Univerzita Karlova
Pedagogická fakulta

DISERTAČNÍ PRÁCE

2020 David Janda

Univerzita Karlova
Pedagogická fakulta

DISERTAČNÍ PRÁCE

Simple categorization of mathematical objects: Examining students' decisions

Jednoduchá kategorizace matematických objektů: zkoumání rozhodnutí žáků a studentů

David Janda

Vedoucí práce: doc. RNDr. Naďa Vondrová, Ph.D.

Studijní program: Pedagogika

Studijní obor: Didaktika matematiky

2020

I, the author, declare that the work "Simple categorization of mathematical objects: Examining students' decisions" is my own and has not been submitted for a degree at any other institution. None of the work has previously been published in this form.

In Prague, 30. 4. 2020

....

Prohlašuji, že jsem disertační práci na téma „Jednoduchá kategorizace matematických objektů: zkoumání rozhodnutí žáků a studentů" vypracoval pod vedením školitelky samostatně za použití v práci uvedených pramenů a literatury. Dále prohlašuji, že tato práce nebyla využita k získání jiného nebo stejného titulu.

V Praze, 30. 4. 2020

....

Acknowledgement

The greatest thanks to Naďa Vondrová, an excellent supervisor and an excellent person. Her support and advice in the domain of mathematics education can perhaps only be matched by her support and advice in my personal life (especially in the hard times). Great thanks to Adrian Simpson for a lot of time spent in the initial times of my research. Great thanks also to my wife Aneta who supported me all these years and never doubted. And finally, great thanks to my colleagues, friends and family.

This thesis became a part of many stories with all of you, including help, support, fun and friendship. Thus, it does my effort was worth it.

Poděkování

Největší díky patří Nadě Vondrové, excelentní školitelce, která se ukázala být ještě lepším člověkem. Její podpora a rada na poli didaktiky matematiky může být srovnávána snad jen s její podporou a radou v osobním životě (zvláště v těch těžkých chvílích). Velké díky Adrianu Simpsonovi za inspirativní diskuse v počátcích mého výzkumu. Velké díky také mé manželce Anetě, která mě všechny ty roky podporovala a nikdy nepochybovala. A konečně, velké díky mým kolegům, přátelům a rodině.

Tato práce se stala součástí mnoha zážitků s vámi všemi, které zahrnují pomoc, podporu, legraci a přátelství. Proto to stálo zato.

ABSTRAKT

Cílem disertační práce je popsat rozhodovací proces žáků a studentů při tzv. jednoduché kategorizaci, neboli rozhodnutí, zda konkrétní objekt je, nebo není prvkem dané kategorie. Tento proces je přitom zkoumán v kontextu kategorií matematických objektů. V teoretické části práce jsou představeny argumenty, proč je zkoumání jednoduché kategorizace matematických objektů důležité pro didaktiku matematiky. Tyto argumenty přitom nevycházejí pouze z dostupné literatury v didaktice matematiky, ale částečně čerpají také z literatury historické, matematické a psychologické. V prakticky zaměřených kapitolách práce je popsán návrh a pilotáž výzkumného nástroje vhodného ke zkoumání jednoduché kategorizace. Dominantními prvky tohoto nástroje je měření binárních odpovědí (ano/ne) respondenta a jeho reakčního času. Tento nástroj je poté využit v hlavní studii se smíšenou, kvalitativně-kvantitativní metodologií. Bylo zjištěno, že pomocí navrženého nástroje je při dodržení vhodných metodologických pravidel možno rozlišit různé přístupy respondentů ke kategorizaci. Navíc byly popsány základní vzory v rozhodovacím procesu respondentů. Těmi jsou například rozdíly v kategorizaci příkladů a „nepříkladů“, rozdíly v kategorizaci jednodušších a komplikovanějších objektů, vztah mezi počty správných odpovědí respondenta a jeho reakčními časy apod. Tato zjištění mohou být důležitá pro další výzkum v této oblasti a následné aplikace. Takovou aplikací může být například porovnávání kvality mentální reprezentace konceptu respondentů pomocí jednoduché kategorizace.

KLÍČOVÁ SLOVA

Katogorizace, kategorie, matematické objekty, matematické koncepty, reprezentace, definice, představa

ABSTRACT

The aim of the thesis is to describe the decision making process of students in the so-called simple categorization, i.e., decision whether a particular object is or is not an element of a category. This process is examined in the context of categories of mathematical objects. The theoretical part of the thesis presents arguments why the study of simple categorization of mathematical objects is important for mathematics education. These arguments are not only based on the available literature in mathematics education, but also partly draw on historical, mathematical and psychological literature. The practical chapters of the thesis describe the design and piloting of a research tool suitable for this research. The dominant elements of this tool are the measurement of the binary answers (yes / no) of the respondent and of his/her reaction time. This tool is then used in the Main study based on mixed, qualitative-quantitative methodology. It was found that with the help of the proposed tool, while adhering to appropriate methodological rules, it is possible to distinguish different approaches of respondents to categorization. In addition, the basic patterns in the decision-making process of the respondents were described. These are, for instance, differences in the categorization of examples and non-examples, differences in the categorization of simple and more complicated objects, the relationship between the number of correct responses of respondents and their reaction times, etc. These findings may be important for further research in this area and subsequent applications. Such an application can be, for example, comparing the quality of respondents' mental concept representations using simple categorization.

KEYWORDS

Categorization, categories, mathematical objects, mathematical concepts, representations, concept image, concept definition

Table of content

Introduction

Consider the task below:

Decide which of the following equations represent a function.

$$x^2 + y^2 = 1$$

$$y = x$$

$$x - y = x + y$$

$$y = \sqrt{-x}$$

This simple task represents a possible way a teacher can examine or evaluate a part of a student's image of a function concept. Of course, the task does not cover the concept in its entirety; nevertheless, it can give the teacher a rough idea about some aspects of the concept in a student's mind. In this task, it is mostly the knowledge of the necessary condition for accepting the expression as a function: that for one element of the domain, there exists one and only one element of the range.

Using a standard label x for independent and y for dependent variables, the solution of the task requires recalling, realization and use of a relatively wide range of knowledge. This includes, among others, the answers to these questions: Which rules of equivalent equations can be applied? What is the domain of the expression $\sqrt{-x}$?, How many numbers fit the equation $x^2 + y^2 = 1$?, etc. However, after all these considerations, there is a simple "yes or no" decision, whether the equation represents a function or not.

This thesis is focused on this kind of decision in the domain of mathematics education, calling it *simple categorization*. The task "to categorize something" is understood as requiring a decision whether an object is a member of the category corresponding to a concept. The theoretical chapter will show that it can sometimes be complicated. In the practical chapters of the thesis (Chapter 2 and Chapter 3), information obtained when students are assigned simple categorization tasks will be examined.

The theoretical chapter of the thesis consists of several sections. Basic terms, serving as starting points of the thesis, such as mathematical concepts and mathematical objects, are discussed in Section 1.1. Three sections follow which represent three interrelated domains connected to the topic of simple categorization of mathematical objects.

Section 1.2 is dedicated to the way mathematical concepts are represented in an individual's mind. However, the literature uses words "represented" and "representation" in various ways. For this reason, three different ways of dealing with a mathematical concept are distinguished. *Concept representation* is focused on the static, current image of a concept in an individual's mind, whereas *concept formation*[1] describes how a concept representation is developed in one's mind. Finally, *concept processing* consists of dealing with a concept by a person at a concrete moment. This distinction is inspired by two sources – a cognitive psychology approach (Sternberg & Sternberg, 2012) and the distinction of three components of understanding (building, having and enacting) presented by Duffin and Simpson (2000). Arguments are given for the examination of mathematical concepts in these three perspectives using simple categorization as a research tool.

Two common views of mathematical concepts are discussed in Sections 1.3 and 1.4. In Section 1.3, mathematical concepts are seen as categories of mathematical objects. Basic characteristics of such categories are described and their representations in the mind of the individual are discussed. In Section 1.4, mathematical concepts are approached from the perspective of their definitions. However, it is important not to see mathematical definition as a static entity, but rather as a product of concept development. Thus, a historical perspective is also considered and relevant ideas of authors who examined the development of mathematical concepts are presented using examples. Next, this perspective is connected to the domain of mathematics education.

The main intention of the practical chapters of the thesis – to examine concept representation of school mathematical concepts through students' categorization – arises from this theoretical analysis and is summarized in Section 1.5. This intention is elaborated into individual research goals presented in Chapter 2. The development of the main research instrument – a simple categorization test – and its validation via two pilot studies is described there.

The Main study is presented in Chapter 3. The respondents were introduced to a new concept in two ways, one group was given the definition followed by examples, and the other was presented with the examples first and the definition later. The validated categorization test and semi-structured interviews were used as research tools. The data was analysed both quantitatively and qualitatively.

The results are summarised and discussed in the context of existing research in Chapter 4. This final chapter also includes limitations and implications of my research.

Appendices include figures used in the Main study.

[1] Concept development refers to the phylogenetical development of a concept and concept formation refers to its ontological development.

Chapter 1

Theoretical background and literature

Literature does not usually define what a mathematical concept and a mathematical object are and uses these terms in their intuitive meaning. One exception is Freudenthal (1986) who analyses them in detail and in their mutual connection and investigates their relation to general epistemology and concrete mathematical theories. Some of the considerations below are based on his work.

1.1. Mathematical concepts and objects

The term *mathematical concept* is understood as a representative of an idea, mostly referring to some objects, categories, properties or definitions (see also Alcock, 2001, pp. 90–91). Moreover, two other views of mathematical concepts are useful for this thesis. In the first, a mathematical concept is accepted as a formalized part of mathematical theories in which this formalization is made by a *definition*. In the second, a mathematical concept is seen as a product of conceptualization – the phylogenetical process in which the idea was gradually shaped in the work of mathematician(s). The development of mathematical concepts in history is complicated; however, common patterns can be identified (Kvasz, 2008). All these views are considered in the following paragraphs.

The term *mathematical object* is understood in accordance with Sfard as a concrete instance of a mathematical concept.[2]

> Indeed, like physicists or biologists, the mathematicians use to talk about a certain universe, populated by certain objects. These objects have certain features and are subjected to certain processes governed by well defined laws. (Sfard, 1991, p. 3)

A common term also used in this context is *mental image* (Freudenthal, 1986; Fischbein, 1993). It is possible to say that mathematical objects are understood as rather independent of an individual, whereas mental image is a current representation of an object in an individual's mind. Thus, it is possible to say that mental image is a representation of an object in an individual's mind, which is the stance mostly taken in this thesis. However, a mathematical object can sometimes be semantically identified with its representation.

[2] There is a useful analogy to an object-oriented programming. Consider, for instance, the analogy of heredity, analogy between dynamic data types and representations of mathematical concepts, or analogy between properties and processes related to the mathematical concepts on the one hand and methods, parameters and properties of objects on the other.

There are various basic cognitive processes behind the representation of mathematical objects. Thinking about a mathematical object includes perception, processing and cognitive development of the representation of the object (Sternberg & Sternberg, 2012). For simplicity, it is assumed here that having an idea of something as an object means being able to create a mental image of this object. However, one has to be aware that the reality is more complicated.

1.2. Concept representation, concept formation and concept processing

Representation of knowledge in the human mind is described by cognitive psychology through theories and suitable cognitive models (Sternberg & Sternberg, 2012). For instance, it is commonly accepted to distinguish between declarative and procedural representation of knowledge.

> Declarative knowledge refers to facts that can be stated, such as the date of birth, (…) Procedural knowledge refers to knowledge of procedures that can be implemented. Examples are the steps involved in tying your shoelaces,… (Sternberg & Sternberg, 2012, p. 271)

In this perspective, representations of mathematical concepts and objects are commonly understood as a part of declarative knowledge. However, procedural knowledge also plays a substantial role in mathematics education, mostly in theories of formation of mathematical concepts (Sfard, 1991; Dubinsky, 1991; Hejný, 2012).

The common statement "a concept is represented in an individual's mind…" evokes an impression that concept representation is something static. Mathematical concepts are commonly perceived in this way – as static parts of mathematical theories whose definitions usually do not change. However, one has to be aware that their representations in an individual's mind are not static. They are dynamic and often change and develop in time. As noted in the Introduction, in order to emphasise these different views, concept representation, concept formation and concept processing will be distinguished and discussed in the following sections.

1.2.1. Concept representation

To describe a concept representation in an individual's mind, the distinction between *concept image* and *concept definition* (Vinner & Hershkowitz, 1980; Tall & Vinner, 1981) is useful.

Concept image and concept definition

The terms concept image and concept definition serve to distinguish two qualitatively different but interrelated types of concept representation. *Concept image* is understood as "the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes" (Tall & Vinner, 1981, p. 2)[3]. *Concept definition* is

[3] As is common in mathematics education theories, the meaning of the concept image and concept definition varies considerably across the literature. For instance, Moore (1994, p. 252) describes the concept image as "the cognitive structure in an individual's mind associated with the concept" and adds that "concept image refers to the set of all mental pictures that one associates with the concept, together with all the properties characterizing them". The adjudgment whether these two definitions are equal depends on the delimitation of the notion *mental picture,* but the latter formulation seems to pave the way for the restriction of the meaning of concept image as a set of examples and non-examples.

characterised as "a form of words used to specify that concept" (Tall & Vinner, 1981, p. 2). The concept definition is, thus, often used in two meanings. First, as a general product of historical development of a mathematical concept. Second, as a product of concept formation in the mind of an individual who considers it or uses it as a definition. When the distinction of these meanings is required, Tall and Vinner use the term *personal concept definition*. Similarly, the term *personal concept image* is used in the literature.

Following the above, a personal concept definition should be considered as a part of a personal concept image. However, authors mostly present them as disjoint parts of an individual's cognitive system due to their effort to present different kinds of processes behind these notions. For instance, Vinner (1983) shows several theoretical ways students proceed in a task. Their interpretation reflects the common teachers' experience – that sometimes students only use their concept image, concept definition, or concept image "consulting" a concept definition, etc. This is why the theory is highly useful for teachers – from the practical point of view, it makes good sense to distinguish a personal concept definition from the rest of the personal concept image.

When considering a personal concept image, a special role is played by *examples* of a concept.

Examples

The term *example* plays an indisputable role in mathematics and mathematics education but is characterised in various ways. In this thesis, the word *example* will be used in accordance with Watson and Mason (2002, p. 239) who specified an example as "anything used as a raw material for intuiting relationships and inductive reasoning: illustrations of concepts and principles; contexts that illustrate or motivate a particular topic in mathematics; and particular solutions where several are possible". The work of Watson and Mason (and other authors working within their framework) amply illustrates the meaning, boundaries and complexity of the term example and its uses (for more detail see Watson & Mason, 2006). In the text above, the phrases *for intuiting relationships* and *inductive reasoning* are important. The former represents the way examples help to understand properties of a defined concept which would be hard to grasp without them. The latter refers to generalization: an ability to extract information about a concept from an object(s).

The potential of the term example, but also formal paradoxes hidden in it, begins to be revealed when one starts to think about different types of examples.[4]

Typology of examples

Various types of examples have valuable functions in mathematics education. Michener (1978) describes *start-up examples* (examples suitable for the introduction of a concept), *reference examples* (examples which are used for the consolidation of some properties of concept), *model examples* (examples which characterize a concept well) and counterexamples ("examples that show a statement is not true", Michener, 1978, p. 367). With a similar meaning to *model examples*, *generic examples* are often understood as ones which allow for (and ideally initiate) generalization (Mason & Pimm, 1984). For instance, Zazkis and Leikin (2008) point out that $\sqrt{2}$ serves as a generic example of irrational numbers.

[4] Watson and Mason based their Theory of exemplification on this idea, exploring and examining the effect of students' working with examples. However, some of their conclusions have been questioned by some other researchers (Iannone et al., 2011).

The *prototype* or *prototypical example* has a similar meaning to a generic example but a rather different origin. The concept of prototype comes from theories of category representations and was introduced by Rosch (1973). Rosch showed that categories of objects are represented in an individual's mind by typical representatives of a category rather than a list of members (prototypes will be further discussed in Section 1.3).

