XVII: *Mean RT for the condition B and the rest of the relevant conditions, listed for each of the version of the condition separately and also separately for right and left target position.*

	Mean RT	Std. Deviation	N
Con. B_Right	1490.7	1269.7	106
Con. B_Left	1353.3	1002.1	106
Con.12_Right	1817.2	1349.9	106
Con.12_Left	1814.4	1312.7	106
Con.21_Right	1414.9	1265	106
Con.21_Left	1372.1	982.3	106
Con.34_Right	1169.5	691.9	106
Con.34_Left	1118.1	542.1	106
Con.43_Right	1934.9	1063.4	106
Con.43_Left	1902.4	902.7	106

Tab. XVIII: *Mean RT for the condition B and the rest of the relevant conditions listed for each of the version of the condition separately but together for right/left target position.*

colour	Mean RT (ms)	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
con B	1422.0	109.1	1205.6	1638.4
con 12	1815.8	124.6	1568.7	2062.8
con 21	1393.5	100.6	1194.1	1592.9
con 34	1143.8	57.2	1030.4	1257.3
con 43	1918.7	88.1	1744	2093.3

4.7.2 Irrelevant conditions

To make sure the effects we found were not a product of a mistake or unforeseen circumstances, we conducted an analysis that compared the crucial condition (condition B, where the linguistic colour boundary is placed) with "irrelevant" conditions. The reason why they are judged irrelevant is because these conditions suggest clear and very probable hypotheses. The hypothesis in question is the following: **Those conditions that contain colours that have more than one colour hue in between them have significantly lower mean reaction times than condition B.** The conditions are 13; 31; 14; 41; 24; 42. The first number refers to the target colour, the second number is the distractor colour. All of the conditions listed were compared to condition B.

As hypothesized, there is a significant difference between all of the above conditions and condition B. The main effect of colour and side were both significant (the effect of colour after the Greenhouse−Geisser correction of degrees of freedom, the Mauchly's Test of sphericity was significant for colour, X^2 (20) = 1091.8, p < .001). The main effect of colour was significant with F (1.17, 123.1) = 51.7, p < .001). Condition B showed significantly slower reaction times in finding a target (Tab. XIX; Fig. 26). This condition seems more difficult because there is a smaller difference between the target and the distractor compared to the conditions where the target and the distractor are further away from each other (separated by another colour hue). This result suggests that the methodology of the study and the analyses are well−chosen.

Interestingly enough, the main effect of size showed that it took participants significantly longer to locate the target if it was located in the right side of the visual field (main effect of side *F* (1,105) = 9.8, p < .05; Tab. XX).

Tab. XIX: *Pairwise Comparisons between con. B and the (so-called) irrelevant conditions.*

(I) colour	(J) colour	Mean Difference (I–J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Con. B	con. 13	537.5*	79.1	.000	291.1	783.8
	con. 31	557.5*	79.0	.000	311.4	803.5
	con. 14	695.3*	94.2	.000	402.1	988.6
	con. 41	663.2*	81.3	.000	410.1	916.3
	con. 24	682.5*	91.7	.000	396.9	968.2
	con. 42	605.1*	84.7	.000	341.3	868.8

Tab. XX: *The difference between mean RT for right and left target position in irrelevant conditions.*

side	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
right	902.4	42.7	817.8	987.0
left	872.7	39.1	795.3	950.2

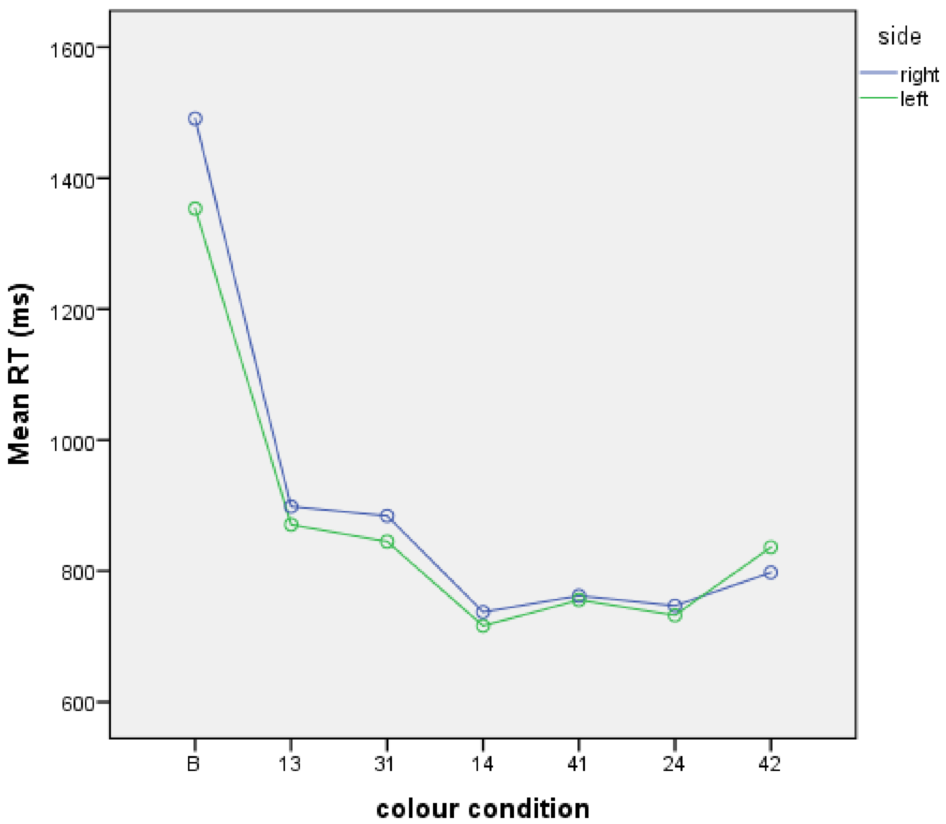


Fig. 26: *The difference between condition B and the rest of the conditions (so-called irrelevant).*

The only condition where reaction times were faster for the right−hand side was condition 42, but the result is non−significant according to Wilcoxon signed−rank test ($Z = -.966$, $p = .334$; Tab. XXI).

Tab. XXI: *The difference between mean RT for left and right target position in condition 42.*

		Mean	SD
42	**right**	797.7	34.5
	left	836.3	37.3

5 DISCUSSION

The main interest in this work is spatial cognitive ability, more specifically its testing through methods that are easily implemented in a classroom either using paper–pencil administration or a simple administration on a computer. The main purpose here was to determine whether the Hidden figures subtest based on Witkin's Embedded Figures test would be a proper method of testing spatial cognitive ability. The study tried to find a correlation between a standardized test of spatial ability (in this case a Mental rotation test) and a Hidden figures subtest, whose construction was strictly based on Witkin's original Embedded Figures test (which could not be used due to copyright reasons).

A literature review done as a base for the theoretical part of this thesis has shown that tests requiring participants to find a simple figure in a complex design have been used as tests of perceptual speed, a part of spatial cognitive ability (Ekstrom, Dermen & Harman, 1976). This finding was quite surprising because Ekstrom (together with collegues Dermen and Harman, 1976) himself addresses Witkin's EFT, and sees a connection between perceptual speed and field independence. However, he still advises the reader not to use EFT as a test of perceptual speed but rather use it to test cognitive styles (according to Witkin's instructions). This fact itself suggests that Ekstrom, Dermen and Harman (1976) see EFT as very similar to his Hidden Patterns test, which has the same objective as EFT. So the question of the extent of the differences between Ekstrom's and Witkin's test of hidden figures remains unanswered in the literature. Our study and its results suggest that tests based on the objective to find simple figures in a complex pattern are all testing the same concept, the same *cognitive ability*. Contrary to Ekstrom's careful suggestion not to use EFT as a test for perceptual speed (spatial cognitive ability), our results suggest the opposite.

The relevance of EFT for modern psychology can be seen in a recent attempt to create a new version of EFT by the Leuven research group concentrating around Johan Wagemans (de–Wit *et al.*, 2017; Huygelier *et al.*, 2018; Chamberlain *et al.*, 2017), which works on the Leuven–EFT (L–EFT) and asking the questions that have been asked here: what does the test based on EFT really measure?

The criticism regarding Field dependence/independence as a cognitive style that was heard from many sources is supported by our results. Previously mentioned research together with more recent articles bring about certain doubts with regard to EFT and the whole concept of field dependence/independence as a cognitive style (e.g. Angeli, Valanides, Polemitou & Fraggoulidou, 2016; Miyake, Witzki & Emerson, 2001; Tinajero & Páramo, 1997; Zhang, 2004).

Bigelow (1971) compared children (aged ten) from different socioeconomic classes using a children's version of GEFT and found that children from higher socioeconomic classes performed more field independently compared to their counterparts from lower socioeconomic classes. According to Bigelow (1971), field independence is a function of growing up in a family oriented more towards the importance of achievement, which is more often the case in higher socioeconomic classes. The author's results also showed that the older the child was, the more field independently s/he performed. This suggests that field independence is an adaptive and advantageous strategy that develops with increasing age.

The notion that field independence is in fact an ability rather than a style support a study done by Emmet, Clifford & Gwyer (2003). They refer to field independence as a "fluid ability" in line with Cronbach & Snow (1981) and Sternberg & Grigorenko (1997). McKenna (1984) provides a very interesting review of the literature on field dependence/independence and EFT, and discusses three possible links of this concept to cognitive ability. He provides a meta–analysis of literature finding correlations between field independence and general ability (general intelligence), where he shows that success in EFT is linked to a nonverbal test of general intelligence. Secondly, he provides broad evidence of a correlation between performance in EFT and spatial cognitive ability, where in some cases the correlation is strong (e.g., GEFT and Spatial Relations test correlated on .79 in Sherman's research 1974 as cited in McKenna, 1984), but generally is around $r = .50$. which in psychology is not weak at all. Lastly, he discusses a correlation between EFT and fluid intelligence in this case represented by the Raven matrices test, which is in general weaker, but still significant (for more detailed information and references see McKenna, 1984). This thorough analysis of experimental literature suggests a link between field dependence/independence and cognitive abilities, which is in line with our results. Even the correlations (in our case around $r = 0.30–0.40$) between EFT–like subtests and Mental rotation subtest suggest that the notion of field independence is a cognitive ability, strictly speaking, part of the spatial cognitive ability.

The hypotheses presented here are confirmed. Not only is there a correlation between a number of correct answers between the EFT–like subtest (Hidden figures subtest) and the Mental rotation subtest, also a significant correlation has been found between the overall success of these subtests.

Another support regarding the criticism and confusion that surrounds FDI cognitive style is the fact that tests created to measure FDI are sometimes used as tests of spatial ability. Quaiser–Pohl, Lehmann & Eid (2004) used Rod–and–Frame tests to test spatial ability in children (together with Mental – Rotations – Tests and Water – Level – Tasks) claiming that the Rod – and – Frame test and EFT are in fact tests of "*spatial – perception ability*" (Quaiser – Pohl, Lehmann & Eid, 2004, p. 96).

To be fair, the cognitive styles, or thinking styles (Zhang, 2002), are not only criticized, some researchers claim they bring a certain new quality about, and a new point of view to personality research, and thus help us understand individual differences (Nigro, Cicogna, D'Olimpio, & Cosenza, 2012; Peterson, Deary & Austin, 2005; Zhang, 2002). It is important to say that these studies defending styles in most cases use other styles than FD/FI for their research and argumentation. This work is not trying to argument against all cognitive styles, but rather specifically tries to show that FD/FI is not a style in that sense and that EFT and tests comparable to it can be used as tests for spatial cognitive ability.

The main relation between the Hidden figures subtest and the Mental rotation subtest were described and debated, however, the results dealing with the Target position subtest have not yet been discussed, and they are neither less important nor less interesting than the rest of the results.

The original study conducted by Gilbert and colleagues (2006) shows significantly faster reaction times in participants to spot the target if the target and the distractor were from different colour categories (in their case blue/green categories). This finding was significant only for the right visual field (Whorf effect). The authors explain this by claiming that language brain regions are in most individuals located in the left hemisphere, therefore language categorization affects visual perception. This study was not the only one that has found this kind of results (e.g.: Gilbert *et al.*, 2006; Gilbert *et al.*, 2008; Regier & Kay, 2009; Xia, Xu & Mo, 2018).

These study results, however, were not supported by our findings. On the one hand, we did not find significantly faster reaction times in trials where the target/distractor pair crossed the colour category boundary, and on the other hand, we found that participants were significantly faster in the trials where the target was located in the left visual field, not the right one. This applied for both kinds of trials where target/distractor crossed the colour category boundary, and for those where the target/distractor pair was from the same colour category. The results we obtained are based on a much larger pool of participants compared to that of

Gilbert *et al.* (2006) with eleven participants, Gilbert *et al.* (2008) with twelve participants, and Xia, Xu & Mo (2018) with thirty participants. The advantage of the research study with fewer participants, in this case, lies mostly in the possibility to design the colour category boundaries personally for each participant in a pilot study (Gilbert *et al.*, 2006), which we did not have the chance to do. We also conducted a pilot study, where recruited participants chose the colour categories. But then we used colour categories with the highest agreement among participants from the pilot study for all the participants in the main experiment. There is a chance that some individuals did not consider the colours we used as the same or different category (accordingly), which would affect their results. We argue that after choosing only the colours with more than 94% agreement about their category among participants in the pilot study, we did not anticipate that more than 6% of the participants in the main experiment would have different notions about the categories. The advantage of our process of choosing the colours is that we were able to have many more participants in the study, which increases the statistical strength of the results and gives us a better opportunity to generalize results compared to studies with fewer participants. The question is why we obtained results that are so different from the study we conceptually replicated.

The explanation of the results suggesting an advantage of stimuli located in the left visual field may lie in the studies examining attention deficits[19] in patients with right−hemisphere damage suffering from spatial neglect (e g.: Chica *et al.*, 2012). Studying neglect patients in comparison to healthy individuals suggests that attention deficit may be one of the crucial problems of the patients' state.

Attention is defined as a cognitive function that *"[...]prepare[s] the individual to respond quickly and accurately to incoming information by selecting relevant and ignoring irrelevant stimuli*" (Bartolomeo, 2006; Bartolomeo & Chokron, 2002; Chica *et al.*, 2012; Mesulam, 1981), which is very close to what we tested in our Target position subtest. Shulman *et al.* (2010) in their study located centres of spatial selective attention in the right hemisphere in agreement with Bartolomeo & Chokron (2002), Bartolomeo (2006) and Chica *et al.* (2011). This means that in larger groups of participants, the Whorf effect might be overpowered by the effect of spatial attentional networks[16] being located (mainly) in the right hemisphere, which may cause a faster location of the target in the left visual field regardless of the colour categories of the target and the distractor.

16 Attention is a complex cognitive function. Here we address the so-called sustained attention, which refers to the ability to willingly control our focused attention over a period of time (Chica *et al.*, 2006; Sturm & Willmes, 2001), together with spatial attention (Shulman *et al.*, 2010).

5.1 Gender differences in performance in spatial tasks

The gender differences (or sex differences; this work uses these two terms interchangeably as the term "sex"[17] and the term "gender"[18] are very often used interchangeably in the literature) were not found which is in line with the hypotheses. It is important to address this issue as it has not been clearly stated in the literature whether gender differences in EFT or in overall spatial tasks are significant. The literature review shows confusing results. Bigelow's study (1971) supports our finding that there is no gender difference in performance in EFT. On the other hand, McGilligan & Barclat (1974) test nine−graders using an intelligence test and Witkin's Rod−and−Frame test and they did find a significant difference between boys and girls in performance in RFT with a clear advantage on the boys' side. Also, the authors found a significant correlation between spatial ability and RFT performance ($r = .50$; $p < .01$). Surprisingly, boys and girls did not differ in spatial ability tests contrary to Kagan's research (as cited in McGilligan & Barclat, 1974) who found boys to be better at spatial tasks, which he explains by cultural and social influence that support boys' involvement in activities providing spatial skills practice. Mefoh, Nwoke, Chukwuorji & Chijioke (2017) tested adolescents using GEFT and a problem−solving task (a puzzle box). Their results showed a significant advantage of field independent participants over the field dependent ones in the puzzle box task. Also, they found significant results suggesting an overall advantage of male participants over female participants in problem−solving tasks. Daseking, Peterman and Waldman (2017) tested a sample of 1425 participants with WAIS−IV and found that females outperformed males in the Processing speed index, which supports research by Keith, Reynolds, Patel and Ridley (2008), who found the same results, together with results showing that males outperform females in visuospatial ability (using Woodcock−Johnson test of cognitive abilities). Based on these findings one could expect that males would outperform females in mental rotation tasks (after all we used a subtest from Woodcock−Johnson test) and in the Hidden figures subtest and that females outperform males in the Target position task (which is a task based on the speed of reactions and processing). Our findings do not support these results as we did not find any differences between genders in any of the tasks. Campbell, Toth and Brady (2018) found only very weak evidence of sex differences in the mental rotation task (cubes task and hands task) but found that male participants need less effort to complete the task (based on the pupil dilatation during the task). Ariel, Lembeck, Moffat and Hertzog (2018) found that female and male

17 e.g.: Daseking, Peterman & Waldman, 2017; McGilligan & Barclay, 1974.
18 Gagnon *et al.*, 2018.

participants did not differ in performance in various spatial tasks, but their confidence in their own performance differed significantly with females being less self–confident about their performance.