Counterexamples can be considered as very important kind of examples, as many authors (such as Michener, 1978; MacHale, 1980; Mason & Pimm 1984; Goldenberg & Mason, 2008) show. Goldenberg and Mason, for instance, exemplify $|x|$ as "a counterexample to the conjecture that all continuous functions are also everywhere differentiable", etc. An important property of counterexamples is that they have their meaning only in relation to the concrete theorem – without the theorem, there is nothing to disprove and the counterexample only becomes an ordinary object. It is possible to see that the nature of counterexamples and, for instance, prototypical examples or start-up examples differs.

Finally, Fischbein (1993) introduces a term *figural concept*, pointing out the possible dual nature of some mental images one possesses. Using an example of an isosceles triangle, Fischbein shows that we are able to make various manipulations with mental images (object representations). Thus, when the nature of the original object is "sensorial", there may be general ideas behind the mental image. To stress that these mental entities have characteristics of concepts, the term figural concept was introduced for them. However, similarly to counterexamples, two different parts of human cognition are mixed in the term figural concept. First, a representation of an object in one's mind and second, access to this representation through particular processes.

Note: Consider terms with the opposite meaning to those named above which, intuitively, would have some non-standard properties – they would be examples which do not allow generalization, for instance. Consider number 2 as "definitely not a generic example" of primes. Moreover, it could also be easily labelled as both, a prototypical and an untypical example (for someone 2 is a good representative of primes even when it does not allow generalization).[5]

1.2.2. Concept formation

The term *concept formation* denotes the development of the concept representation in an individual's mind. Vinner (2002) uses the words "concept formation" for the process when the personal concept image is developed and points out the moment when "the image is formed, the definition becomes dispensable". He adds that it "will remain inactive or even be forgotten when handling statements about the concept in consideration" (p. 69).

Tall (2001) describes how the personal concept definition can allow for the development of personal concept image.

> The more sophisticated thinker notices properties of structures and relationships between them. Formal thinking begins when selected properties are isolated and used as concept definitions from which other properties may be deduced by mathematical proof. (Tall, 2001, p. 202)

[5] This discussion offers good potential: Consider, for instance, whether there is a difference between prototypes and generic examples on the one hand and non-generic and atypical examples on the other. Another example is that of a function which is not continuous. It is not easy to find a prototypical discontinuous function and if we find one, it will not be a prototype of discontinuous functions for everyone, etc.

Tall and Vinner (1981) use the term *conflicting factor* for the conflict between various parts of personal concept image.[6] They explore this phenomenon on the concepts of limits and continuity and present two important conclusions. First, the conflicts between different parts of personal concept image can cause a development of personal concept image (they explain under what circumstances it can happen). And second, even if conflicting parts of personal concept image are presented (and sometimes even evoked), it does not necessary lead to the learning process.

Similarly, Duffin and Simpson (1993) describe part of concept formation using the terms *natural, conflicting* and *alien learning experience* where each of these experiences gives rise to different responses.

Process-object paradigm
A common idea of concept formation is that we improve our concept representation when we process mental tasks somehow connected to the concept – solving problems, working with examples, proving claims, etc. This is captured by the so called process-object paradigm.

Several theories of concept formation follow the process-concept paradigm. For instance, in Dubinsky's APOS theory (Dubinsky, 1991; Dubinsky & McDonald, 2001), four phases are distinguished in the process of gaining mathematical knowledge: *actions*, *objects*, *processes*, *schemas*. Similarly, Sfard's Theory of Reification (Sfard, 1989; Sfard, 1991) is built on the idea that concept formation proceeds from the processual nature to the structural nature through three stages of *interiorization*, *condensation* and *reification*. Similarly, in his theory of generic models, Hejný (2012) describes the development of mathematical concepts in an individual's mind as a process from *isolated models* through *generic models* to *abstract knowledge.*

Many special cases which make the basis of a mathematical concept have a processual rather than conceptual character. A typical example is that of function. In schools, function is often introduced as "a dependence of something on something", but there is also a conceptual perspective. For instance, when particular categories of functions or their properties are presented, the representation of a function as an object mostly prevails. It is very hard to swap between these two perspectives and use one for the purposes of the other.

1.2.3. Concept processing

When discussing ways mathematicians deal with mathematical problems, Poincaré (1952, p. 120) writes:

> Many children are incapable of becoming mathematicians who must none the less be taught mathematics; and mathematicians themselves are not all cast in the same mould. We have only to read their works to distinguish among them two kinds of minds – logicians like Wierstrass, for instance, and intuitionists like Riemann. There is the same difference among our students. Some prefer to treat their problems "by analysis," as they say, others "by geometry".

What Poincaré saw from the perspective of mathematicians, Vinner and Tall see from the students' perspective when they use a framework of concept image and concept definition.

[6] Alongside personal concept image and personal concept definition, they introduce other special terms, for instance, 'personal concept definition image' and discuss conflicts between them. However, the main point of their article is the existence of conflicts themselves, their presentation on the concept of continuity and discussion of how to deal with them in an educational process.

Vinner (1983) examines a concept of function and describes various ways students are processing their decisions when solving mathematical tasks. For instance, one might expect that a student's solution will be formulated based on consulting both concept image and concept definition. However, Vinner also discusses the possibility that a student only uses his/her personal concept image for the solution, ignoring formal properties of the concept required by the definition. The opposite case, when one disregards his/her image and processes the solution based on the personal concept definition, is also possible.

Theories of a similar nature which distinguish among various approaches to a decision can also be found in psychology. Dual process theories represent the case. For instance, Kahneman (2011) distinguishes between two systems representing different cognitive processes. On the one hand, System 1 represents fast, unconscious and intuitive processes. On the other hand, System 2 represents rather slower and conscious processes characterized by controlled reasoning.

It is not possible to simply associate processes based on concept image to System 1 and processes based on concept definition to System 2, in the same way that it is not possible to say that Riemann thought through his concept image and Weierstrass through his concept definition. However, all the ideas presented show that concrete mental processes can be approached differently and that it is relevant to examine the nature of these approaches.

1.2.4. Two approaches to categorization

From the above, it is clear that it makes sense to distinguish between two different kinds of mathematical concept representations. According to the literature, these two kinds of concept representations are reflected in an in-the-moment approach to the concept (e.g., while solving a mathematical problem) from the subjective-intuitive (Poincaré), educational (Tall & Vinner) and cognitive (Kahneman) perspectives.

It is therefore reasonable to assume that the concept representation can be built in these two ways. In the context of concept formation, mathematics education literature focuses on the process-concept paradigm and conflicts in the cognitive system that initiate a shift in concept formation. Thus, the question arises whether differences in concept representation and concept processing could be observed if the concept is formed in the two ways.

As this work is concerned with students' concept images, the question arises as to what means we have to investigate them. Students' categorization seems to offer one suitable means of investigation as it provides some insight into their concept images by being relatively simple to observe in the students' behaviour. Besides observing the accuracy of such categorization, we can gather more information about the decision process by measuring reaction times.

1.2.5. Reaction time studies

Mathematics education studies using reaction times are scarce. For example, Vamvakoussi et al. (2012) gave the respondents a task to decide if a statement about rational numbers is true or false. The authors observed if the results differed when the statements fit the intuitive image or not; specifically, that respondents are influenced by their knowledge of natural numbers when comparing rational numbers represented by fractions. A similar approach was implied in the experiment of Babai et al. (2006) where the authors were focused on the comparison of areas and perimeters of two plain objects, with the same main conclusion.

Studies based on measurement of reaction time have two possible implications for this thesis – methodological and didactical-mathematical. Ashby et al. (1994) allow respondents to categorize simple plain objects into two set categories. Each respondent was tested multiple times, where each session consisted of 300 figures. Reaction times of respondents varied from 0.5 to 3 seconds. The authors interpreted their results within frameworks of several theories of categorization and showed that prototypical objects did not influence the reaction time. Such findings are, however, probably dependent on the context and methodology of the experiment because some other authors (Rips et al., 1973; Rosch, 1999) showed that when respondents were asked to verify if a figure is an element of a category, the fastest positive answers belonged to prototypes.

Many studies focus on the length of processing visual information. They examine the reaction times of human subjects, deciding whether something is or is not a member of a category. VanRullen and Thorpe (2001) found the reaction times to be mostly between 150ms and 600ms (median was slightly above 350ms; a figure was shown for 20ms). Another study with a similar methodology (Thorpe et al., 1996) found reaction times mostly between 200ms and 1 000ms (median was 445ms). Thus, only the reaction times lesser than these values should be considered as theoretically problematic for physiological reasons.

Ratcliff (1993) focuses on long reaction time data outliers and describes that what can be an outlier in one context does not necessarily have to be an outlier in another context. A common way of dealing with outliers is to trim reaction times greater than the mean reaction time of the group plus three standard deviations (as used in Vamvakoussi et al., 2012) but other authors (Leys et al., 2013) highlight disadvantages of this approach and recommend the use of median and absolute deviation.

Baayen and Milin (2010) discuss several approaches to data trimming; concretely how to deal with outlying data. The authors caution against trimming data unnecessarily. They also discuss different approaches to data transformation based on the reaction time distribution. In accordance with those recommendations, I have chosen to do no data trimming if there are no serious reasons to do so (for further discussion on concrete experiment methodology see Sections 2.6.3 and 3.4).

1.3. Categories

1.3.1. Classical categories

The term *category* is mostly understood as a collection of objects and is often used in relation to human understanding and work with categories of objects. How categories are represented in a person's mind is one of the questions of cognitive psychology and is connected to other topics of this domain such as perception, structure of memory, etc. (Sternberg & Sternberg, 2012).

Classical understanding of a category is based on the idea that to be a member of a category means to have some common properties with the other members of the category (Lakoff, 1987). Nevertheless, a scientific enquiry into category representation showed that this is not the only way of dealing with categories. Historically, various models of category representation and its use were developed, some of which are described in this section.

Mathematical concepts with their exact definitions match the above concept of categories precisely. Mathematicians are trained to think about categories in the same way. Thus, we cannot be surprised by a teacher's tendency to also think in this way when introducing the concepts of school mathematics. However, premature formalization of working with categories can be a source of various complications in the educational process. Thus, the problem may be in the teacher's inadequate work with categories rather than in the meaning of the mathematical concept.

From this perspective, substantial differences in work with categories can be seen in all grades of school. For instance, Alcock and Simpson (2002) argue that one of the reasons students' transition to university mathematics is difficult lies in the way categories are dealt with.

> School mathematics primarily involves calculations performed upon specific mathematical objects. For example, students are required to integrate a specific function or solve a specific differential equation. Even the few proofs encountered at this level (in the U.K. context) have this property: students are asked to prove by introduction that this formula gives the sum of the first n terms of this series or prove that this trigonometric identity is equivalent to that one.
>
> Proof at university goes beyond this. Work with specific objects is still required: students are asked to find the limit of a given sequence or to find the rational number that is represented by a given infinite decimal. However, they must now also work with entire categories of objects. (Alcock & Simpson, 2002, p. 29)

The idea that the accent moves more to the work with entire categories rather than concrete objects in higher grades of education is important. It shows that dealing with categories (and the way of changing our approach) forms an implicit part of formal mathematics education.

1.3.2. Properties of classical categories

Two great mathematical ideas are reflected in the concept of a classical category. The first is the development of the concept of set, which was explored by Cantor in the second half of the 19th century and first formalized in the ZFC axiomatic set.[7] The second is the process of logical formalization and axiomatization of mathematical theories as it lends precision to a definition.

The concept of set allows us to regard a definition as a construct which establishes a category of mathematical objects with appropriate properties. Nevertheless, without doubt, the classical view of categories was available before Cantor[8] – the concept of set was used informally and intuitively. However, the formalization of the concept of set allowed for a better organization of mathematical concepts and objects (Freudenthal, 1986) and precision of their perception. For instance, Frege (living at the turn of the 19th and 20th centuries) distinguished functional relationships from other mathematical objects. However, from the perspective of set theory, both functions and, for instance, real numbers are interpretable as sets and examined as such.

The analogy between a category and a set is useful when we try to formulate basic features of a category in the classical meaning. However, even if a property seems to be clear from the mathematical point of view, its educational interpretation does not have to be.

[7] Zermelo–Fraenkel set theory with the axiom of choice included.
[8] In fact, the origin of this thought can be found in Aristotle and his text Categories.

- Categories consist of individual objects

To understand an object as a member of a category, one must recognize it as an individual object.[9] For instance, basic binary operations such as addition, subtraction, multiplication and division are not often understood or perceived as objects (unlike the symbols representing them). Students often understand them in a processual way rather than as objects.

The formation of mental images of concrete objects behind the definition also requires refraining from external object representations. For instance, a precise mental image of a quadratic function represented by a graph of a parabola requires a realization of implicit information (that the graph only visualizes a part of the whole) and idealization of some properties of the represented object (that the curve has no width). The basics of this idea are provided by the psychology of perception (Sternberg & Sternberg, 2012).

- Categories are determined by a definition

A definition represents the list of necessary and sufficient conditions for an object to become an element of a category. A student should be aware that it is always possible to decide about the membership of an object using the definition.

Moreover, as definitions have a stipulative character (Vinner, 2002), the same necessarily applies to the corresponding categories – we define only those concepts and deal only with those categories for which we have a reason to do so. This will also be discussed in the following section.

- Categories are sharp

It is possible to decide about any mathematical object whether it satisfies the definition or not, i.e., whether the object satisfies the conditions established by the definition.

- Categories are not internally ordered

There is no inner structure among members of a category. Individual members of a category are thought of without any additional context and thus, no one is more 'important' than the other.

The classical model is as useful as it can be 'dangerous' for a mathematics teacher. On the one hand, it represents a kind of a clear and precise view of mathematical concepts and appropriate objects. The experienced mathematics teacher is (more or less intuitively) aware that the definition forms a category with the properties discussed even if the imagination or visualization fails, which can give rise to a feeling of confidence. On the other hand, a teacher can easily succumb to the impression that to understand a concept in the way described is simple and therefore easy or trivial for students.

Research on categories in the second half of the 20th century showed that they have many potential mental representations, depending on their nature, development, an individual's experience, etc. Thus, to think about a category representation in its classical form would be an oversimplification of the real state.

[9] An idea mentioned by Freudenthal when he notes that the constitution of mental objects precedes concept attainment (Freudenthal, 1986, p. 33).

1.3.3. Representations of categories

Wittgenstein

Wittgenstein's discussion of the concept of a game can be considered the first formal discussion of a non-classical category representation. Wittgenstein (1958) described several ways in which this concept differs from a classical category. *Family resemblance* expresses the situation when for every common property of two (or more) members of a category, another element(s) can be found for which the property absents. In other words, it is not possible to identify a set of properties which would characterise all elements of a category. Accepting family resemblance, one could say that a category of games can be delimited as a "logical sum of a corresponding set of sub-concepts" (Wittgenstein, 1958, p. 32), for instance, *children's games*, *sports game*s, *competition games,* etc.

However, Wittgenstein adds that there are categories which do not fit even this delimitation. First, it is not possible to set the precise boundaries of a category which would fit all the perspectives – it is not possible to decide exactly what still satisfies the game concept and what does not: "We do not know the boundaries because none have been drawn. To repeat, we can draw a boundary – for a special purpose." (ibid., p. 33) As he states, it is possible to delimit the boundary of a category with a purpose; however, this boundary does not have an exact character. Rather, it has an artificial nature. Second, Wittgenstein also shows that category boundaries are not supposed to be static and can change over time. The extensibility of category boundaries by the development of the concept is thus possible. A good example is provided by Lakoff who extended the category of games by videogames (Lakoff, 1987, p. 16–17).

Fuzzy sets

The introduction of fuzzy sets (Zadeh, 1965) provides a potential mathematical model of the blur of category boundaries and a measurement of the centrality of category members. However, in order for the category to be well interpreted as a fuzzy set, it has to have a parameter of category membership which can be measured well. Categories of "rich people" or "tall men" (Lakoff, 1987) are good examples of those – the criterion of category membership is relatively clear and possible to interpret as a characteristic function of fuzzy set membership. This model represents significant progress compared to the classical model. However, category representation in an individual's mind appears to be more diverse.

Prototype theory

An important contribution to the domain of category representation arrived with Prototype Theory (Rosch, 1973). Rosch provides an interpretative framework of categories, some of whose members are more typical than others for an individual. Such members are called *prototypes*. The task "decide whether an object is a member of a category" is then processed mostly by comparison with the prototypes of a category.

In mathematics education, the idea of prototypes was first introduced by Hershkowitz (1989) when she showed that a student's category representations of at least some mathematical concepts correspond to the prototype theory and that these representations can influence his/her performance in related mathematical tasks. Hershkowitz examined geometrical objects and showed that figures similar to triangle prototypes but not fitting the definition were more often considered to be members of the category of triangles than atypical triangles that fit the definition.

Since the 1980s, prototype theory has been proven to be a valid and useful framework for category representations of mathematical concepts. Moreover, the word *prototype* has been taken on as a common term in mathematics education. Thus, it is not possible to discuss the role of prototypes in mathematics education literature in its entirety. The following paragraphs elaborate on two studies in the context of prototypes with a connection to the topic of the thesis.

Alcock and Simpson (2017) examined an interference between defining/explaining and classifying in the case of sequences. Before being asked to classify sequences, one group of students was asked to explain a concept of an increasing sequence, the second was asked to define it and the third was given the exact definition. It was found that the first two groups classified the sequences better than the third group.

In their other work, Alcock and Simpson (2002) compared different ways of prototype construction and appropriate category development in both everyday and technical contexts. In their experiment, respondents reasoned about the relationship between two categories – bounded and convergent sequences. The authors described three kinds of respondents' reasoning in tasks from the domain of sequences: *generalising*, *property abstraction* and *working from definitions*. Using the first kind of reasoning, a respondent generated a conjecture from the prototype and generalized it for the whole category. Using the second, a respondent abstracted a property from the prototype and reasons about the category using this property. Using the third, a respondent reasoned based on the definition. While the first two are based on dealing with categories in everyday contexts, the third corresponds to dealing with classical categories.

Other category representations

Since the introduction of the concept of prototype, many authors have worked on the development of the theory. A detailed overview is provided by Kruschke (2005), who classifies theories of categorization according to three parameters. First, Krushke distinguishes whether the theory describes a category as content-specified or boundary-specified giving the following example. The category of 'skyscrapers' can be delimited by its boundary (let us say that the ratio of height to width is greater than 1.62) or by its content (if the building is more similar to one than other). Second, the author distinguishes three types of category representations – whether they are specified by a *global summary* or by *piecemeal components*: "the two descriptions of skyscrapers given above were global summaries, insofar as a single condition defined the boundary or the content of each category" (Kruschke, 2005, p. 185). Thus, the category represented by piecemeal components is represented by more than one condition. Third, Kruschke distinguishes whether the categorization process is *strict* or based on *graded similarity*. This distinction encourages interpretation using sets and fuzzy-sets respectively.

To name a concrete theory, Shin and Nosofsky (1992) developed a scaling definition of similarity assuming that each object can be represented as a point in multidimensional space and similarity between objects is represented as their distance.[10] The above and similar experiments led to the formulation of RULEX Theory (from Rule-Plus-Exception, Nosofsky, Palmeri, & McKinley, 1994) which, simplified, is based on "a model which searches, in order, for simple rules, then more complex logical rules, and finally for rules plus stored exceptions that successfully discriminate the category members" (Kurtz, 2015). The RULEX theory can

[10] This idea is interesting from the mathematical perspective when one considers various types of metrics for the measurement of the distance.

be applied to some rather complex concepts in mathematics. Consider, for instance, the concept of function where the representation of a category of functions can be made by continuous functions plus some special cases, such as Dirichlet's function (Tall & Vinner, 1981).

1.4. Mathematical concepts and definitions

The definitions of mathematical concepts have been formulated at least since Euclid's *Elements*. The development and formulations of definitions have changed and evolved in line with the development of concepts themselves. However, the discussion of modern definitions, their basic properties and nature is bound to the beginnings of formal logic. Since one realizes that the development of modern set theory and formal logic went together, being investigated by the same great mathematicians, it is no coincidence that we can date the beginnings of the classical interpretation of cognitive categories and formal definitions to the same period. From then up to today it is possible to find many ideas in connection with the development of mathematical concepts and their definitions. Some of these ideas will be presented in concrete examples, in a form beneficial to the topic of the thesis, for it is these examples that show what role definitions play (or should play) in mathematics education.

Poincaré (1952) considered the role of definitions in learning and teaching when critically assessing the balance between a formal logical and intuitive approach to mathematical concepts. The phrase "it is by logic that we prove, but by intuition that we discover" (Poincaré, 1952, p. 129) expresses Poincaré's perspective that both views of mathematics should be cultivated because both can lead (and indeed led in history) to great discoveries. The work of Lakatos (1976) describes one of these cases.

Example 1: Concept of a polyhedron as a product of a mixture of different ways of thinking
Lakatos describes a complicated historical development of the concept of a polyhedron. He ascribes the beginning of this concept formation to Euler's polyhedron formula[11] and continues to the various definitions of a polyhedron. Among others, he describes a number of methods mathematicians use in their work. An important one is the use of counterexamples to disprove a theorem (some nonconvex polyhedron as a counterexample for Euler's formula applied on general polyhedra).

Kvasz (2008) revisits strategies described by Lakatos, distinguishing two approaches in dealing with a theorem disproved by a counterexample. The first lies in ignoring counterexamples as "monsters" (do not think about a nonconvex polyhedron as a polyhedron). The second consists of restricting the theorem to ensure validity (restrict Euler's formula only on convex polyhedra). Both cases are problematic. In the first, the theorem does not apply to all the objects considered. In the second, the theorem is possibly restricted too much (Euler's formula also holds for some nonconvex polyhedra). Kvasz concludes that both these approaches are needed when looking for a theorem as general as possible.

[11] The number of vertices (v), edges (e) and faces (f) of a polyhedron satisfies the formula $v + f - e = 2$. The statement was first mentioned by Euler in 1758. Euler's formula is satisfied for convex polyhedra. The number on the right of the formula is called Euler's characteristic and various nonconvex polyhedra have various Euler's characteristic (including 2).

1.4.1. Historical development of mathematical concepts

The previous example leads us directly to the history of the development of mathematical concepts. Mathematicians working in this domain are commonly focused on the description of historical development of concrete concepts,[12] domains of mathematics,[13] or on general patterns between mathematical theories in different contexts.

Kvasz's (2008) work belongs among the latter. He describes the development of mathematical concepts in the linguistic perspective by interpreting "changes in mathematics as changes of the *language* of mathematics" (Kvasz, 2008, p. 7). The author distinguishes three kinds of significant changes in the development of mathematics: *re-coding*, *relativisation* and *re-formulation.* Re-coding represents changes in symbolism and relativisation represents a change in relation between the symbols and objects they express.

> A triangle in Euclid, in projective geometry, in Lobachevski, in Klein, or in Poincaré looks the same; it is constructed following the same rules. Nevertheless, in each of these cases it is something rather different because it has different properties and different propositions can be proven about it. (Kvasz, 2008, p. 9)

The third change, called re-formulation, describes how the basic primitives (axioms) and principles are formulated in accordance with the construction of the theory (shown on the different kinds of axiomatizations of Euclidean geometry). Among others, Kvasz points out the complexity of the development of some concepts.

Example 2: Complexity of the conceptualization – equations

For instance, Kvasz shows a complexity of the development of an equation concept when describing individual stages of the historical development of the language of algebra since al-Khwárizmí. Kvasz summarizes that solving an equation requires several steps:

- to find a rule written in an ordinary language, which makes it possible to calculate the root of the equation,
- to find an expression of the symbolic language, which makes it possible to express the root in terms of its coefficients,
- to find a factorization of the polynomial form,
- to reduce the given problem to an auxiliary problem of lesser degree, etc. (Kvasz, 2008, p. 199, shortened)

Kvasz's analysis of the concept of equation brings three important perspectives to this thesis. First, at the beginning of mathematical concepts, there is usually a problem; the development of a concept is often a consequence of its solution. Second, the manipulation with the concept changes over time and the changes are not only formal and marginal, but also fundamental. Thirdly, the development of mathematical concepts can hardly be considered to be finished.

The fact that the development of a mathematical concept is a complex and longitudinal process which can evoke progress of seemingly unrelated parts of mathematics is well known. Example 2 also shows that even when there are time periods when a concept seems to be well known and

[12] For instance, the concept of continuity (Nunez & Lakoff, 1998) or the concept of function (Kleiner, 1989). It is noteworthy to mention the work of Piaget and his colleges (Piaget et al., 1977) who also provide a detailed description of the history of the function concept.

[13] For instance, mathematical analysis (Edwards, 1979).

understood, new discoveries or representations may lead to the revision of the seemingly stable concept and to the formulation of new definitions within new mathematical theories. Without doubt, it is complicated for a teacher to follow this developmental perspective in school mathematics. The following example shows a typical simplification of the historical development for educational purposes.

Example 3: Misleading motivation for the introduction of complex numbers
A usual source of motivation for the introduction of complex numbers is the discussion about the existence of roots of a general quadratic equation $x^2 = ax + b$, where x is the unknown and a and b parameters. This equation can be interpreted as finding a line which intersects a parabola (with the vertex in the origin). Three cases can happen: a line will intersect the parabola in two points, one point or there is no common point. Needham (1998) points out that this is entirely in agreement with the algebraic solution which yields an appropriate number of roots in the real numbers and shows that the true motivation for considering other numbers comes with a special set of cubic equations. Consider, for instance, the cubic $x^3 = 15x + 4$. Geometrically interpreted, we investigate whether the cubic $y = x^3$ intersects the line $y = = 15x + 4$. At least one such intersection obviously exists. However, Cardano's formula (introduced in the 16th century) leads to the solution $\sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$. Needham notes that this was the paradox which allowed Bombelli to develop basic operations with complex numbers.[14]

However, if any motivation is presented at all when complex numbers are introduced to secondary school students, it is probably based on the former motivation (no matter if it is based on geometric or algebraic representations). This is not the only case when a historical development of a concept can be simplified in school mathematics (whether it is beneficial or not). But it is another example of how we deal with mathematical concepts, interpreting them in the most intuitive form and trying to mediate this interpretation for students. Nevertheless, one has to be aware that it is not necessarily the most effective way, which moves us back to the domain of mathematics education.

1.4.2. Definitions from the point of view of mathematics education

In contrast to non-technical scientific domains, definitions of mathematical concepts are usually understood as stipulative (in contrast to lexical) (Brown, 1998; Alcock & Simpson, 2017). Thus, to define a concept means to use a new or existing term in a concrete context (a mathematical theory) – the term becomes meaningful in this context. However, it does not mean that a term cannot have its meaning outside the theory – consider, for instance, the concept of continuity.

Example 4: Concept of continuity
The concept of continuity is essential for students' understanding of the concepts of calculus. Tall and Vinner (1981, p. 17) examined concept images of novice university students and described various conflicts between these images and the formal definition. Their conclusions are important (especially in the framework of concept image and concept definition) but not surprising. Students tend to have an intuitive and informal personal concept image of continuity before they actually meet the concept of function.

[14] For further information and full explanation of this motivation see Needham (1998, p. 4–5) and Stillwell (2010).

Thus, when introducing the continuity concept, a teacher often both explains a complicated calculus concept and expresses how to formalize it in a closed logical system. The attribution of the meaning to the words ‘to be continuous’ is problematic in the context of functions.

Example 5: Explanatory power of the definition of function
Example 4 represents a situation in which a mathematical definition is associated with an intuitive, real life concept. In contrast, the concept of function changes its meaning several times during primary and lower and upper secondary schools. It begins with intuitive definitions such as ‘the dependence of one quantity on another’, continues with emphasizing the existence of only one element from the domain for each element of the range and mostly ends with the definition based on a Cartesian product of two sets.

It is useful to mention the study of Vinner and Dreyfus (1989) who compiled a questionnaire where respondents decided whether presented graphs represented a function or not, found functions with some special properties and defined the concept of function itself. The authors identified various kinds of definitions provided by the respondents and showed that those respondents who formulated a “modern” definition of function[15] often did not use it when dealing with other tasks.

From my perspective, an important problem in dealing with mathematical concepts is observable in the case of function concept. Students are taught definitions (for instance, a function as a subset of Cartesian product of two sets) whose explanatory power is never fully utilized (or is utilized only for a few basic, artificially created problems). Thus, the definition loses its meaning and students’ attention understandably shifts to concrete representatives of the concept and their intuitive properties.

The examples described in this chapter show that the development, meaning, and use of mathematical concepts and their definitions can often be non-intuitive (Example 1), in contrast with their historical development (Examples 2 and 3), in contrast with many students’ images of them[16] (Example 4) and their potential sometimes remains unfulfilled (Example 5). The emphasis on the exactness of a mathematical definition can thus be easily overestimated and the intuitive component of its understanding can be underestimated.

Moreover, it is clear that, in a practical teaching context, roles and forms of mathematical concepts and their definitions change for different school grades. In the lower grades, mathematical concepts are presented in a similar way to any ordinary everyday concepts; an unprecise concept definition is accepted. However, in the higher grades of education, the exactness of definition improves. This also shows that both intuitive and exact approaches to work with definitions necessarily intermingle in the mind of an individual.

1.5. Summary

Let us summarize the considerations presented in this chapter which provides the theoretical background of my research.

[15] The authors use the term Dirichlet-Bourbaki definition described as „correspondence between two nonempty sets that assigns to every element in the first set (the domain) exactly one element in the second set (the codomain)” (p. 357).
[16] But they can also be in contrast with teachers’ images.

Section 1.2 was devoted to cognition connected to mathematical concepts. Concept image and concept definition were discussed as two cognitive representations of mathematical concepts in an individual's mind, which differ meaningfully. In the concept image, a substantial role belongs to examples and non-examples,[17] mathematical objects which allow one to induce information and serve usefully as a source of properties of an appropriate concept. Different kinds of examples were presented and classified, based on their nature: cognitive and educational.

Section 1.3 provided an overview of cognitive representations and processes related to categories. A classical category was described as a representative of a logical structure and of the compactness of mathematical concepts. This, however, does not reflect properly how it is represented and used in an individual's mind. Cognitive psychology provides plenty of models of category representations which often fit mathematical concepts too, as shown in some experiments in the domain of mathematics education. These two perspectives provide two distinctive views of the mathematical concepts.

In Section 1.4, the phylogenesis and ontogenesis of mathematical concepts and their definitions was taken into account. Several examples were provided to show that in school mathematics, mathematical concepts and their definitions are often not used in accordance with their historical development or theoretical exactness.

Based on the above, the following principles were determined which underlie my research in the practical chapters of the thesis:

- Mathematical concepts as defined notions are products of a longitudinal phylogenetic development. The development and formalization of some mathematical concepts are not separated but rather progress simultaneously. The formalization of some concepts can initiate and facilitate the formalization of other concepts.
- One of the possible models of concept representation is the distinction between concept image and concept definition. Concept definition is understood as a part of concept image. However, they are distinguishable both in terms of static representations of a concept and of the current student's approach to the concept (for instance, when solving a task).
- The process-object paradigm is often used to describe concept formation in didactical theories. This paradigm also represents an imaginary bridge between concept formation and concept processing.
- A notion of categories in a classical way is understood as a product of the sequential formalization of mathematics. The interpretation of a mathematical concept as a category of mathematical objects was significantly extended by the logical formalization of mathematics and introduction of set theory.
- School mathematics is often seen in a structural perspective as a world of defined objects with relations between them which are exact and based on the rules of logic.
- The classical category and its members have several properties corresponding to the concept of a set: it consists of individual objects, it is determined by its definition, it has sharp borders and it has no internal structure.

[17] The term non-example is understood in the thesis as an object which does not fit the appropriate definition while an example does.

- The interpretation of and thinking about many mathematical concepts as a classical category creates an implicit part of the present curricula and is also implicitly considered as a substantial part of understanding the concept.
- Teachers are able to interpret most mathematical concepts from the point of view of classical categories. However, this interpretation may not be easy for students.
- A teacher's work with formal definitions and appropriate categories does not often fit the logical and historical development of mathematical concepts.
- The representation of mathematical categories in the minds of students is not primarily classical. There are various category representation models which, more or less, fit various kinds of concepts.

In addition to the individual observations described above, several times I have encountered a potential dual approach to dealing with mathematical objects in various contexts. Following this idea, it is reasonable to presume that this theoretical dual approach will also be reflected in the categorization of mathematical objects. In the context of these observations, I consider it relevant to ask the following question, which is a starting point for the practical chapters of this thesis:

> Can an individual's approach to a categorization task reveal information about his/her concept image and concept definition?