Contrary to our findings, Xu, Kim & Lewis (2016) found a significant male dominance in a spatial ability task (The Purdue Visualization of Rotations test).

An interesting point of view is presented by Sokolowski, Hawes and Lyons (2019) who explain the difference between genders in spatial and mathematical skills by a significantly bigger spatial anxiety in females. This anxiety and self–doubt in females could cause avoidance of using spatial strategies in solving problems (not only math problems), which leads to less practice of spatial ability in females. This could be the key to the difference in mathematics and spatial skills between genders (Sokolowski, Hawes & Lyons (2019). Reilly, Neumann and Andrews (2016) conclude that the difference between genders in various cognitive abilities (not only spatial skills) is probably caused by various factors together (hormones, gender, culture, social believes about genders, nurture, etc.). Reilly and colleagues (2016) for example found out that gender-identity was a stronger predictor of performance in various cognitive tasks than biological sex.

As we have not found any differences between genders in any of the tasks, we conclude that even though gender differences are still debated and shown in research, there is no strong and clear evidence of what may cause them. The fact we have not found any gender differences in our sample and the other research studies that also did not show any differences between genders (e.g.: Ariel *et al.*, 2018) may suggest that if the differences were partly caused by sociocultural influence and nurture (Meyer, Cimpian & Leslie, 2015), times might be changing (see Krajčíková, 2017). The debate about biological and physiological determinism of gender-specific abilities is still open (Pérez – Fabello & Felisberti, 2018; Reilly, Neumann & Andrews, 2016).

5.2 Limitations

In this chapter, we would like to point out some limitations, procedures or guidelines of this research that could have affected the results. We would like to acknowledge these limitations and suggest solutions for future research. **Participants** were selected with the help of faculty staff, who recruited their students, and with the help of HUME lab, which supports researchers and provides a pool of potential participants. This method is not the best one to get a heterogeneous sample regarding gender, age and field of study. At the beginning of this study, the idea was to compare participants from technical fields to those form the humanities, but we did not find enough participants from the technical side because the recruited participants were mostly from the Faculty of Arts of Masaryk University. The attempt to compare the students of humanities and those of engineering was made by Pérez–Fabello and Felisberti (2018), who found differences between strategies used by students to solve various cognitive tasks. Psychology and art students were prone to use object image processing while engineers used a spatial processing strategy. This suggests that there might be differences, however, we did not have the opportunity to examine them.

Furthermore, participants often participated in the study as part of a course, where the participation in a research study was expected and required. In this case, we have to consider *a participation bias*, which refers to the fact that a certain reward can attract a certain type of people. This tendency usually applies to research studies that recruit participants from the general population, and the type of reward offered for participation can play an important role in the recruitment of participants (Hsieh & Kocielnik, 2016).

The only reward in our study was a course credit (or part of a course credit), which raises the question of how the presence/absence of reward affects motivation, hence the performance. According to Sharp, Pelletier and Lévesque (2006), offering reward in form of course credits gets more students to participate in a study and can help create variability of demographic characteristics in the sample. Reward attracts more extrinsically motivated participants (motivation is external). If there is no reward offered for participation, the sample may contain more intrinsically oriented people, who are more self–motivated. Sharp and colleagues (2006) conclude that offering reward in form of course credits (in university setting) creates bigger and more representative samples compared to a situation where no reward is offered and only highly self–motivated individuals are willing to participate. In our case, the majority of participants were participating to get course credits, but there were still those who were recruited by word of mouth from the author's social circle. Those would be the self–motivated individuals, who participated despite no reward was offered. The ideal scenario would be if all the participants had the same reward offered for participation, but considering how difficult it was to collect a sample this big (original N = 113), we accepted everyone who met our inclusion criteria.

The **testing situation** was set in a standardized setting in a HUME lab II laboratory with standard lighting and the same computer monitors for all participants. The two computers used for testing had different keyboards. This should not be a problem, as the only subtest that required keyboard use was the Target position subtest. It is important to say that these specific keyboards may have caused the mistakes that some participants made in responding during this particular subtest (pressing a different key than instructed), although it is very unlikely and the number of participants who made this mistake is negligible.

The laboratory offers two testing spots, so we decided to test two participants at once. After we had advertised the experiment, potential participants could sign up for a specific date and time. Each testing block had two openings. After the first round of testing, we realized that some blocks were filled with only one participant, but the majority of blocks were still filled with two participants. So some participants were tested in pairs (each having their own computer) and some were tested individually. The problem here could be social facilitation (Strube, 2005; Zajonc, 1965) in testing two participants at the time. A solution to the problem could be to examine the differences between those participants who were tested separately (individually) and those tested in pairs. Unfortunately, the research assistants who were present at the testing were not instructed to make a note which participants were tested in what kind of setting (individual/pair), because we did not expect to test anyone individually. The number of participants tested individually was low, so if there was any effect of social influence, it would be manifested in the vast majority of participants. We did not include the social facilitation effect into our analyses.

The **order of the subtests** was the same for each participant, which could cause a higher level of tiredness among participants in the last subtest. A better way of dealing with this would be to create subgroups of participants and change the order of subtests for each subgroup. This method was not used due to the difficulty of the programming of the battery. We acknowledge this problem but do not find it serious, because we did not compare the results of the participant with the norms.

The **subtests** used in the study were carefully chosen. The first one that mimics Witkin's EFT is a **Hidden figures test**, which was based on Witkin's original test (1950) in its group form (GEFT). The reason why the original test could not be used is explained in *chapter 4.1.1*. The short version of EFT with 12 items (compared to the complete version of the test with 24 items) showed the same psychometric characteristics as the complete version (Jackson, 1956) and highly correlates with the 24 item version (r = .99). Using the adjusted form of EFT and not the original one was a solution that was the next best after we found out that the original was not available for research purposes. The research question does not ask for the original EFT; the study tries to find an answer, whether the EFT–like test (finding simple figures in complex designs) is in fact a test of spatial cognitive ability. Thus we consider not using the original EFT a good decision given the circumstances. The subtest we used originally consisted of 12 items (the same as the short version of EFT), but we had to exclude one item because it did not discriminate among participants (the item was extremely difficult) and it resembled one of the Witkin's items to o much. This decision, in our opinion, did not affect the results as we did not compare the scores of our participants with norms.

The **Target position subtest** was based on Gilbert *et al.*'s, (2006) study, which examines questions of language universality. The original study used different colour categories (blue/green), which could have an effect on the results. We first conducted a pilot study where our participants created the colour categories that we used afterwards for the main experiment. This process provided a better ecological validity when it comes to the experiment itself. While Gilbert & colleagues (2006) tested only 11 participants, we were able to test much more individuals after we had carefully chosen the colour categories during the pilot study. Furthermore, using the blue/green colour boundary is tricky, because the human eye can discriminate best between various green and red colour hues (Osorio & Vorobyev, 2005), which gives the colour green a clear advantage when using blue/green stimuli (Gilbert *et al.*, 2006). It is important to note that the colour category purple appears as an eighth colour category in different languages, while green appears fourth and blue right after green as a fifth category (Berlin & Kay, 1969 as cited in Kay & McDaniel, 1978) This could mean a disadvantage for the purple stimuli. Also, Gilbert *et al.* (2006) allow each participant to select a colour boundary, while we have conducted a pilot study where different individuals selected the colour categories (to avoid priming). This strategy was chosen because we wanted to be able to test more participants. This strategy was successful, compared to Gilbert's study (*et al.*, 2006) we were able to test more than 100 individuals. However, one could argue that the process of selecting the colour hues in the pilot study by different individuals than those participating in the main experiment (and selecting blue/purple colour categories) could have an effect on the results.