Chapter 2

Research design and research instruments

2.1. Research goals

In the theoretical chapter above, the reasons for investigating students' categorization of mathematical objects were given. The simple categorization (the decision whether or not an object is a member of a category in question) represents a basic task in this domain. The above theoretical analysis implies several open questions in the field of mathematics education. The intention to observe whether respondents use different cognitive approaches in making their decision is one of these questions, and serves as a starting point for the practical part of this thesis.

The research design can be divided into several steps.

Step 1: Design a method that allows us to explore if and under which conditions it is possible to observe two different approaches to making a decision during simple categorization. These two approaches are delineated by the terms concept image and concept definition.

Step 2: Validate this method, showing how it is used and discussing the possibilities and limitations of its use.

Step 3: Identify the respondents, contexts or concepts (or concrete figures), where one approach to a particular decision prevails.

Step 3 can be investigated using different methods. One would be to choose a sample of respondents, process the simple categorization test with an arbitrarily chosen concept, and then observe possible differences among respondents. However, this would mean that we would not get any information about the respondent's process of acquiring the concept. This prompted two methodological decisions for the planned Main study. First, it was decided to use a new concept developed for the purposes of the study, which opened up the possibility to investigate the above process (learning from definition vs. learning from examples and non-examples).

Second, it was decided that two different ways of concept formation would be developed. The first follows an idea of exact definition, appropriate respondent's personal concept image and a theoretically exact approach to categorization. The second follows an idea of intuitive concept formation based on the distinction of examples and non-examples, and a theoretically intuitive approach to categorization.

Step 4 involves describing possible relationships between the results of the third step and the quality of personal concept image and personal concept definition of a respondent. If possible, the aim is to establish the designed method as a way of measuring the 'quality' of the personal concept image and personal concept definition of the respondent.

I will start with the description of how the categorization test was designed.

2.2. Simple categorization test design

As mentioned above, the simple categorization test is a construct developed with the purpose of gathering data about respondents' reactions when categorizing objects. Similar tests are used in psychology. Alongside investigating the content of the respondents' response, their reaction time[18] is measured as it provides additional information about their reasoning during the categorization (Ashby et al., 1994; VanRullen & Thorpe, 2001). Two types of content were used in the tests in my study.

2.2.1. Structure of the test

The categorization test was designed in an online application Testable.com which was developed as a methodological tool for recording manipulation with objects on a computer screen and for measuring the reaction time of this manipulation. Raw data can be downloaded from the database as a .csv file. Data includes the time from when the image was displayed to the respondent's reaction, and whether the answer is right or wrong.

Identification and concentration activity

At the above webpage, the respondents identified themselves by an ID number. Afterwards, they were instructed that they would first be participating in a training activity: the letters "A" or "N" (for Czech words "ano" and "ne" meaning "yes" and "no", respectively) would be presented. The task was to press a corresponding key on the keyboard. A random letter was then presented five times in a row with no time break. This task was included to focus the respondents on the test.

Note: The concentration activity was included in the research design only after Pilot study 1, following the identification of the first example effect (see Section 2.5.3). It is included here for the sake of completeness of the research design.

Categorization of figures

After the previously described activity, instructions appeared on the computer screen that 20 figures would be presented to the respondents who were asked to press "A" if the graph represented an object of the category in question or "N" if not. The test consisted of ten examples and ten non-examples; however, respondents were not aware of this ratio. Their order was random for each respondent. No upper time limit was used for the respondent's reaction. There was also no delay between pressing the key and the appearance of the next figure.

2.2.2. Content of the categorization test

Various existing concepts were considered for the purpose of the study. The concepts defined by properties of functions at the secondary school level were considered for the following reasons. First, since secondary school, students are familiar with the representation of functions as both graphs and algebraic expressions. Second, concepts based on the properties of functions are usually easily visualizable and, thus, easily observable from examples and non-examples –

[18] Reaction time and response time are commonly taken to be the same. The first term prevails in the thesis.

there is a chance that respondents learning through examples and non-examples will develop an adequate concept image.

On the other hand, respondents can potentially have preconcepts of some concepts (consider, for instance, the concept of continuity as described in Section 1.4.2) or partial knowledge about concepts from previous education. This led me to the idea of a concept developed solely for the purposes of the study.

Therefore, two versions of the test, related to two different mathematical concepts were used in the pilot studies: the concept of a Tall function (in original "vysoká funkce") which was developed for the purposes of the study and the concept of an injective function.[19] Whereas the first concept will be discussed in the following sections, the second is commonly understood, and thus is represented only by examples and non-examples used in the test (see Appendix B).

Concept of a Tall function and its place in the study

For the purposes of a simple categorization test, the concept of a Tall function was developed. The concept refers to an intuitive property which is similar to some global properties of functions (for instance, an injection of a function or distinction between a bounded and unbounded function). Its definition, examples and non-examples and potential connected issues will be discussed in the following paragraphs

Definition of a Tall function

The definition of a Tall function was formulated in this form:

> Every function whose maximum is greater than the absolute value of its minimum will be called a Tall function.

Discussion of the definition and potential objections

A possible objection due to the definition in the presented form is that the case of non-existence of maximum and/or minimum is not included. It is implicit that in the case that the maximum or minimum of the function (or both) does not exist, the definition is not fulfilled and an appropriate function is not a Tall function.

All examples related to the concept of a Tall function as used in the simple categorization test are presented (ordered from e1 to e10) in Fig. 1, and all non-examples are presented in Fig. 2 (n1 to n10). The same figures were used in each test (in both pilot studies and the Main study); nevertheless, they were always presented in a random order.

As can be seen from the graphs, they have some properties which could play a role when one is deciding whether a graph represents a Tall function or not. I will discuss them in the following paragraphs.

[19] The function f is said to be injective provided that for all x and y from the domain, whenever $f(x) = f(y)$, then $x = y$.

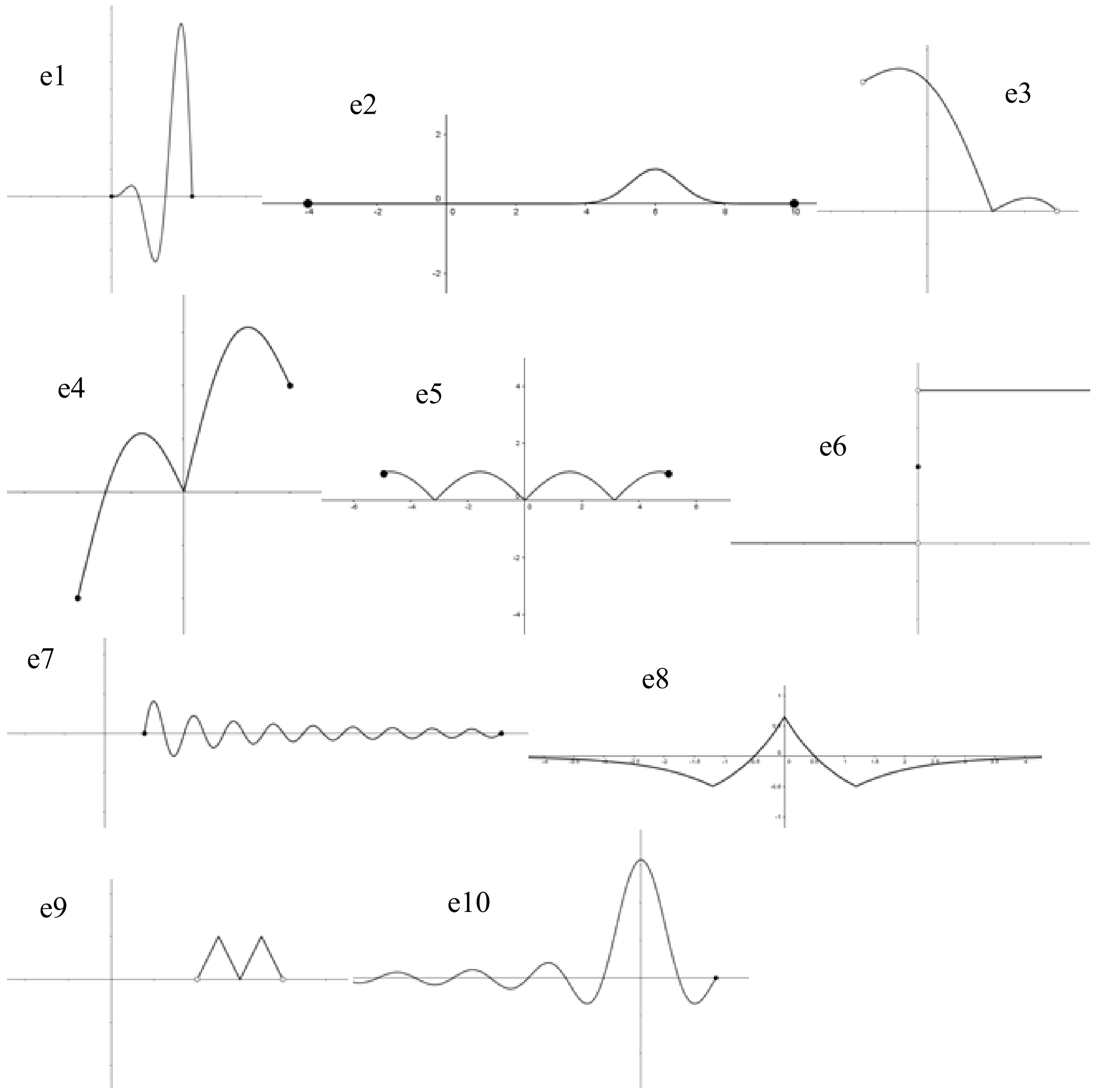


Fig. 1: Examples (e1–e10) of a Tall function used in a simple categorization test

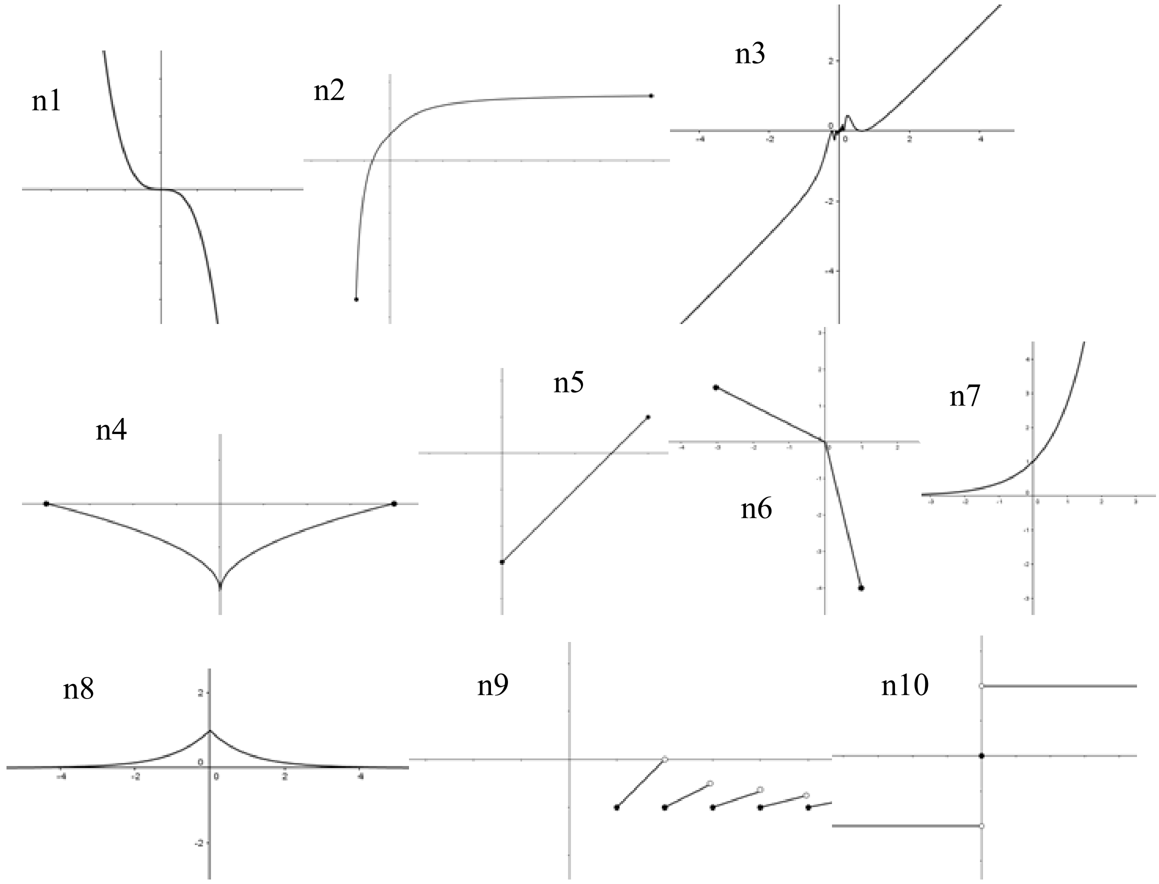


Fig. 2: Non-examples (n1–n10) of a Tall function used in the simple categorization test

Existence of maxima/minima
The existence of maxima and minima is taken as a necessary condition for a function to be considered as a Tall function. However, when the maximum or minimum does not exist, there are two possible causes. First, a function is unbounded from above (or from below). This is the case in figures n1, n3 and n7. Second, a function is bounded from above or below by a concrete value, but never reaches it. This is the case in n7, n8 and n9. The case of n9 differs from the others since a respondent has to be aware of the meaning of filled and empty points which represent inclusion or exclusion of the point to the graph of the function.

In some cases, a potential misunderstanding is not connected to the property of a function but to its representation, in this case a graph.

Approaching/non-approaching an axis:
Consider figures e8, n7 and n8, where it is not clear if the function reaches zero for x going to infinity or minus infinity. These graphs were included intentionally because it is meaningful to observe how respondents deal with such ambiguous situations.

Comparison of values:
A similar issue arises with figures e7 and e8, which represent functions where the comparison of maximum and minimum may be ambiguous. It is then on the respondent how he/she will decide without additional information (e7) or if he/she will check values on y-axis (e8).

Visualisation of part of a graph

In the case of figure n9, it is not clear how the graph continues for increasing x depicted outside the picture. Thus, it is important to realize in what ways the respondents can interpret the continuation of the graph out of the picture.

2.2.3. Limitations of the categorization test

All the issues described above provide potential problems which have to be solved somehow (or simply ignored) when a respondent categorizes figures and, thus, can influence the results of experiments, representing potential limitations of the test.

It has been noted that even if it is possible to clearly decide whether a function belongs to a category, each respondent can perceive the graph as a representation of a different function. In the ordinary education process, this aspect is compensated for by a *context* and *convention* (a teacher specifies the concept image shaping in students' minds, highlights important features, diminishes irrelevant ones, etc.). However, in the experiments presented in this thesis there was no starting discussion, nor agreement about the understanding of graphs (deliberately because our pilot interviews and other experience showed that these discussions either increase confusion of respondents or differences between them, or both, see below). Thus, during the data analysis, awareness of whether differences between reactions can occur due to this effect is relevant and will be taken into account.

2.3. Technical remarks

The word *figure* in the following text primarily refers to graphs of functions (that is, either the example or the non-example). The abbreviation "Fig." is used only to refer to the pictures displayed in the text.

The term *personal concept image* is used in the sense of a respondent's total cognitive structure connected to the appropriate concept, in accordance with Tall and Vinner (1981). The word "personal" refers to a concrete respondent in the study.

The term *personal concept definition* is understood as a very specific part of the personal concept image which represents a definition of a corresponding concept. For further discussion of the relationship of these two terms see Section 1.2.1.

In the text below, the phrase "Respondent decided through (using) his/her personal concept definition" means that in my view, a respondent made a decision (conscious or unconscious) considering (some or all) aspects of his personal concept definition. Similarly, the phrase "Respondent decided through his personal concept image" means that the respondent decided rather intuitively, without consulting his personal concept definition. It does not necessarily mean that the respondent decided impulsively, unconsciously or without deeper consideration.

When it is said that the figure was categorized "correctly", this only means that the answer was in line with my categorization. There are some figures whose categorization can be questionable and dependent on additional information as discussed above.

2.4. Summary

With the research goals presented above and a developed research instrument, Pilot study 1 was designed to check the potential of the simple categorization test. It was necessary to observe behaviour of respondents when they were processing the test and identify potentially problematic issues. As described below, the test was shown to be usable for my research goals. Moreover, several critical issues were identified and ways to deal with them formulated. Pilot study 2 was designed with two main purposes. First, to explore possible ways to analyse data obtained in the simple categorization test. Second, to explore ways to avoid the influence of issues observed in the first Pilot study on the data obtained. And third, to pilot a simple categorization test in the situation it was designed for: a quantitative testing of categorization of mathematical objects.

2.5. Pilot study 1

The simple categorization test, as presented in Section 2.2, was developed to examine how a respondent deals with categories of mathematical objects. The goal of the first Pilot study was to observe the respondents' behaviour when processing the test and describe potentially problematic issues.

2.5.1. Methodology

The Pilot study was realized with 4 respondents who were experienced mathematics teachers. Respondents with rich experience in mathematics were chosen for two reasons. First, because of their ability to give good feedback on the concepts and figures in question. And second, because of their higher ability of metacognition – it was assumed that these respondents would talk about their reasoning about mathematical objects more easily than students

The Pilot study was realized in a silent room, only a respondent and the author were present. The experiment was designed as a semi-structured interview, with the simple categorization test included. The interview consisted of several parts.

- The concept of a Tall function was discussed including the most important properties of the concept as presented in Section 2.2.
- The simple categorization test was introduced, the process of testing was described and the task to decide whether the presented graph of a function is an example or non-example of a Tall function was introduced.
- The respondent was asked to pay attention to their cognition during processing of the test.
- The respondent completed the simple categorization test in the online environment with figures representing examples and non-examples of a Tall function.
- After the test, each figure in the test was discussed one by one in terms of any cognitive issues which the respondent might have experienced when doing the test.

The interviews lasted from one to three hours.

2.5.2. Data and their analysis

The data consisted of results of the test in terms of the respondents' decisions and reaction times and transcripts of interviews. The transcripts were analysed in terms of cognitive issues reported by the respondents and the data concerning decisions and reaction times were scrutinised in order to point to any problems.

2.5.3. Results and implications for future research

In the Pilot study, I identified some potentially problematic issues which might have influenced the respondents' decisions. In the following paragraphs, I present these issues and suggest ways of dealing with them in future research.

Blackout effect: An effect of unconscious delay of the respondent's reaction

The analysis of the interviews uncovered an interesting phenomenon. The long reaction time was not always caused by the respondent's inability to categorize a figure, but in their own words, they were "stuck". In other words, they did not ponder the figure, they were simply "thinking of nothing". This phenomenon will be called the "blackout effect" here. Unfortunately, within the methodology of the study, there is no possibility to identify and exclude the blackout effect from the collected data.[20] Thus, in the following studies, I presume that the blackout effect is distributed equally in my data and, therefore does not influence the data significantly.

Previous figure effect

In the first Pilot study, some respondents mentioned that the long delay in their reaction was not caused by thinking about the current figure but about the previous one. There are several possible causes of the "previous figure effect" but similarly to the blackout effect, there is no possibility of its identification without adding an additional cognitive load to the respondent. Thus, it will again be presumed that the previous figure effect is distributed in the data equally.

Effect of awareness of measuring reaction times

The respondents' awareness of the measurement of reaction times was identified as a very stressful factor by three participants. Therefore, I decided to observe this phenomenon more closely and include it as a variable into the second Pilot study.

Effect of the first figure(s)

When observing the respondents, it became obvious that the decision about at least the first (but probably a few first) figures is executed more slowly. First, I decided to include the observation of this effect in the second Pilot study. Second, because this effect has no simple cognitive nature, I decided to reduce its influence by randomizing the figures presented for each respondent and by developing a simple identification and concentration activity to make respondents more concentrated at the beginning of the test (see Section 2.2).

Explicit instructions and discussion of its negative effect

All the above issues represent potential limitations[21] of a simple categorization test. Thus, it would be useful to try to separate and measure the influence of these effects. Nevertheless, the more we draw respondents' attention to these problems, the more complicated the situation would be for them as they would have to allocate part of their cognitive resources to paying attention to them. Thus, it was decided that in future studies, no explicit information will be

[20] A potential solution might be that a respondent simply says he/she has just experienced such a blackout. However, this is not a good option. First, some blackouts might not be identified by the respondent. Second, the person's concentration on the identification of blackouts requires some cognitive resources and thus can influence the results even more.

[21] But not all the limitations – for a deeper discussion of limitations of the simple categorization test as transpired after the Main study, see Section 4.76.

provided about potentially problematic issues so as not to draw the participants' attention to them including:

- information about the number of figures presented,
- information about the ratio of examples and non-examples,
- information if the graph presented truly represents a graph of a function in all cases,
- information if there are any time-breaks between the presented figures.

2.6. Pilot study 2

The second Pilot study[22] was designed with three purposes in mind. First, to validate the concept of a Tall function as relevant for the Main study. Second, to investigate possibilities and limitations of the analysis and interpretation of the data gathered by measuring reaction times of the simple categorization. Third, based on the conclusions of Pilot study 1, to investigate the influence of 'Effect of awareness of measuring reaction times' and the 'Effect of the first figure(s)'.

2.6.1. Research questions

Based on the above goals and the findings of the first Pilot study, the following research questions were formulated.

P_RQ1. Which examples and non-examples fit the purposes of the Main study?
P_RQ2. How accessible is the concept of a Tall function to secondary school students?
P_RQ3. How do the reaction time data refer to the personal concept image and personal concept definition of the concept of a Tall function?
P_RQ4. What influence does the awareness of respondents of measuring reaction times have on their accuracy and reaction time in simple categorization?
P_RQ5. What influence does the order of presented figures have? Particularly, what difference is there between the first and the first few figures in comparison to the rest?

2.6.2. Research sample and research design

The research sample consisted of 22 students of the second grade of secondary school in Prague (16–17 years old) from one class. The study was conducted in June 2016, at the end of the school year in which the concept of function had been taught in detail. No students or gathered data were excluded from the study. The classroom where the study was conducted was equipped with computers.

The study consisted of three activities which were organised by myself.

Activity 1: The definition of a Tall function and development of appropriate concept image
The respondents were introduced to the main idea of the study, namely that they would be taught a new concept by definition and presented with several examples with explanation, and afterwards they would solve some tasks. The respondents were also instructed that they were not allowed to ask any questions regarding the concept in question or communicate with or explain the concept to anyone else.

[22] While I designed the study and collected the data, part of the statistical analysis was made in cooperation with Derek Pilous. These results were presented in (Pilous & Janda, 2017). Here, I will only present the results connected to the focus of the thesis and elaborate them.

Then, the respondents were collectively introduced to the concept of a Tall function by definition and several examples and non-examples of graphs of a Tall function. Next, some concepts connected to the graphical representation of functions was revised with them. The following questions were discussed and illustrated by prototypical graphs of functions:

“What is the Cartesian coordinate system and what is a graph of a function?”

“What is the meaning of filled or empty points?” The principle of inclusion or exclusion of a specific point was discussed in relation to Fig. 3a and Fig. 3b.

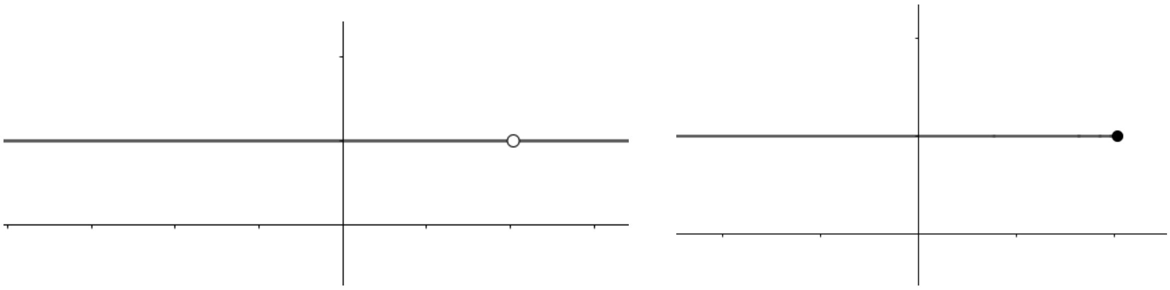


Fig. 3a, 3b: Graphs for the discussion of the meaning of filled/empty points

“What does it mean that ‘a function goes to the infinity’?” The explanation was based on the linear function $y = x$.

“What does it mean that ‘a function is approaching / approaching and never reaching concrete value’?” The difference was discussed in relation to graphs presented in Fig. 4a and 4b.

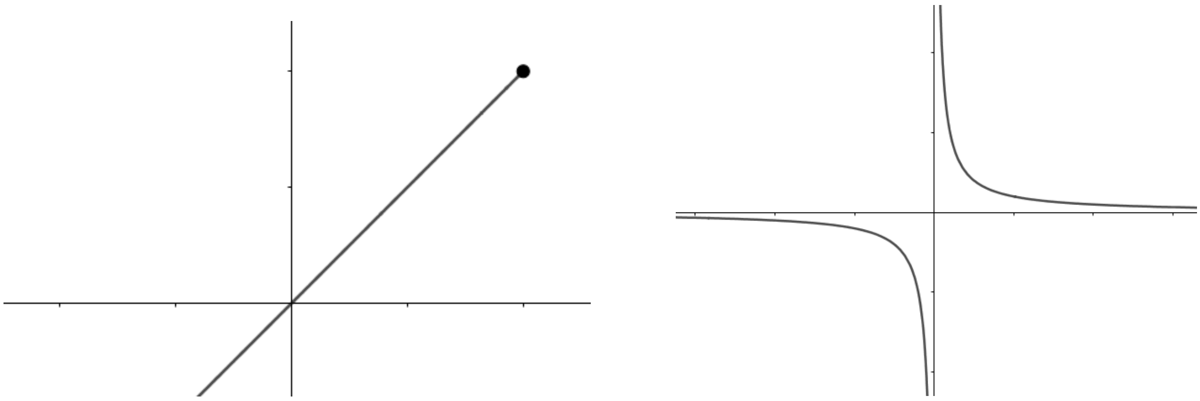


Fig. 4a, 4b: Graphs for the discussion of “approaching/approaching and never reaching a value”

“What is maximum and minimum?” Both definitions were explained in relation to graphs presented in Fig. 5a, 5b, 5c and 5d. The question of multiple maxima/minima was also discussed.

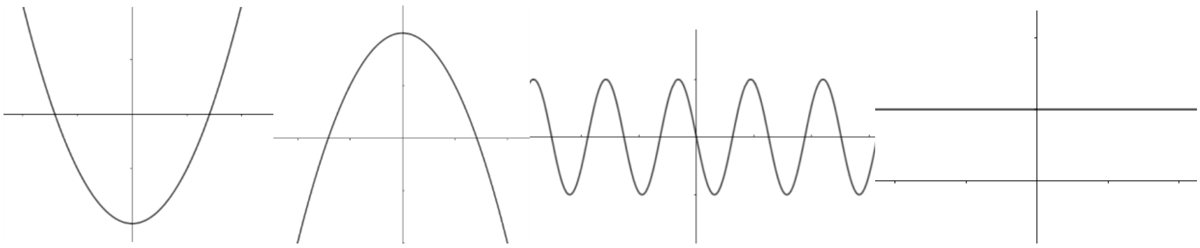


Fig. 5a, 5b, 5c, 5d: Graphs for the discussion of concept of maximum and minimum and of the possibility of more than one maximum or minimum

“Does a function always have a maximum and minimum?” Both cases, i.e., an unbounded function and a bounded function which never reaches the boundary value were discussed.

“What is an absolute value and an absolute value of a minimum?” The process of finding the absolute value of a minimum of a function was expressed in the graph of function $y = |x| - 2$ with the restricted domain from -3 to 3 (see Fig. 6).

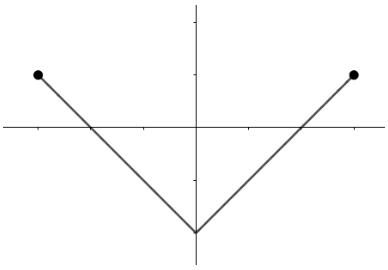


Fig. 6: Graph of function used for explaining a process of finding an absolute value of a minimum of a function

Next, critical issues of a Tall function concept described in detail in Section 2.2.2 were discussed with the students. After that, the respondents were instructed to participate in an online test using an internet browser.

Activity 2: Simple categorization test with respondents unaware of measuring reaction time
All participating respondents processed the categorization test as described in Section 2.2 under my supervision.

Activity 3: Simple categorization test with respondents aware of measuring reaction time
After the first simple categorization test, the respondents were informed that their reaction time was being recorded. Next, they repeated the whole process consisting of the identification and concentration activity and the simple categorization test as described in Section 2.2. The order of the figures in the first and second trial differed for the same respondent, as it was random.

2.6.3. Analysis of the data

The study was conducted with 22 respondents who reacted to 20 figures in each of the two tests, which gave us information about the accuracy and reaction time for each of 440 responses. Several methods were used for the data analysis.

Reaction time data outliers
In accordance with Baayen and Milin (2010), a trimming of outliers was considered and rejected because even long outliers can be a natural result of the decision process. If long outliers were trimmed, we could not be sure that only values representing spurious processes were excluded. From this perspective, the outliers could actually provide additional information about both respondents and figures used in the test.

As for the fastest reactions, their values (0.67s in the first trial, 0.41s in the second trial) can be seen as acceptable considering that values around 0.40s for a decision about categorization are possible, as discussed in Section 1.2.5.

Distribution of data
In accordance with the literature (Ratcliff, 1993; Baayen & Milin, 2010), reaction time data distribution is mostly considered as ex-gaussian, log-normal, inverse normal or Gamma. Thus,

one possible way to normalize the data is their logaritmization and use some statistical methods on the logaritmized values. Therefore, the log-normal distribution of reaction time data was presumed. The analysis of the second Pilot study data from this perspective and its discussion is beyond the scope of this text. It can be found in (Pilous & Janda, 2017).

Geometrical mean vs. median
When referring to the reaction time data, researchers mostly use the median of reaction times (Leys et al., 2013) for the analysis. Because the distribution of reaction time data is close to log-normal, the use of a geometrical mean is also possible. In this thesis, I primarily use the geometrical mean and in some cases, I present both values for comparison. When statistical tests comparing the mean values are used (independent sample t-test or paired sample t-test), it is done on geometrical means of reaction times because of the assumed log-normal distribution of the data.

Normalization of the data due to a respondent's general ability to react
It is natural that the approach of each respondent differs. For instance, he/she decides (consciously or unconsciously) how much effort they will invest to be fast and how much to be accurate. In other cases, some respondents can prioritize a faster reaction when the others want to be as accurate as possible. Thus, we must be aware of these effects on the data. Where necessary, I decided to objectivize the relative reaction time by ranking reaction times of particular figures of a particular respondent from the slowest to the fastest.[23]

Statistical tests and significance
As discussed above, a paired sample t-test demands normal distribution of data. To achieve this, geometric means of reaction times of a particular respondent were used. However, with such a small sample size, results of statistical tests have to be interpreted carefully. Thus, they were used only in a few cases in the second Pilot study to support the presented findings rather than to draw main conclusions.