The last limitation of the study is linked to **the analyses** and the data we used to calculate the correlations. After careful consideration, we decided to create a new variable that reflects the overall success (*Mean RT after penalisation*) of a participant in the Mental rotation subtest and the Hidden figures subtest. This variable is strongly dependent on the participant's number of correct answers in each subtest. We acknowledge this fact and realize that this variable reflects the number of correct answers more than the reaction times, which could have a certain effect on the results. The reason why we proceeded this way is explained in chapter *4.6.3. The Reasoning behind creating a new variable.*

5.3 Applications and future research

EFT (Witkin, 1950) and tests based on finding simple figures in complex designs have been neglected and seemingly disappeared from test batteries in the last decades. This might partly be caused by the fact that different tests based on the same premise are used to test different characteristics. Some of them are tests of cognitive style (Witkin's EFT, 1950), some tests of Perceptual speed (Ekstrom, Dermen & Harman, 1976), and some are forgotten even though they represented the base for subsequent tests (Gottschaldt as cited in Witkin, 1950). This work's main goal has been to show that tests of this kind are neither a historic artifact nor too confusing to be used in practice. The focus was set on EFT (Witkin, 1950), but the results can be generalized to the tests operating on the same principle, finding simple figures in complex designs.

There has been a calling for the test that would test spatial cognitive ability in an academic setting or as a part of entrance exams for a specific field, for instance, architecture or design (Cho, 2017).

"*We believe that there is a need to design artistic spatial imagery measurement tests to access the students' skills in mental image processing, including image rotation[...]*" (Pérez–Fabello & Felisberti, 2018, p. 131).

And even though Pérez–Fabello and Felisberti (2018) are in fact talking about the need for a new *style ("[...]image cognitive styles such as an artistic spatial cognitive style related to the spatial location of objects [...]*" Pérez–Fabello & Felisberti, 2018, p. 137) and a method to test this *style*, what we propose is to use EFT–like tests for testing spatial ability manifesting in mental imagery processing, perceptual speed and mental rotation skills.

The research problem that remains for the future is the nature of the stimuli, which has not been examined here. It is necessary to explore whether all the stimuli used for EFT–like tests have the same characteristics, to consider the difference between colourful vs. monochromatic pictures (Witkin's EFT contains both as well as our version used here), to examine the difficulty of each item, the order of the items and the number of the items (as Witkin's original test contains 24 items, the short version 12 items). It would be ideal to create new norms, which are not separate for female and male participants but would differentiate different age groups (Cummings & Murray, 1987; Jackson, Messick & Myers, 1964; Lis & Powers, 1979).

Finally, addressing the Target position subtest, which was a conceptual replication of Gilbert *et al.* (2006), the results show that the linguistic categorization of colours did not help our participants to distinguish between colour hues from different colour categories. For future research, it would be interesting to include participants whose mother tongue differs and find out if the results are comparable to those obtained here. As this part of the study can be considered basic research (opposed to applied research), there is no clear application of the results. But it does not mean there will not be one in the future. So far, this study presents a contribution to the body of knowledge about the universality of colour categories.

CONCLUSION

The main goal of this thesis was to bring some clarity to the controversial concept of the field dependence/independence cognitive style (Witkin, 1950). The main focus was on the Embedded Figures test (EFT), that supposedly determines whether a person is field dependent or independent. The theoretical part is divided into several chapters. The first chapter describes cognitive styles in general and explains the differences between a style and an ability. It also explains the main patterns of figure–ground organization in visual perception. Lastly, it describes the field dependence/independence cognitive style, tests that were designed to "measure" this so-called style in an individual, and, most importantly, provides an overview over the criticism of FD/FI as a cognitive style, which is crucial for the empirical part of this thesis.

The voices of criticism of the FD/FI style frequently claim that the FD/FI style is not a style but rather an ability. Some authors talk about its link to general intelligence, some claim it is a spatial cognitive ability measured by tests such as EFT, which was originally constructed to test FD/FI.

The second chapter, inspired by the criticism of FD/FI, provides an overview of definitions of spatial cognitive ability, its link to theories of intelligence and how spatial ability is tested. It finds links between FD/FI and spatial cognitive ability, and debates the possibility that FDI is in fact part of spatial cognitive ability.

In our research, we used a methodology described by Gilbert *et al.* (2006) as control subtest. The method originally tested a hypothesis of the universality of colour categories. We conducted a conceptual replication of Gilbert's task. That is the reason why the third chapter focuses on language universality and the universality of language colour categories. This chapter described Sapir–Whorf's hypothesis, then closely explains Gilbert's study and his methodology. Lastly, the link between Gilbert's study and the rest of the thesis is explained.

The empirical part explains in detail the methodology, tests that were used, and the experimental setting. The test battery consisted of three subtests, the Hidden figures subtest (test based strictly on EFT), the Mental rotation subtest (a standardized test of spatial ability), and a Target position subtest (conceptual replication of Gilbert's methodology, which was linked to spatial ability as it required a left – right decision about the position of the target). It was hypothesized that participants' performance in the Hidden figures subtest would correlate with their performance in the Mental rotation subtest, which was confirmed. It was also hypothesized that the Target position subtest results would correlate with the rest of the battery, even though we expected a weak correlation. This hypothesis was not confirmed. We did not find any significant differences between genders in any of the subtests.

The results are debated in the discussion part of the thesis, where we concluded that the criticism of the FD/FI cognitive style, which claims that this style is in fact an ability, more specifically a part of a spatial cognitive ability, is based on solid ground. Based on our results we claim (as some before us, too) that the FD/FI is not a cognitive style but a factor of spatial cognitive ability. Due to this, we suggest using EFT and tests comparable to EFT as a tests of spatial ability.

The conceptual replication of Gilbert's study showed very interesting results. The results, however, differed from those of the original study. The results obtained here do not support Gilbert's results and his claims that the language of colour categorization helps to visually discriminate between colours. Even though the Target position subtest was "only" a control subtest in our battery, we find the results concerning this subtest very interesting.

We acknowledge the limitation of this research, which is discussed thoroughly. Lastly, the implication of the results and an inspiration for future research are debated.

To conclude, the value of this thesis lies in its suggestion to use EFT and tests based on it as a test of spatial cognitive ability in a university setting as part of an entrance examination for technical fields, in the job market as part of interviews for technical jobs and as a practice for people who are planning to improve their spatial skills.

LIST OF ABBREVIATIONS

BAT	Body–Adjustment test
CHC	Cattel–Horn–Caroll
Cf	Flexibility of closure
CFA	Confirmation factor analysis
Cs	Closure speed
EFT	Embedded Figures test
FD	Filed dependent
FI	Field independent
FD/FI	Field dependent/independent
GEFT	Group Embedded Figures test
Gv	Visual processing (C–H–C terminology) or Broad visual perception (Carroll terminology)
Il	Perceptual illusions
Im	Imagery
L–EFT	Leuven Embedded Figures test
Le	Length estimation
LRJT	Left – right judgment task
Mv	Visual memory
Pi	Serial perceptual integration
Pn	Perceptual alternations
RAT	Room–Adjustment test
RFT	Rod–and–Frame test
RRT	Rotating–Room test
S–W hypothesis	Sapir–Whorf hypothesis
Sr	Spatial relations or Speed rotation
Ss	Spatial scanning
Vz	Visualization

LIST OF TABLES

LIST OF FIGURES

REFERENCES

Alazmi, L., Gadsby, G. E., Heneghan, N. R., & Punt, T. D. (2018). Do trunk−based left/right judgment tasks elicit motor imagery? *Musculoskeletal Science and Practice, 35,* 55–60.

Ando J., Ono, Y., & Wright, M. J. (2001). Genetic structure of spatial and verbal working memory. *Behavior Genetics, 31*(6), 615–624.

Angeli, C., Valanides, N., Polemitou, E., & Fraggoulidou, E. (2016). An interaction effect between young children's field dependence−independence and order of learning with glass−box and black−box simulations: Evidence for the malleability of cognitive style in computer−supported learning. *Computers in Human Behavior, 6,* 569–583.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo−Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3.

Ariel, R., Lembeck, N. A., Moffat, S., & Hertzog, C. (2018). Are there sex differences in confidence and metacognitive monitoring accuracy for everyday, academic, and psychometrically measured spatial ability?. *Intelligence, 70,* 42–51.

Asch, S. E., & Witkin, H. A. (1948). Studies in space orientation. II. Perception of the upright with displaced visual fields and with body tilted. *Journal of Experimental Psychology, 38*(4), 455.

Ausburn, L. J., & Ausburn, F. B. (1978). Cognitive styles: Some information and implications for instructional design. *Ectj, 26*(4), 337–354.

Baddeley, A. D., & Hitch, G. J. (1974). *Working memory*. In G. H. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory*. New York, Academic Press.

Bartolomeo, P. (2006). A parietofrontal network for spatial awareness in the right hemisphere of the human brain. *Archives of Neurology, 63*(9), 1238–1241.

Bartolomeo, P., & Chokron, S. (2002). Orienting of attention in left unilateral neglect. *Neuroscience & Biobehavioral Reviews, 26*(2), 217–234.

Berenbaum, S. A., Bryk, K. L. K., & Beltz, A. M. (2012). Early androgen effects on spatial and mechanical abilities: Evidence from congenital adrenal hyperplasia. *Behavioral Neuroscience, 126*(1), 86.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross−cultural psychology: Research and applications*. Cambridge University Press.

Bigelow, G. S. (1971). Field dependence−field independence in 5 – to 10 – year – old children. *The Journal of Educational Research, 64*(9), 397–400.

Borich, G. D., & Bauman, P. M. (1972). Convergent and discriminant validation of the French and Guilford−Zimmerman spatial orientation and spatial visualization factors. *Educational and Psychological Measurement, 32*(4), 1029–1033.

Breckenridge, J. D., McAuley, J. H., Butler, D. S., Stewart, H., Moseley, G. L., & Ginn, K. A. (2017). The development of a shoulder specific left/right judgement task: validity & reliability. *Musculoskeletal Science and Practice, 28,* 39–45.

Brualdi, A. C. (1996). Multiple Intelligences: Gardner's theory. *Practical Assessment, Research & Evaluation, 5*(10), 3.

Campbell, M. J., Toth, A. J., & Brady, N. (2018). Illuminating sex differences in mental rotation using pupillometry. *Biological psychology, 138,* 19–26.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor−analytic studies*. Cambridge University Press.

Casasanto, D. (2008). Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought. *Language Learning 58*(1), 63–79.

Casasanto, D. (2005). Crying "Whorf". *Science, 307,* 1721–1722.

Cogné, M., Taillade, M., N'Kaoua, B., Tarruella, A., Klinger, E., Larrue, F.,... & Sorita, E. (2017). The contribution of virtual reality to the diagnosis of spatial navigation disorders and to the study of the role of navigational aids: A systematic literature review. Annals of physical and rehabilitation medicine, 60(3), 164–176.

Constantinidis, C., & Wang, X. J. (2004). A neural circuit basis for spatial working memory. *The Neuroscientist, 10*(6), 553–565.

Conway, A. R., & Kovacs, K. (2015). New and emerging models of human intelligence. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(5), 419–426.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. *Visual information processing,* 75–176.

Cooper, L. A. (1975). Mental rotation of random two−dimensional shapes. *Cognitive psychology, 7*(1), 20–43.

Cooper, L. A. (1976). Demonstration of a mental analog of an external rotation. *Perception & Psychophysics, 19*(4), 296–302.

Cooper, L. A., & Shepard, R. N. (1975). Mental transformation in the identification of left and right hands. *Journal of Experimental psychology: Human perception and performance, 1*(1), 48.

Cronbach, L., & Snow, R. (1981). *Aptitudes and instructional methods. A handbook for research on interactions.* NewYork: Irvington Publishers.

Cummings, A. L., & Murray, H. G. (1987). Psychometric data on the Group Embedded Figures Test for a sample of adult learners. *Perceptual and motor skills, 65*(2), 583–586.

Čeněk, J. (2015). *Interkulturní rozdíly ve zrakovém vnímání*. Doctoral Dissertation, Masaryk University. Retrieved from: https://is.muni.cz/th/f4aw4?info = 1;zpet = %2Fvyhledavani %2F %3Fsearch %3DInterkulturn %C3 %AD %20rozd %C3 %ADly %20ve %20zrakov %C3 %A9m %20vn %C3 %ADm %C3 %A1n %C3 %AD %20agenda:th %26start %3D1

Daseking, M., Petermann, F., & Waldmann, H. C. (2017). Sex differences in cognitive abilities: analyses for the german WAIS−IV. *Personality and Individual Differences, 114,* 145–150.

David, L. T. (2012). Training effects on mental rotation, spatial orientation and spatial visualisation depending on the initial level of spatial abilities. *Procedia−Social and Behavioral Sciences, 33,* 328–332.

Davidoff, J., & Warrington, E. K. (2001). A particular difficulty in discriminating between mirror images. *Neuropsychologia, 39*(10), 1022–1036.

De-Witt, L. H., Huygelier, H., Van der Hallen, R., Chamberlain, R., & Wagemans, J. (2017). Developing the Leauven Embedded Figures Test (L–EFT): Testing the stimulus features that influence embedding. *PeerJ, 5*(2862).

Delta–E Calculator. (n.d.). Available from: http://colourmine.org/delta–e–calculator.

Depth Perception. (n.d.). [Picture]. Available from: https://bit.ly/2M9Ogsh.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences, 104*(3), 1097–1102.

EasyRGB. (n.d.). Available from: https://www.easyrgb.com/en/.

Ekstrom, R. B., Dermen, D., & Harman, H. H. (1976). *Manual for kit of factor–referenced cognitive tests (Vol. 102)*. Princeton, NJ: Educational Testing Service.

Emmett, D., Clifford, B. R., & Gwyer, P. (2003). An investigation of the interaction between cognitive style and context reinstatement on the memory performance of eyewitnesses. *Personality and Individual Differences, 34*(8), 1495–1508.

Eschen, H., Kötter, T., Rodeck, R., Harnisch, M., & Schüppstuhl, T. (2018). Augmented and Virtual Reality for Inspection and Maintenance Processes in the Aviation Industry. *Procedia Manufacturing, 19,* 156–163.

Eysenck, M. W., & Keane, M. T. (2008). *Kognitivní psychologie*. Praha: Academia.

Fitzgibbons, D., Goldberger, L., & Eagle, M. (1965).Field dependence and memory for incidental material. *Perceptual and Motor Skills, 21,* 743–749.

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. (2008). Lateralization of categorical perception of colour changes with colour term acquisition. *PNAS 105*(47), 18221–18225.

Gagnon, K. T., Thomas, B. J., Munion, A., Creem–Regehr, S. H., Cashdan, E. A., & Stefanucci, J. K. (2018). Not all those who wander are lost: Spatial exploration patterns and their relationship to gender and spatial memory. *Cognition, 180,* 108–117.

Gardner, H. (1999). Dimenze *myšlení: teorie rozmanitých iteligencí*. Praha, Portál, sro.

Geary, D. C. (2014). Evolved sex differences in modern context. *New Male Studies: An International Journal, 3*(5), 20.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of colour discrimination. *Brain and language, 105*(2), 91–98.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R.B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *PNAS, 103*(2), 489–494.

Goldstein, A. G., & Chance, J. E. (1965). Effect of practice on sex – related differences in performance on Embedded Figures. *Psychon. Sci., 3,* 361–362.

Goodenough D. R. (1976). The role of individual differences in field dependence as a factor in learning and memory. *Psychological Bulletin, 83*(4), 675–694.

Goodenough, D., & Witkin, H. (1977). Origins of the field–dependent and field–independent cognitive styles. ETS *Research Bulletin Series, 1977*(1), 1–80.

Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science, 306,* 496–499.

Górska, R. (2005). Spatial Imagination an Overview of the Longitudinal Research at Cracow University of Technology. *Journal for Geometry and Graphics, 9*(2), 201–208.

Grajewski, D., Górski, F., Zawadzki, P., & Hamrol, A. (2013). Application of virtual reality techniques in the design of ergonomic manufacturing workplaces. *Procedia Computer Science, 25,* 289–301.