2.6.4. Results

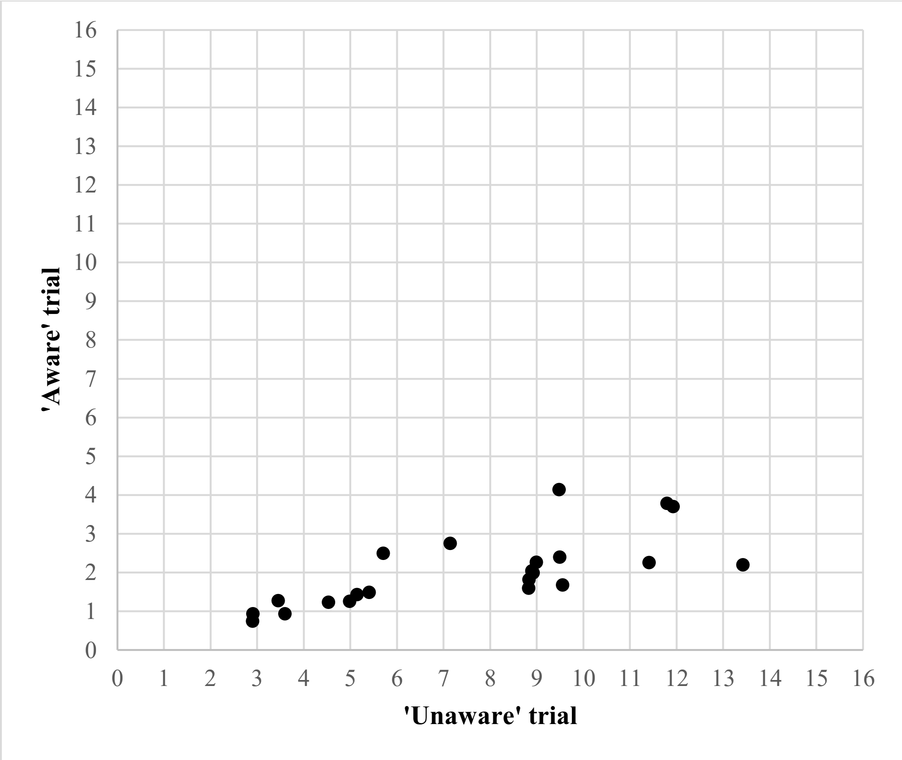
Effect of awareness of measuring reaction time
The results of the first trial (in which the respondents were unaware of reaction time measurement, 'unaware' trial) and the second trial (in which they were aware of this measurement, 'aware' trial) differed in several ways. The reaction times in the first trial were in the range from 0.67s to 79.14s, the reaction times of the second trial were in the range from 0.41s to 14.08s. The geometrical mean of all the reaction times in the first trial was 6.89s and in the second trial it was 1.83s. Even though there might be a learning effect in play, the large difference in times suggests that the respondents probably tried to be as quick as possible, as expected. This is noteworthy as they had been informed that the reaction time was being measured; however, they had not been explicitly instructed to be as fast as possible.

The difference in reaction times between the first and second trial is obvious[24] and was expected in accordance with the results of the first Pilot study. Nevertheless, some noteworthy observations can be made when looking closer at the results. As can be seen from Graph 1, all

---

[23] This approach was used in the Main study (see Section 3.8) as it was considered suitable due to the follow-up connection to the qualitative results.
[24] This conclusion was also proved statistically. The difference in reaction times between the trials was significant on the level of 99% with p-value $1.6 \cdot 10^{-9}$ ($t = 10.096$) using paired sample t-test.

the respondents reacted faster in the second task than in the first (if there was no change in reaction times, the dots would be on the diagonal). In addition, those respondents who tended to be fast when unaware that reaction time was being measured also tended to be fast when aware of it. A cautious conclusion can be made that under both conditions, some respondents are generally faster than others. This gives further credit to the theoretical normalization due to a respondent's general reaction ability which is possible to access, for instance, by division of measured reaction times by the geometrical mean of all reaction times of the appropriate respondent or by using the ranking of reaction times described above.



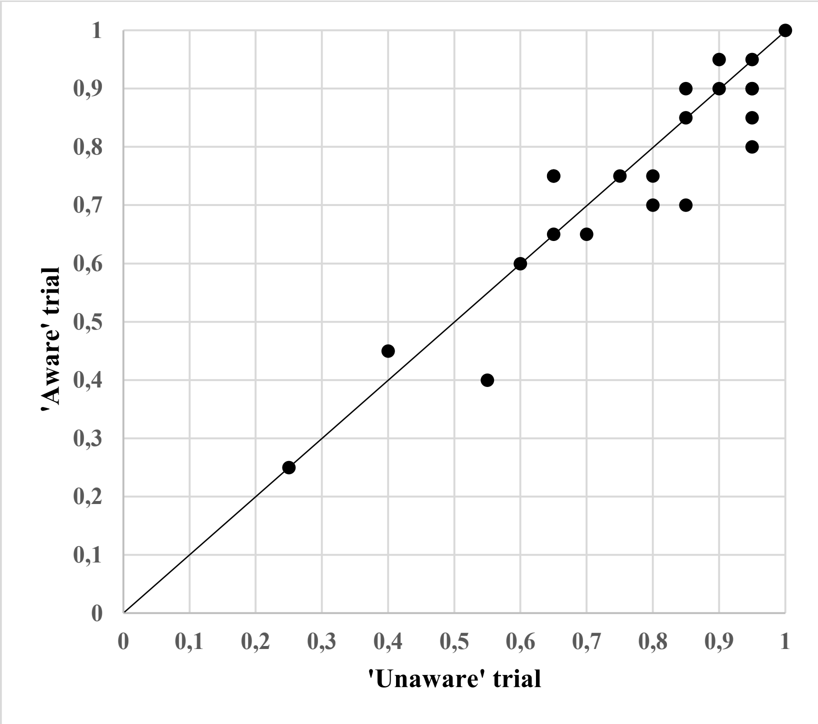
Graph 1: Comparison of reaction times of respondents between 'unaware' and 'aware' trials

The accuracy of respondents and the learning effect

The information on the accuracy in the categorization task reflects the respondent's personal concept image of a Tall function and is therefore related to P_RQ2 and P_RQ3 in Section 2.6.1. The accuracy did not vary much between both tasks. The average accuracy was 76.8% in the first trial and 74.5% in the second.[25]

The comparison of accuracies in the first and second trial is presented in Graph 2. The accuracy was actually slightly worse in the second trial, most likely due to the awareness of reaction time measurement. This suggests that the effect of the awareness of reaction time measurement is stronger than the fact that the respondents completed the same test twice (with a different order of figures).

[25] As described in Section 2.22, the task consists of true/false questions and as such, the accuracy of a randomly deciding hypothetical respondent is 50%.



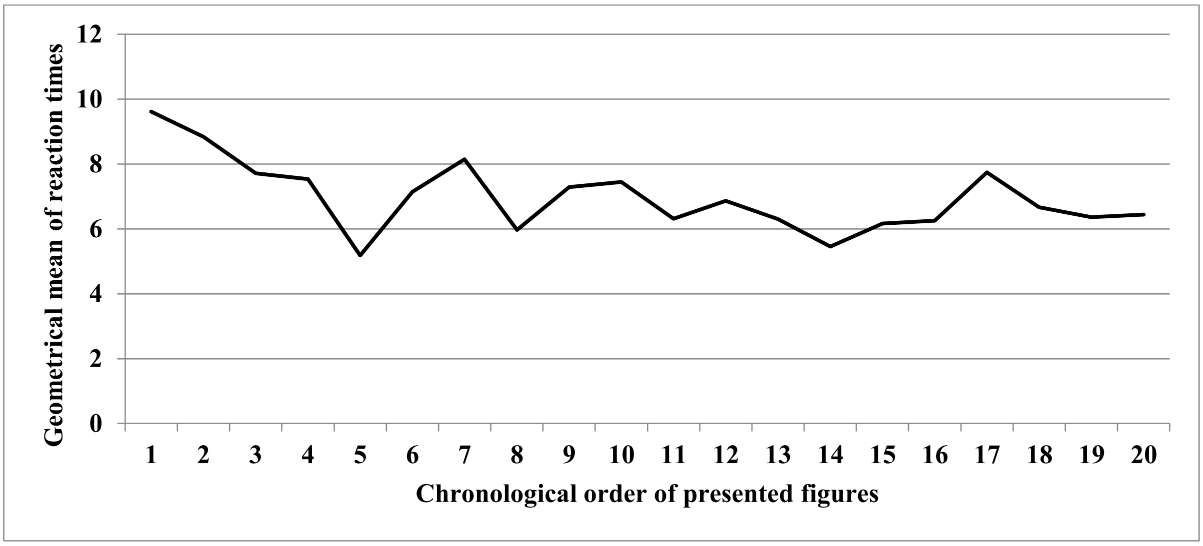
Graph 2: Comparison of average accuracy of respondents between ‘unaware’ and ‘aware’ trials

It can also be seen from Graph 2 that four respondents improved their accuracy (their results are above the diagonal) and the accuracy of eight respondents remained the same (on the diagonal). It is important to clarify here that this does not mean that their answers were necessarily the same in both trials. In fact, the casewise change of a concrete respondent’s decisions (categorizing a concrete figure once correctly and once incorrectly) was relatively high – the ratio of this cross-trial change was 11.8%. Only 2 respondents from 22 did not change their decision for any of the figures. Thus, the concept of a Tall function was shown not to be stable in the minds of most respondents who changed their decisions in a relatively high number of cases.

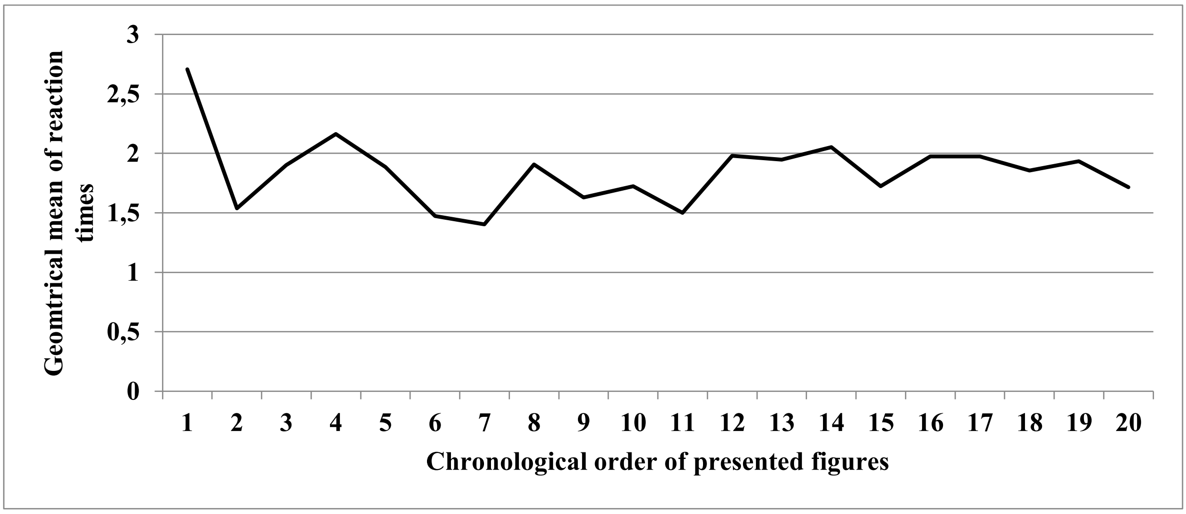
A relatively high casewise change provides two possible means of interpretation. First, the respondents profoundly changed their personal concept image of a Tall function which led to various changes in their decision (they re-learned concrete critical aspects of graphs or realized particular aspects of the definition). However, this interpretation is improbable as the respondents spent dramatically less time on their decision making in the second trial and no feedback was provided to them on their decisions during the study. Moreover, the average accuracy in the second trial hardly changed; thus, it would be a “change” of the respondents’ personal concept image rather than its “improvement”. The second interpretation posits that the respondents reacted in the second trial as if they saw the figures for the first time and their previous experience did not seem to influence their decision much. However, this conclusion is speculative.

Effect of the first figure(s)

Next, a possible influence of the order in which figures were presented will be considered. In both trials, the first figure presented was identified as the slowest on average. In the first trial, the geometrical mean of reaction times of the first figure was 9.61s while the geometrical mean of all reaction times was 6.89s. In the second trial, the geometrical mean of reaction time of the first figure was 2.70s and the geometrical mean of all reaction times was 1.83s. Such a difference can hardly be random. An overview of the geometrical means of reaction times of all the figures in chronological order[26] can be seen in Graphs 3 and 4.



Graph 3: Comparison of the geometrical means of presented figures in chronological order in the first trial



Graph 4: Comparison of geometrical means of presented figures in geometrical order in the second trial

Difference between examples and non-examples

Even though it was not originally planned, another interesting observation was made when comparing reaction times of examples and non-examples. The times for the latter were significantly shorter (Table 1). This difference was statistically significant (on the confidence level of 95% with $p = 9.63 \cdot 10^{-4}, t = 3.939$ in Trial 1 and with $p = 1.26 \cdot 10^{-3}, t = 3.819$ in the second trial). Such a large difference can be explained by a different complexity of the decision when the respondents must consider whether the figure is an example or a non-example of a Tall function. To show that the figure is a non-example of the relevant category, it is

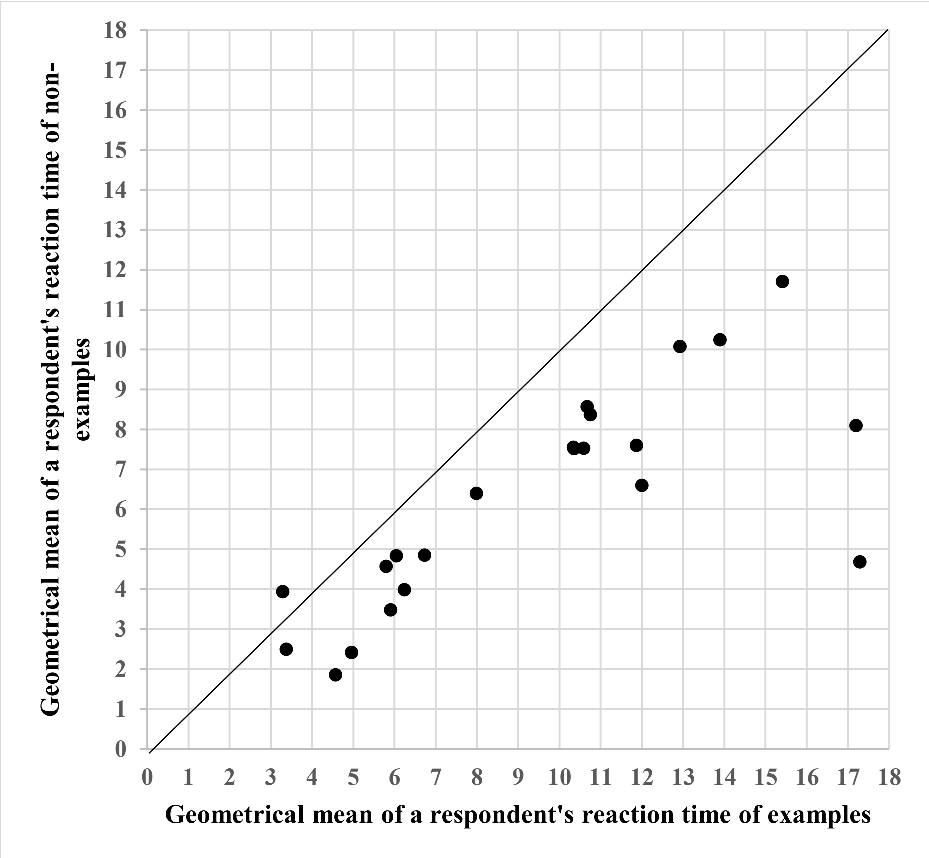
[26] The geometrical mean was calculated from the reaction times of all the figures presented as the first, the second, etc. figure in the sequence.

sufficient to find one aspect of the graph showing that the figure does not belong to the category. On the other hand, to prove that the figure is an example can be challenging.