Guisande, M. A., Páramo, M. F., Tinajero, C., & Almeida L. S. (2007). Field dependence independence (FDI) cognitive style: An analysis of attentional functioning. *Psicothema, 19*(4), 572–577.

Hao, X., Wang, K., Li, W., Yang, W., Wei, D., Qiu, J., & Zhang, Q. (2013). Individual differences in brain structure and resting brain function underlie cognitive styles: Evidence from the Embedded Figures Test. *PLoS ONE, 8*(12), 1–9.

Hardiess, G., Mallot, H. A., & Meilinger, T. (2015). Virtual reality and spatial cognition. *International Encyclopedia of the Social and Behavioral Sciences, 2nd edn., Wright JD (Ed.)*, 133–137.

Hardin, C. L. (2013). Berlin and Kay Theory. *Encyclopedia of Colour Science and Technology,* 1–4.

Healy, S. D., Harkins, D. A., & Ray, S. (2010). Battle within: Role of cognitive style in conflict resolution strategies of male college students . *American Journal Of Psychological Research*, *6*(1), 157–192.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective−taking spatial abilities. *Intelligence, 32*(2), 175–191.

Hergovich, A. (2003). Field dependence, suggestibility and belief in paranormal phenomena. *Personality and Individual Differences, 34,* 195–209.

Hock, H. S., & Tromley, C. L. (1978). Mental rotation and perceptual uprightness. *Perception & Psychophysics, 24*(6), 529–533.

Hochberg, J., & Gellman, L. (1977). The effect of landmark features on mental rotation times. *Memory & Cognition, 5*(1), 23–26.

Hsieh, G., & Kocielnik, R. (2016). You get who you pay for: The impact of incentives on participation bias. In *Proceedings of the 19th ACM Conference on Computer−Supported Cooperative Work & Social Computing*, 823–835.

Huygelier, H., Van der Hallen, R., Wagemans, J., de-Wit, L., & Chamberlain, R. (2018). The Leuven Embedded Figures Test (L−EFT): Measuring perception, intelligence or executive functions? *PeerJ, 6* (4524).

Chamberlain, R., Van der Hallen, R., Huygelier, H., Van de Cruys, S., & Wagemans, J. (2017). Local−global processing bias is not a unitary individual difference in visual processing. *Vision Research, 141*, 247–257.

Chiang, T. (1998). *Story of your life. Stories of your life and others*, Picador, 117–78.

Chica, A. B., de Schotten, M. T., Toba, M., Malhotra, P., Lupiáñez, J., & Bartolomeo, P. (2012). Attention networks and their interactions after right−hemisphere damage. *Cortex, 48*(6), 654–663.

Cho, J. Y. (2017). An investigation of design studio performance in relation to creativity, spatial ability, and visual cognitive style. *Thinking Skills and Creativity, 23*, 67–78.

Chraif, M. (2013). Gender differences in mental rotation at young Romanian students at psychology−a pilot study. *Procedia−Social and Behavioral Sciences, 78,* 692–696.

Ichise, R. (2018). A Cognitive Architecture Consisting of Human Intelligence Factors. *Procedia Computer Science, 123,* 165–170.

Imai, M. (2017). *Jazyk a myšlení*. Charles University in Prague, Karolinum Press.

Jackson, D. N. (1956). A short form of Witkin's Embedded−Figures Test. *Journal of abnormal and social psychology, 53,*. 254–255.

Jackson, D. N., Messick, S., & Myers, C. T. (1964). Evaluation of group and individual forms of embedded figures measures of field−independence. *Educational and Psychological Measurement, 24,* 177–192.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology, 8*(4), 441–480.

Kay, P. & Kempton, W. (1984). What is the Sapir−Whorf Hypothesis? *American Anthropologist, 86,* 65–79.

Kay, P. & McDaniel, Ch. K. (1978). The Linguistic Significance of the Meanings of Basic Colour Terms. *Language, 54*(3), 610–646.

Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence, 36*(6), 502–525.

Khatib, M., & Hosseinpur, R. M. (2011). On the Validity of the Group Embedded Figure Test (GEFT). *Journal of Language Teaching & Research, 2*(3), 640–648.

Khodadady, E. (2012). Field-dependence/independence cognitive style and performance on the IELTS listening comprehension. *International Journal of Linguistics, 4*(3), 622–635.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition, 29*(5), 745–756.

Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: Toward an integrated framework of cognitive style. *Psychological Bulletin, 133*(3), 464–481.

Krajčíková, L. (2014). *Field independence as a spatial cognitive ability: Correlation study of verbal and spatial short–term memory.* Master's Thesis, Masaryk University. Retrieved from: https://is.muni.cz/th/dqjgb?info = 1;zpet = %2Fvyhledavani %2F %3Fsearch %3Dkraj %C4 %8D %C3 %ADkov %C3 %A1 %26start %3D1

Krajčíková, L. (2017). *„Gender gap" v* kontexte zahraničného a tuzemského vysokého školství. In Doktorské štúdium– Súčasný stav a perspektívy I.: Zborník vedeckých prác. 1. vyd. Prešov: Prešovská univerzita, pp. 82-92. ISBN: 978–80–555–1835–0. URL: http://www.pulib.sk/web/kniznica/elpub/dokument/Lukacova1.

Lindsey, D. T. & Brown A. M. (2002). Colour Naming and the Phototoxic Effects of Sunlight on the Eye. *Psychological Science, 13*(6), 506–512.

Lis, D. J., & Powers, J. E. (1979). Reliability and validity of the Group Embedded figures test for a grade school sample. Perceptual and Motor Skills, 48, 660–662.

Lohman, D. F. (1996). Spatial ability and *g*. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Ludwig, I., & Lachnit, H. (2004). Effects of practice and transfer in the detection of embedded figures. *Psychological Research, 68,* 277–288.

Lutchmaya, S., Baron–Cohen, S., & Raggatt, P. (2001). Foetal testosterone and vocabulary size in 18–and 24–month–old infants. *Infant Behavior and Development, 24*(4), 418–424.

Mackey, A. (2012). Why (or why not), when and how to replicate research. *Replication research in applied linguistics*, 2146.

MacLeod, C.M., Jackson, R.A. & Palmer, J. (1986). On the relation between spatial ability and field dependence. *Intelligence, 10,* 141–151.

Marks, D.F. (1999). Consciousness, Mental Imagery and Action. *British Journal of Psychology* (90), 567–585.

Martín–Gutiérrez, J., Contero, M., & Alcañiz, M. (2015). Augmented reality to training spatial skills. *Procedia Computer Science, 77,* 33–39.

Massa, L. J., Mayer, R. E., & Bohon, L. M. (2005). Individual differences in gender role beliefs influence spatial ability test performance. *Learning and Individual Differences, 15*(2), 99–111.

McGilligan, R. P., & Barclat, A. G. (1974). Sex differences and spatial ability factors in Witkin's "differentiation" construct. *Journal of Clinical Psychology, 30*(4), 528–532.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10.

McKenna, F. P. (1984). Measures of field dependence: Cognitive style or cognitive ability? *Journal of Personality and Social Psychology, 47*(3), 593.

Mefoh, P. C., Nwoke, M. B., Chukwuorji, J. C., & Chijioke, A. O. (2017). Effect of cognitive style and gender on adolescents' problem solving ability. Thinking Skills and Creativity, 25, 47–52.

Messick, S. (1976). Personal consistencies in cognition and creativity. In Messick and Associates, S. (Ed.), *Individuality and Learning*. San Francisco: Jossey-Bass.

Messick, S., & Damarin, F. (1964). Cognitive styles and memory for faces. *The Journal of Abnormal and Social Psychology, 69*(3), 313.

Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 10*(4), 309–325.

Meyer M, Cimpian A., & Leslie S−J. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology, 6,* 1–12.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent−variable analysis. *Journal of experimental psychology: General, 130*(4), 621.

Miyake, A., Witzki, A. H., & Emerson, M. J. (2001). Field dependence−independence from a working memory perspective: A dual−task investigation of the Hidden Figures Test. *Memory, 9*(4/5/6), 445–457.

Moskowitz, D. S., Dreyer, A. S., & Kronsberg, S. (1981). Preschool children's field independence: Prediction from antecedent and concurrent maternal and child behavior. *Perceptual and Motor Skills, 52*, 607–616.

Neumann, D. L., Fitzgerald, Z. T., Furedy, J. J., & Boyle, G. J. (2007). Sexually dimorphic effects of acute nicotine administration on arousal and visual−spatial ability in non−smoking human volunteers. *Pharmacology Biochemistry and Behavior, 86*(4), 758–765.

Nigro, G., Cicogna, P. C., D'Olimpio, F., & Cosenza, M. (2012). The role of visual perceptual style and personality disorder traits in event−based prospective memory. *Personality and Individual Differences, 53*(7), 912–916.

Nisbett, R., & Masuda, T. (2003). Culture and point of view. *Proceedings of the National Academy of Sciences, 100*(19), 11163–11170.

Osorio, D., & Vorobyev, M. (2005). Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and colour vision. *Proceedings of the Royal Society of London B: Biological Sciences, 272*(1574), 1745–1752.

Oxford dictionaries [online]. [cit. 2018–07–04]. Available from: https://en.oxforddictionaries.com/definition/

Panek, P. E., Funk, L. G., & Nelson, P. K. (1980). Reliability and validity of the Group Embedded figures test across the lifespan. *Perceptual and Motor Skills, 50,* 1171–1174.

Pérez–Fabello, M. J., & Felisberti, F. (2018). Object–spatial imagery in fine art, psychology, and engineering. *Thinking Skills and Creativity, 27,* 131–138.

Peterson, E. R., Deary, I. J., & Austin, E. J. (2005). Are intelligence and personality related to verbal–imagery and wholistic–analytic cognitive styles? *Personality and Individual Differences, 39*(1), 201–213.

Preiss, M. (1999). Paměťový test učení. Bratislava: Psychodiagnostika.

Pylyshyn, Z. W. (1979). The rate of “mental rotation” of images: A test of a holistic analogue hypothesis. *Memory & Cognition, 7*(1), 19–28.

Pylyshyn, Z. (2003). Return of the mental image: are there really pictures in the brain? *Trends in cognitive sciences, 7*(3), 113–118.

Quaiser–Pohl, C., Lehmann, W., & Eid, M. (2004). The relationship between spatial abilities and representations of large–scale space in children– a structural equation modeling analysis. *Personality and Individual Differences, 36*(1), 95–107.

Reilly, D., Neumann, D. L., & Andrews, G. (2016). Sex and sex-role differences in specific cognitive abilities. *Intelligence, 54,* 147–158.

Reger, G. M., McGee, J. S., van der Zaag, C., Thiebaux, M., Buckwalter, J. G., & Rizzo, A. A. (2003). A 3D virtual environment rod and frame test: the reliability and validity of four traditional scoring methods for older adults. *Journal of clinical and experimental neuropsychology, 25*(8), 1169–1177.

Regier, T., & Kay, P. (2009). Language, thought, and colour: Whorf was half right. *Trends in Cognitive Sciences 30*(10), 1–8.

Richardson, J. A., & Turner, T. E. (2000). Field dependence revisited I: Intelligence. *Educational Psychology, 20*(3), 255–270.

Rittschof, K. A. (2010). Field dependence−independence as visuospatial and executive functioning in working memory: implications for instructional systems design and research. *Educational Technology Research and Development, 58*(1), 99–114.

Roberson, D. & Hanley, J. R. (2010). *Relatively speaking: An account of the relationship between language and thought in the colour domain*. In: Malt, B & Wolff, P. (eds.) *Words and the Mind: How words capture human experience*. Oxford University Press, New York, pp. 183-198. ISBN 9780195311129

Román, F. J., Abad, F. J., Escorial, S., Burgaleta, M., Martínez, K., Álvarez-Linera, J.,... & Colom, R. (2014). Reversed hierarchy in the brain for general and specific cognitive abilities: A morphometric analysis. *Human brain mapping, 35*(8), 3805-3818.

Ryder, A., Levy, S., Levine, D., Lunder, K., Linde, D. (Producers) & Villeneuve, D. (Director). (2016). *Arrival* [Motion picture]. USA: Paramount Pictures.

Sharp, E. C., Pelletier, L. G., & Lévesque, C. (2006). The double−edged sword of rewards for participation in psychology experiments. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 38*(3), 269.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three−dimensional objects. *Science, 171*(3972), 701–703.

Shepard, S., & Metzler, D. (1988). Mental rotation: effects of dimensionality of objects and type of task. Journal of experimental psychology: *Human perception and performance, 14*(1), 3.

Shi, Y. (2018). The puzzle of missing female engineers: Academic preparation, ability beliefs, and preferences. *Economics of Education Review, 64,* 129–143.

Shulman, G. L., Pope, D. L., Astafiev, S. V., McAvoy, M. P., Snyder, A. Z., & Corbetta, M. (2010). Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *Journal of Neuroscience, 30*(10), 3640–3651.

Schneider, J., & McGrew, K. (2013). The Cattell−Horn−Carroll (CHC) model of intelligence v2.2: A visual tour and summary. *Institute for Applied Psychometrics (IAP), 1,* 03–13.

Smetackova, I. (2015). Gender stereotypes, performance and identification with math. *Procedia−Social and Behavioral Sciences, 190,* 211–219.

Sokolowski, H. M., Hawes, Z., & Lyons, I. M. (2019). What explains sex differences in math anxiety? A closer look at the role of spatial processing. *Cognition, 182,* 193–212.

Solar, D., Davenport, G., & Bruehl, D. (1969). Social compliance as a function of field dependence. *Perceptual and Motor Skills, 29*(1), 299–306.

Sternberg, R. J., & Grigorenko, E. L. (1997). Are cognitive styles still in style? *American sychologist, 52*(7), 700–712.

Strube, M. J. (2005). What did Triplett really find? A contemporary analysis of the first experiment in social psychology. *The American journal of psychology, 118*(2), 271–286.

Sturm, W., & Willmes, K. (2001). On the functional neuroanatomy of intrinsic and phasic alertness. *Neuroimage, 14*(1), 76–84.

Šikl, R. (2012). Zrakové vnímání. Grada Publishing as.

Tascón, L., Castillo, J., León, I., & Cimadevilla, J. M. (2018). Walking and non−walking space in an equivalent virtual reality task: Sexual dimorphism and aging decline of spatial abilities. *Behavioural brain research, 347,* 201–208.

Taub, G. E., & McGrew, K. S. (2014). The Woodcock−Johnson Tests of Cognitive Abilities III's cognitive performance model: Empirical support for intermediate factors within CHC theory. *Journal of Psychoeducational Assessment, 32*(3), 187–201.

The original version of Kanisza square. (2011). [Picture]. Available from: https://bit.ly/2RsekVs.

Tinajero, C., & Páramo, M.F. (1997). Field dependence−independence and academic achievement: a re−examination of their relationship. *British Journal of Educational Psychology, 67,* 199–212.

Tinti, C., Chiesa, S., Cavaglià, R., Dalmasso, S., Pia, L., & Schmidt, S. (2018). On my right or on your left? Spontaneous spatial perspective taking in blind people. *Consciousness and Cognition, 62,* 1–8.

Toivainen, T., Papageorgiou, K. A., Tosto, M. G., & Kovas, Y. (2017). Sex differences in non−verbal and verbal abilities in childhood and adolescence. *Intelligence, 64,* 81–88.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three−dimensional spatial visualization. *Perceptual and motor skills, 47*(2), 599–604.

Vasta, R., & Liben, L. S. (1996). The water−level task: An intriguing puzzle. *Current Directions in Psychological Science, 5*(6), 171–177.

Vecera, S., Vogel, E., & Woodman, G. (2002). Lower region: A new cue for figure−ground assignment. *Journal of Experimental Psychology: General, 131*(2), 194–205.

Vernon, P. E. (1950). *The structure of human abilities*. T. & A. Constable Ltd., Edinburgh.

Waber, D.P. (1977). Biological substrates of field dependence: Implications of sex difference. *Psychological Bulletin, 84*(6), 1076–1087.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M. *et al.* (2012 a). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure−ground organization. *Psychological Bulletin, 138*(6), 1172–1217.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012b). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological bulletin, 138*(6), 1218.

Wang, C., Tian, Y., Chen, S., Tian, Z., Jiang, T., & Du, F. (2014). Predicting performance in manually controlled rendezvous and docking through spatial abilities. *Advances in Space Research, 53*(2), 362–369.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on colour discrimination. *Proceedings of the National Academy of Sciences, 104*(19), 7780–7785.