Table 1: The difference between reaction times measured for categorization of examples and non-examples

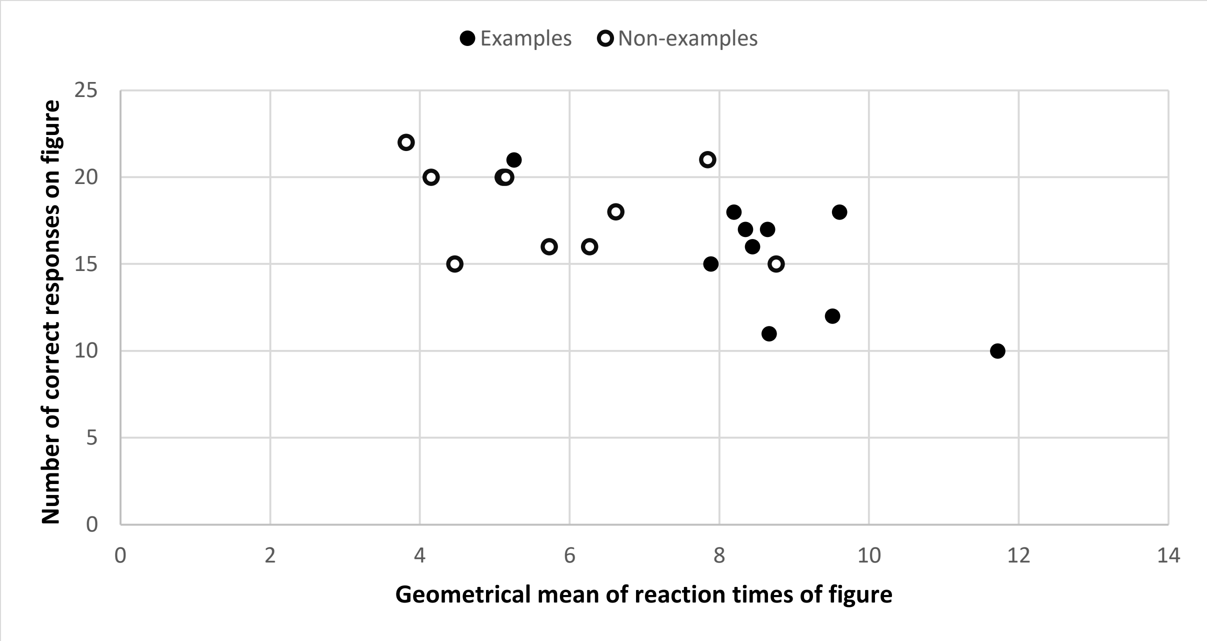
	1st trial			2nd trial		
	min RT	max RT	geometrical mean	min RT	max RT	geometrical mean
Examples	0.67	79.14	8.48	0.50	14.08	2.14
non-examples	0.80	32.53	5.6	0.41	11.59	1.56

The above results are supported by Graph 5 which shows a comparison of differences between reaction times of decisions about examples and non-examples of concrete respondents. Only one respondent decided about examples faster than non-examples when measured by the geometrical mean.



Graph 5: Difference between geometric means of reaction time of examples and non-examples of concrete respondents

Graph 6 presents other results concerning examples and non-examples. In accordance with the previous results, there is a noticeable difference between sets of examples and non-examples in terms of reaction times (the horizontal axis). Moreover, the difference in accuracy can also be seen (the vertical axis). Non-examples seem to be 'easier' to decide about and this naturally leads to shorter reaction times.



Graph 6: The relationship between the geometrical mean of reaction times of figures and the number of correct answers in the first trial

It appears that if a respondent has a well-developed personal concept image of a concept, there is a chance that the difference in reaction times for examples and for non-examples will be significant. If he/she has an undeveloped personal concept image, this difference can be relatively small.

The correlation[27] between the difference of the mean reaction times[28] and the accuracy of the respondent in the first trial was found to be 0.39 and thus, this observation is worth exploring in the Main study.

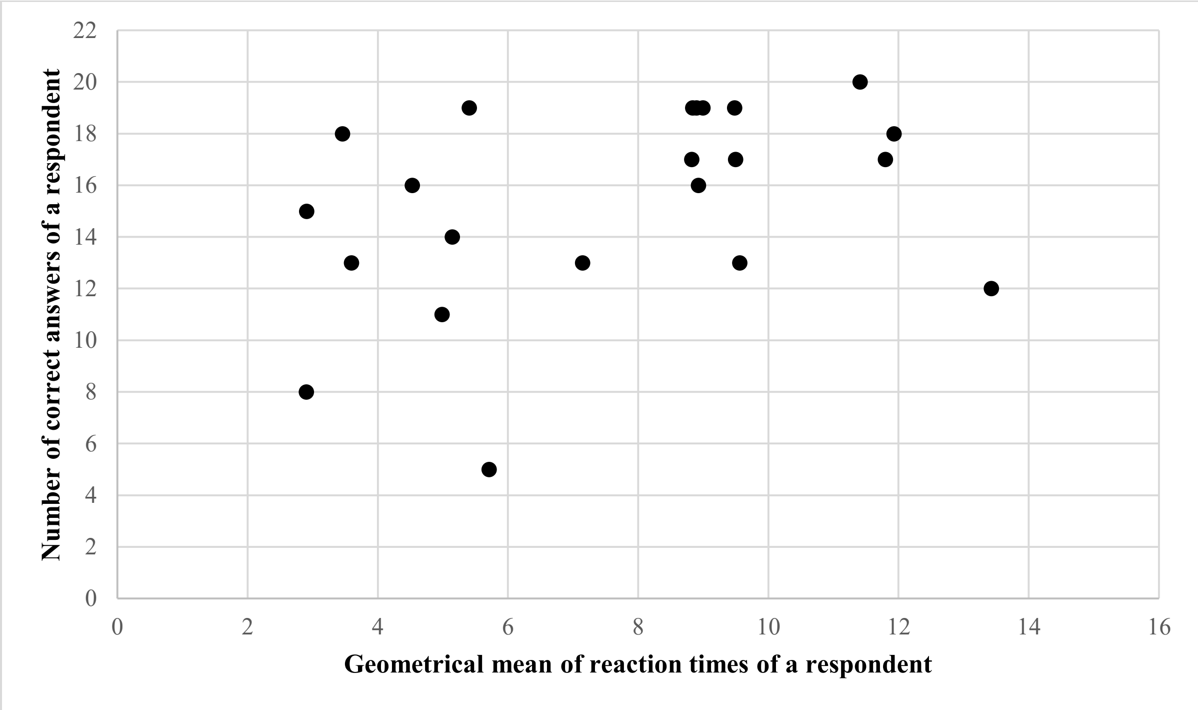
Approach to the decision from the respondent's perspective

The relationship between the respondent's reaction time and the number of correct answers in the first trial is presented in Graph 7.

As can be seen, the respondents with a rather long average reaction time tend to have a high number of correct answers. Moreover, it is possible to split the respondents into two groups based on the geometrical mean of their reaction times (alongside the mean reaction time between 7s and 8s). A tentative hypothesis can be made that these two groups of respondents differ in their approach to the task. These results are promising with respect to the main goals of the thesis. However, more data which would bring new information about a respondent's decision process is needed to gain an insight into the nature of this difference.

[27] Pearson's correlation coefficient was used taking into account the fact that geometrical means should be distributed approximately normally, see also (Pilous & Janda, 2017).

[28] The difference was counted by the expression 'geometrical mean of reaction time of example − geometrical mean of reaction time of non-example' for one respondent.



Graph 7: The relationship between the geometrical mean of reaction times for figures and the number of correct answers in the first trial

Results for concrete figures
As can be seen from Graph 6, there are three non-examples with a very short average reaction time and one example with a very long average reaction time. These figures are presented in Fig. 7. The quantitative data does not provide us with an explanation as to what caused such short or long reaction times for these concrete figures. Interviews with respondents would be needed to get an insight into particular aspects which they considered in their decision process.

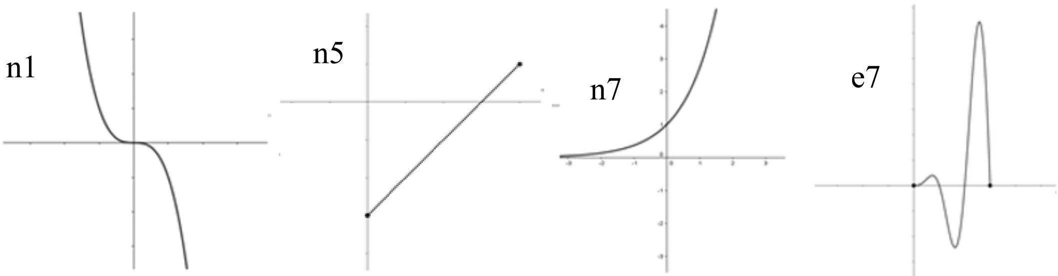


Fig. 7: Non-examples with very short reactions times (n1, n5, n7) and an example with a very long reaction time (e7) measured by geometrical mean

2.6.5. Conclusions and implications for the Main study

Several differences between the ‘unaware’ trial and the ‘aware’ trial were found.

First, the difference in reaction times itself was very high even though the respondents were only told that the reaction time would be measured but they were not explicitly instructed to be as fast as possible. Thus, no such information should be given to respondents in the Main study (P_RQ4).

Second, a relatively high number of changed decisions between both trials (figures once categorized correctly and once incorrectly by a concrete respondent) and almost unchanged accuracies between both trials show that the respondents' personal concept image was somehow changed but not improved. In this perspective, the results were probably influenced by significantly faster reactions in the 'aware' trial.

Third, a relatively high effect of the first and second figures was identified. Reaction times for the first figure (and for the second figure in the second trial) were longer compared with all other figures. Therefore, the discussion of the first example effect is a necessary part of any methodological framework similar to the one used in this study (P_RQ5).

Fourth, significant differences of reaction times were identified between examples and non-examples. An observation was made concerning the difference between the process of decision making for the figure which is an example of a Tall function or which is its non-example. Moreover, I presume that this difference is observable through the reaction times. To test this, a bigger sample is needed.

Last but not least, none of the figures included in the test was identified as inappropriate or problematic (P_RQ1). The difficulty of the task was shown to be adequate for the observation of personal concept image of a Tall function for secondary school students (P_RQ2). Thus, no changes were made in the content of the test in the Main study. However, from the results obtained it was clear that the task planned for the Main study where respondents should learn about the concept on their own and only by one of the ways (the definition or examples and non-examples) would be too difficult for respondents with mathematical knowledge at the secondary school level. Therefore, respondents with greater mathematical experience should be considered for such an approach.

Finally, some results of the Pilot study suggested a need for a qualitative approach to the description of respondents' decision about concrete figures (P_RQ3). Thus, the Main study will include a qualitative part, too.

From perspective of the research instrument discussed in Section 2.2, the simple categorization test was to be compliant with the requirements and problematical parts of its use described. Thus, the methodology of the Main study was planned, consisting of two groups where the concept of a Tall function will be introduced in a different way to respondents from another group. During the analysis of data obtained in Pilot study 2, I realized that for the identification of patterns between the results of the reaction time test and concept representation in a respondent's mind, it would be necessary to focus on the decision about concrete, critical figures in order to describe well the respondent's decision making process. Therefore, an interview with the respondent after each reaction time test was planned.

Chapter 3

Main study

The two pilot studies presented in Chapter 2 validated the simple categorization reaction time test (see Section 2.2) as an eligible research tool to reach the goals described in Section 2.1, revealed the way in which the test should be applied and what its limitation might be. The knowledge gained in the pilot studies was used to design the Main study.

3.1. Research questions

Step 3 in the research design (described in Section 2.1) is conceptualized by two research questions:

RQ1. Is it possible to distinguish two pathways of decision making in a categorization test by measuring the reaction times and accuracy?

RQ2. If so, are there any respondents, contexts or concepts (or concrete examples) where one of the decision-making paths prevails?

The third research question was formulated partly in regard to Step 4:

RQ3. Which differences are possible to observe when the personal concept image of a Tall function is developed through examples and non-examples on the one side, and through the definition on the other?

To be able to generalize some results gained with a Tall function to other concepts, an additional test was included. The concept of an injective function, known to students from the secondary school, was chosen as it is semiotically close to a Tall function concept, and as such, it might provide results worth comparing. Following these intentions, two more research questions were added:

RQ4. Which phenomena identified in the simple categorization test based on a Tall function will be identified in the simple categorization test based on an injective function, too?

RQ5. What kind of relationship exists between the reaction time of respondents in the test focused on the concept of a Tall function and an injective function?

Finally, the results of the second Pilot study led to an interesting observation describing possible differences between examples and non-examples. Thus, the last research question is:

RQ6. Which differences between examples and non-examples can be described for both concepts used (a Tall function, an injective function) based on the simple categorization test data?

3.2. Respondents

The results of the second Pilot study influenced the selection of the research sample. Respondents with presumably better mathematical knowledge were needed as the respondents' personal concept image in the Pilot study 2 was not satisfactory in terms of accuracy despite the fact that the concept of a Tall function was presented to them by both definition and several examples and non-examples, including those with problematic aspects. As the concept of a Tall function was to be presented in only one way (either through the definition or examples and non-examples) in the Main study and the respondents were to construct the personal concept image on their own (without my assistance), the task became even more difficult. Thus, more mathematically able respondents were selected: 21 first-year students (six male and 15 female volunteers), future mathematics teachers, at one of the universities in the Czech Republic.

First, the study design was trialed with three respondents and the experience gained informed the rest of the Main study. It was planned to split the remaining 18 respondents randomly into two groups, referred to as a Definition group (Group D) and an Image group (Group I). However, after the sessions with the first ten respondents, the necessity to gather more qualitative information about Group I respondents was noted. While Group D respondents reacted consistently in the interviews, the reactions of Group I respondents varied. Thus, it was decided that the study would be processed with seven respondents following the procedure of the Definition group and with eleven respondents following the procedure of the Image group. Both procedures are described in detail below. Each of 18 respondents is referred to by a three-digit number.

The experiment was realized in December 2016 and January 2017, at the end of the term, when basic courses of higher mathematics[29] at the university had been completed by the students.

3.3. Research design

The study was designed as a sequence of five activities whose order differed for the respondents of the two groups. The activities were presented to them individually in a semi-structured interview format. Each session of the experiment was realized in a silent room with a PC, with only the respondent and the interviewer, i.e., myself, present. The whole session was video-recorded and later partially transcribed. The consent of the respondents to the video-recording was obtained. The work took each respondent between 30 and 70 minutes.

Fig. 8 and 9 present an overview of activities for the two groups of students. The word 'session' will be used for a series of activities. Thus, each respondent took part in one session.

[29] Including goniometric equations and inequations, elements of predicate logic and synthetic geometry.

Fig. 8: Scheme of the chronological order of activities for Group D respondents (definition group scenario)

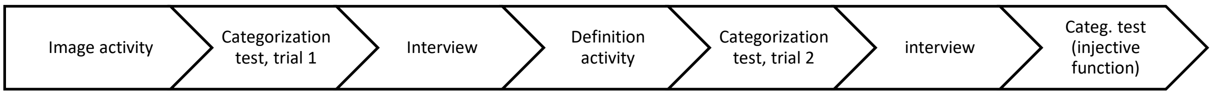


Fig. 9: Scheme of the chronological order of activities for Group I respondents (image group scenario)

3.3.1. Definition activity

The *definition activity* was very simple; it consisted of the presentation of the definition of a Tall function[30] to a respondent in a written form followed by an oral explanation. The respondent was asked to think about the presented concept for five minutes at most. The introduction of the definition activity was roughly as follows.

`Interviewer: What is this all about… in the same way as there are even functions, injective functions and so on, I will show you a definition which determines a set of functions. I will let you think for a while, 5 minutes maximum, and after that, you will take a test in which the graphs of some functions will be presented and you will have to decide if that graph represents the defined function or not. After that, we will discuss it and three more activities will follow.`

No questions were allowed.

3.3.2. Image activity

The aim of the *image activity* was to develop a personal concept image of a Tall function in the mind of a particular respondent. It consisted of the presentation of examples and non-examples of a Tall function. The figures used in the image activity were designed to represent the problematic aspects necessary to be considered (see Sections 2.2.2 and 2.6.2) when one decides about the membership of a figure in the appropriate category (see Appendix A). The following transcript is an exemplary introduction of the image activity with respondent 222. All the sessions followed the same scenario and differed only slightly.

`Interviewer: All we will talk about is a set of functions. Similarly, as there are injective functions or odd functions, I will show you a group of functions and we will talk about them. (…) Now, I will show you several graphs of functions, some of them fit the definition (I call them examples) and some of them do not fit the definition (I call them non-examples). What I need from you… I will present several graphs at once. Tell me when you are ready. The images will remain on the desk. Later, I will give you 5 more minutes to think about them. And then, we will play a game: you can draw five graphs of a function and I will tell you for each, if it represents the Janda function`

[30] Unlike in the second pilot study where the concept of a Tall function was explained to respondents in advance, in the Main study, it was the respondents' task to get to the meaning of a Tall function concept. To prevent any pre-concepts and possible connotation caused by the word *tall*, the pseudonym *Janda function* instead of a Tall function was used during the whole session.

```
or not or that I can't tell without additional information. The Janda function
is defined somehow and the point is how good an image you will develop about
this function. And why are we doing all this? After this activity, you will
take a computer test, where you will have to decide whether the presented
graphs represent the Janda function or not. When you finish the test, some
other activities will follow.
```

3.3.3. Simple categorization tests

After the definition and the image activity, the simple categorization test followed as described in Section 2.2, whose organisation was informed by the pilot studies in two ways. First, as the 'Effect of the first figure(s)' for the group data was found in the second Pilot study, the individual sessions were to be organized[31] which, presumably, would mean that a respondent could concentrate on a task more. This could lessen the influence of the effect. Second, in accordance with the results of the second Pilot study, it was decided that the respondents would not be aware of the reaction time measurement. The graphs in the two trials were the same but they were allocated to the students in a random order.