Witkin, H. A. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality, 19*(1), 1–15.

Witkin, H. A. (1965). Psychological differentiation and forms of pathology. *Journal of Abnormal Psychology, 70*(5), 317–336.

Witkin, H., & Berry, J. (1975). Psychological differentiation in cross−cultural perspective. *ETS Research Bulletin Series, 1975*(1), 1–100.

Witkin H., & Goodenough D. R. (1977*). Field Dependence Revisited.* Princeton, New Jersey: Educational Testing Service, 1–53.

Witkin, H. A., Goodenough, D. R., & Oltman, P.K. (1979). Psychological differentiation: Current status. *Journal of Personality and Social Psychology, 37*(7), 1127–1145.

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1977). Field−dependent and field−independent cognitive styles and their educational implications. *Review of educational research, 47*(1), 1–64.

Wolff, P. & Holmes, K. J. (2011). Linguistic Relativity. *WIREs Cognitive Science, 2,* 253-265.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III*. Itasca. IL: Riverside.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2003). *Woodcock–Johnson III diagnostic supplement to the tests of cognitive abilities*. Itasca, IL: Riverside, 10.

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review, 15*(4), 763–771.

Xia, T., Xu, G., & Mo, L. (2018). Bi−lateralized Whorfian effect in colour perception: Evidence from Chinese Sign Language. *Journal of Neurolinguistics, 49,* 189–201.

Xu, X., Kim, E. S., & Lewis, J. E. (2016). Sex difference in spatial ability for college students and exploration of measurement invariance. *Learning and Individual Differences, 45,* 176–184.

Xue, J., Li, C., Quan, C., Lu, Y., Yue, J., & Zhang, C. (2017). Uncovering the cognitive processes underlying mental rotation: an eye−movement study. *Scientific Reports, 7*(1), 73–76.

Yeo, R. A., Ryman, S. G., Thompson, M. E., van den Heuvel, M. P., de Reus, M. A., Pommy, J., ... & Jung, R. E. (2016). Cognitive specialization for verbal vs. spatial ability in men and women: Neural and behavioral correlates. *Personality and Individual Differences, 102,* 60–67.

Zacks, J. M., Mires, J., Tversky, B., & Hazeltine, E. (2000). Mental spatial transformations of objects and perspective. *Spatial Cognition and Computation, 2*(4), 315–332.

Zajonc, R. B. (1965). The requirements and design of a standard group task. *Journal of Experimental Social Psychology, 1*(1), 71–88.

Zhang, L. F. (2002). Measuring thinking styles in addition to measuring personality traits? *Personality and Individual Differences, 33*(3), 445–458.

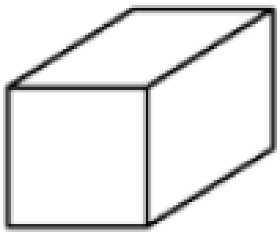
Zhang, L., (2004). Field−dependence/independence: cognitive style or perceptual ability?−validating against thinking styles and academic achievement. *Personality and Individual Differences, 37,* 1295–1311.

Zimney, K. J., Wassinger, C. A., Goranson, J., Kingsbury, T., Kuhn, T., & Morgan, S. (2018). The reliability of card−based and tablet−based left/right judgment measurements. *Musculoskeletal Science and Practice, 33,* 105–109.

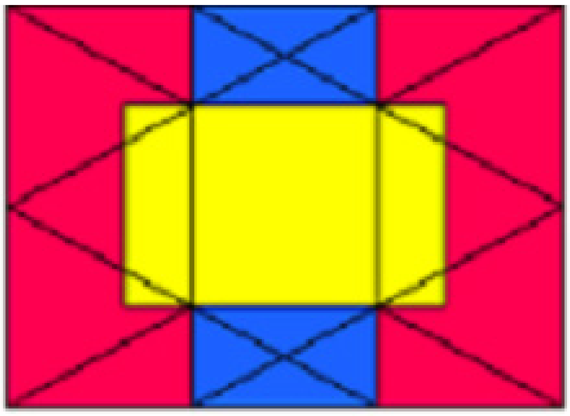
ATTACHMENTS

Attachment 1

The example of the material of the Hidden Figures subtest.



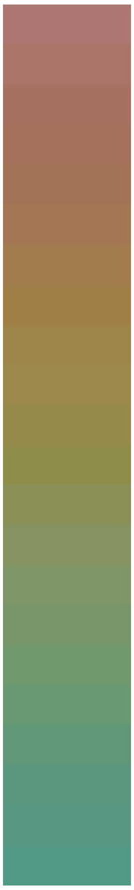
Simple figure

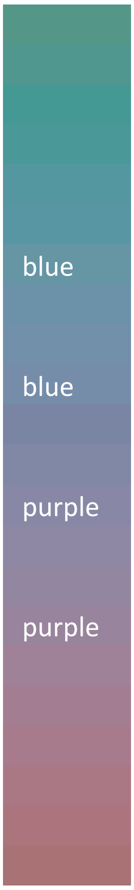


Complex design

Attachment 2

The colours used at the pilot study for the Gilbert *et al.* (2006) replication. Chosen colours are marked.





rgb(172. 118. 114)		rgb(84. 151. 137)
rgb(170. 116. 105)		rgb(80. 151. 143)
rgb(167. 113. 97)		rgb(69. 153. 149)
rgb(164. 114. 92)		rgb(75. 152. 152)
rgb(163. 115. 86)		rgb(85. 151. 160)
rgb(163. 119. 82)		rgb(89. 150. 163)
rgb(164. 123. 78)	blue	**rgb(102. 149. 164)**
rgb(161. 125. 72)		rgb(108. 146. 169)
rgb(159. 132. 74)		rgb(114. 144. 170)
rgb(156. 138. 76)	blue	**rgb(118. 141. 170)**
rgb(149. 137. 76)		rgb(121. 133. 162)
rgb(143. 140. 74)		rgb(128. 136. 165)
rgb(139. 144. 88)	purple	**rgb(136. 136. 166)**
rgb(133. 147. 96)		rgb(143. 136. 164)
rgb(126. 149. 102)		rgb(147. 134. 160)
rgb(120. 150. 106)	purple	**rgb(152. 132. 156)**
rgb(113. 152. 109)		rgb(160. 130. 152)
rgb(105. 153. 115)		rgb(163. 126. 146)
rgb(96. 152. 119)		rgb(167. 123. 140)
rgb(91. 150. 126)		rgb(170. 119. 132)
rgb(89. 151. 132)		rgb(172. 117. 126)
rgb(84. 153. 136)		rgb(169. 115. 117)

Attachment 3

The normality tests (Kolmogorov−Smirnov test and Shapiro−Wilk test) for the variable *Mean RT for each version of the condition* of the Target position subtest.

Tests of Normality for the versions of the condition A: 12 and 21.

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
Mean RT_12	.237	106	.000	.532	106	.000
Mean RT_21	.250	106	.000	.521	106	.000

Tests of Normality for the versions of the condition B: 23 and 32.

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
Mean RT_23	.262	106	.000	.406	106	.000
Mean RT_32	.243	106	.000	.422	106	.000

Tests of Normality for the versions of the condition C: 34 and 43.

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
mean 34	.201	106	.000	.735	106	.000
mean 43	.150	106	.000	.839	106	.000

Tests of Normality for the versions of the condition D: 13 and 31.

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
Mean RT_13	.187	106	.000	.639	106	.000
Mean RT_31	.209	106	.000	.596	106	.000

Tests of Normality for the versions of the condition E: 14 and 41.

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
Mean RT_14	.156	106	.000	.755	106	.000
Mean RT_41	.227	106	.000	.565	106	.000

Tests of Normality for the versions of the condition F: 24 and 42.

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
mean 24	.166	106	.000	.732	106	.000
mean 42	.208	106	.000	.640	106	.000

Attachment 4

The spatial cognitive ability factors according to different authors and they resemblance in the work of other authors.

Hegarty & Waller's (2004) factors	Lohman's (1996) factors
Mental Rotation factor	Visualization
	Speed Rotation
Perceptual Speed factor (Miyake, Witzki & Emerson.2001; Ekstrom, Dermen & Harman, 1976)	Closure Speed
	Closure Flexibility
	Perceptual Speed
Spatial orientation factor	Not discussed