The last activity was the same simple categorization test focused on the *injective function* (figures used in this test differed from the previous tests). Figures used in the test can be found in Appendix B.

3.3.4. Designing questions for interviews

After both categorization tests with a Tall function, a semi-structured *interview* with a respondent was conducted. The interviewer followed a prepared scenario of questions and tasks which were partly informed by the experience from the two pilot studies. However, more experience was needed of how to ask questions and set tasks to get as much information as possible about the respondent's reasoning. Thus, after conducting sessions with the first three respondents according to the Image group scenario and before proceeding to the work with other respondents, the sessions with the three respondents were transcribed and analysed qualitatively to:

- identify appropriate examples and non-examples of a Tall function to be used in the image activity and determine their suitable number,
- identify an appropriate way of presenting figures to a respondent,
- further specify questions for the interview,
- observe, compare and evaluate the difficulty of the image activity.

None of the three respondents formulated a mathematical definition; thus, question 2 below was added. When asked to "Draw five examples of the Janda function" and "Draw five non-examples of the Janda function", the respondents only drew a few examples and non-examples which were significantly different. Therefore, the required number of examples and non-examples to be drawn was reduced from five to three (questions 3 and 4). The original intention was to let the respondents describe their decision process about each of the figures in the test, which proved problematic. The respondents were not able to say how exactly they decided. To prevent questionable and fictional descriptions from the respondents, it was decided to discuss only those figures which were recalled by a particular respondent, following the task to re-draw the graphs presented to him/her in the test (question 8 and 9).

[31] In Pilot study 2, the preparatory session was organised together for the whole group of students.

Thus, the following questions were used with the remaining 18 respondents in both interviews.

1. How would you define the Janda function?
2. How would you describe the Janda function to someone who knows nothing about this concept?
3. Draw three examples of the Janda function. You can use whatever you have seen earlier.
4. Draw three non-examples of the Janda function.
5. Draw an untypical example if you can think of any.
6. Describe how you were deciding during the test when categorizing presented figures – how the decision process ran.
7. Would you say that you decided automatically or that you thought a lot about it?
8. How many graphs presented in the test would you be able to re-draw now? Draw them.
9. How did you decide about each of those re-drawn graphs?

3.3.5. Designing tasks for the image activity

When carrying out the image activity with the first two respondents, figures from Appendix A (G1–G6) were presented to them all at once, only divided into examples and non-examples. The approaches of these two respondents when dealing with the figures differed. The first respondent observed and examined all the examples first and all the non-examples afterwards. The second respondent started to organize all the figures into smaller groups following his own criteria.[32] Moreover, the course of the first three sessions revealed that the large number of figures (8 examples and 11 non-examples) was hard to manage for the respondents. It was observed that the respondents did not pay attention to all of the figures. These three sessions led the interviewer to changing the design of the image activity to avoid different approaches of respondents.

First, the figures were presented in the form presented in Appendix A (i.e., group G1 as the first, then group G2, etc.) and introduced by words: "Here are the examples and here are the non-examples." Second, groups of figures (G1–G4) were created primarily based on the shape of graphs to stress critical aspects of graphs which were necessary for creating a good image of a Tall function concept (see Section 2.2.2). A bigger number of non-examples stem naturally from the necessity to present aspects that a Tall function does not have, compared to those which it has. Third, the number of figures used in the image activity was reduced to 4 examples and 7 non-examples (groups G1–G4); the excluded figures (groups G5 and G6) can also be found in Appendix A.

All the figures were presented in the same way and order to all remaining 18 respondents.

3.4. Analysis of data

The study is of a mixed design, the data was analysed in both a qualitative and quantitative way.

All the interviews were repeatedly watched and the relevant sections were transcribed and analysed in a qualitative way. The goal of this analysis was to describe a respondent's process of decision making when categorizing the figures in the simple categorization test. Specifically, how the respondents formulated their definitions, how they described the concept of a Tall

[32] Only by observing his activity, it is possible to say that some of these criteria were '*Shape of the graph*' and '*Open/Closed domain*', but this was not the focus of the study (see Section 3.6.2).

function, what examples and non-examples they used for their descriptions and how they described their decision making process during the simple categorization test. This information was used for the description of concrete aspects of graphs which were followed by the respondents during the categorization. These aspects were classified and the extent to which they influenced the respondent's decision was observed.

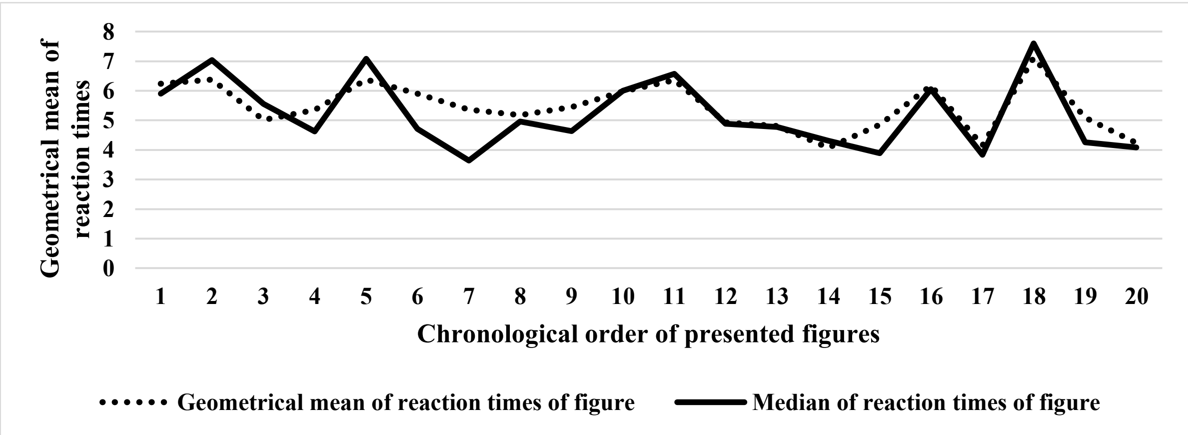
The quantitative data of the three simple categorization tests was obtained in the same way as described in the second Pilot study (see Section 2.2.1). Data were analysed using the quantitative methods based on those described in the second Pilot study (see Section 2.6).

The results of the qualitative and quantitative analyses were compared and discussed as follows. First, the reaction time of a respondent was normalized using the systems of ranks: each reaction time of each figure was ranked from the fastest to the slowest one (numbered from 1 to 20 respectively). Second, the average reaction time rank for each figure was calculated, using the concrete ranks from all the respondents.[33] Based on the results obtained, problematic and otherwise noteworthy figures were selected and confronted with individual aspects of graphs as results of the qualitative analysis.

3.5. Quantitative results

3.5.1. Effect of the first figure(s)[34]

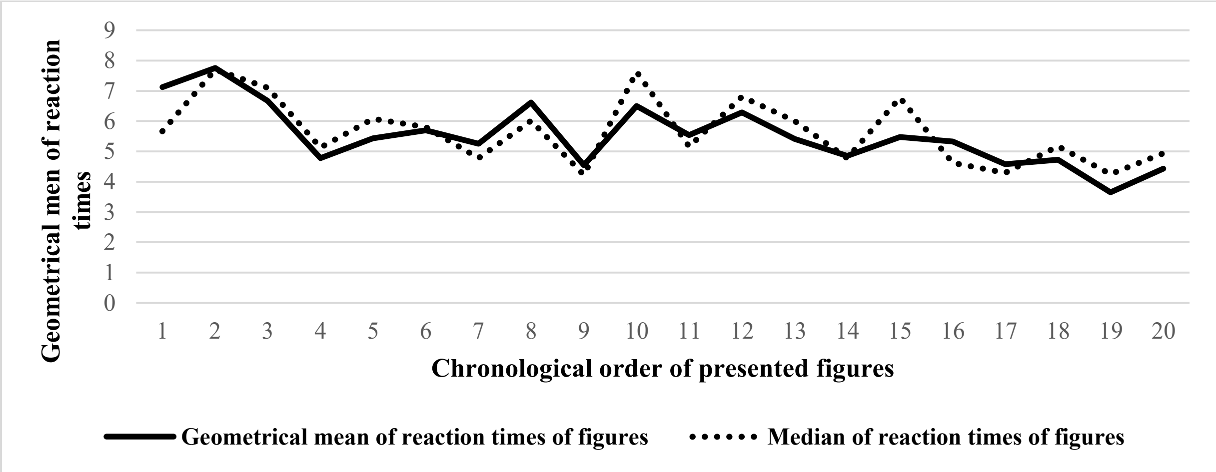
It is natural to start with results which can influence the nature of the ensuing analysis. The 'Effect of the first figure(s)' was not as strong as in the second Pilot study (see Section 2.6.4) as can be seen in Graph 8 (Trial 1) and in Graph 9 (Trial 2). Both graphs also depict medians of reaction times to show the relations between the values of geometrical mean and median of reaction times (for comparison – it can be seen that the medians acquire slightly higher differences).



Graph 8: Comparison of the geometrical means and medians of the presented figures in the chronological order in the first trial

[33] For instance, when respondent A categorized, say, figure e1 as the fastest (rank 1) and respondent B categorized it as the third fastest (rank 3), the average rank of figure e1 would be 1.5. The average rank was counted from all 18 ranks of all 20 figures.

[34] It is worth noting again that in this section both the geometrical means and medians are counted from the reaction times to different figures because they were shown to each respondent in a different (random) order. The values and graphs presented serve only to assess the impact of the 'Effect of the first figure(s)'.



Graph 9: Comparison of the geometrical means and medians of presented figures in the chronological order in the second trial

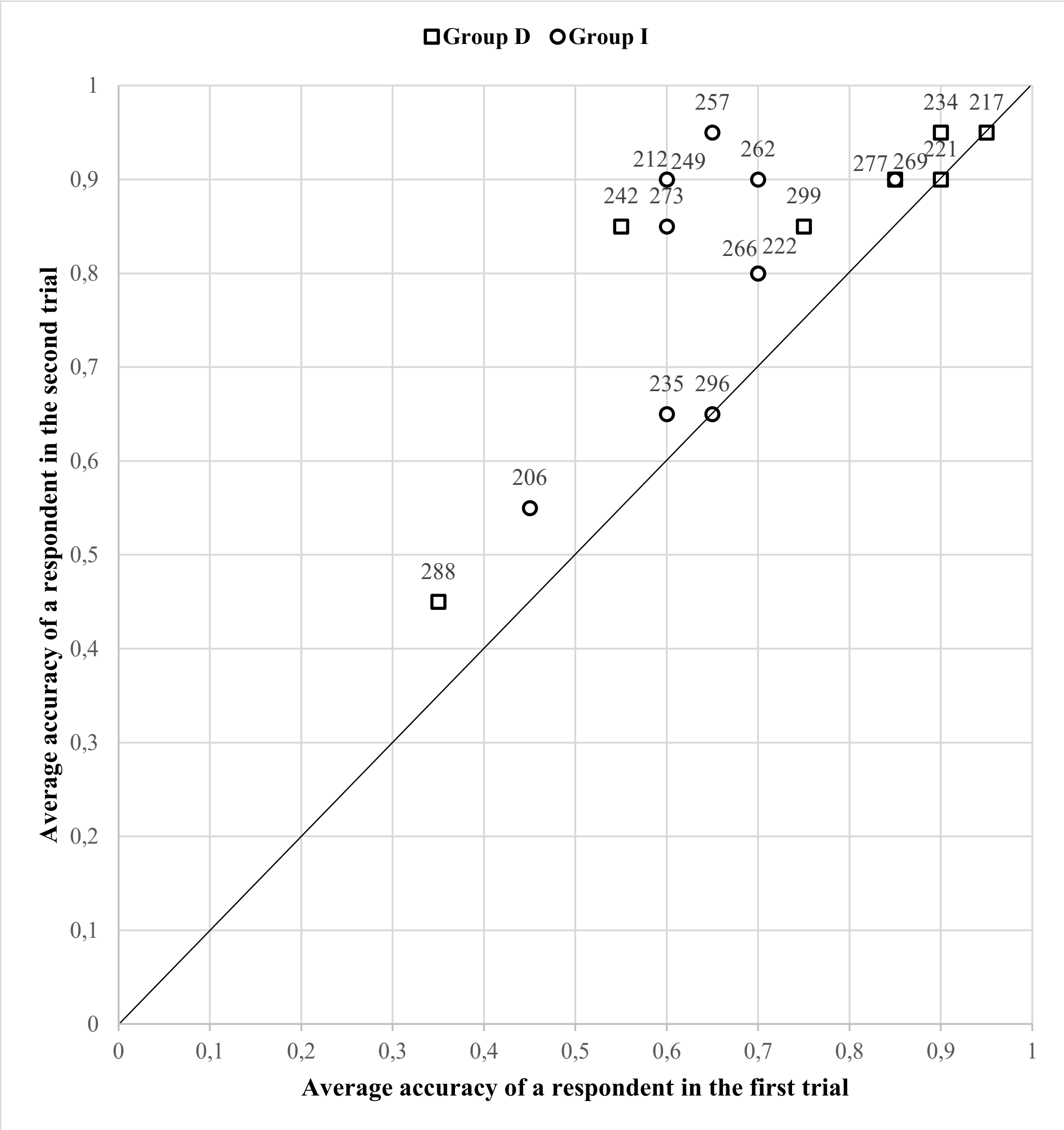
First, neither the geometrical mean nor the median of the first figure presented to the respondents was the highest one in both trials. Therefore, the decision about the first figure was not the slowest one. Second, the visual analysis of both graphs shows that only in the second trial did reactions tend to be faster. These results could be the consequence of using a different methodology from that in the Pilot study, and also of the different sample of respondents. Namely, I believe that the individual session with respondents influenced the results positively. Considering these results, no data trimming was made, because the values did not differ as much as in the Pilot study.

3.5.2. Differences between the Image and Definition groups – accuracy

The differences between the Image group and the Definition group from the perspective of accuracy will be discussed first.[35] As can be seen in Graph 10, almost all the respondents (with the exception of three) increased their accuracy in the second trial. The accuracy of the three other respondents remained almost unchanged (however, this does not necessarily mean that their personal concept image did not change). Moreover, it is possible to see a group of six respondents (five from Group I – respondents 257, 262, 273, 249, 212 – and one from Group D – respondent 242) whose accuracy increased for more than two correct answers.

When comparing accuracies, casewise changes of the respondents' decisions and their number should also be taken into account. The term 'casewise change' refers to the fact that a respondent changed his/her decision about a concrete figure between the first and second trial. The number of such changes was 5.22 on average (ranging from zero changes of respondent 221 to 10 changes of respondents 266 and 296). The respondents changed their decision in 26.1% cases (30.5% respondents of Group I and 19.3% respondents of Group D). Relatively high differences between groups are not surprising taking into account the fact that Group D respondents had better accuracy than Group I respondents, and thus a more stable personal concept image.

[35] It is worth mentioning here that the categorization test has a true/false nature and thus, theoretically, random answers give us the value of 0.5 for an average accuracy.



Graph 10: Comparison of average accuracies of respondents in both trials

An overview of respondents' numbers of changes can be found in Table 2a and Table 2b. *Positive changes* are changes when a respondent answered incorrectly in the first and correctly in the second trial, and *negative changes* vice versa. A deeper discussion of the results presented in Table 2a and Table 2b linked to the qualitative results can be found in Section 3.7.

Table 2a: Number of changes of Group D respondents' decisions between the first and second trial

Group	**Group D**						
respondent	217	221	234	242	288	269	299
positive changes	1	0	1	6	3	2	2
negative changes	1	0	0	0	1	1	0
total changes	2	0	1	6	4	3	2

Table 2b: Number of changes of Group I respondents' decisions between the first and second trial

Group	**Group I**										
respondent	206	222	235	257	262	266	273	296	277	249	212
positive changes	6	4	3	6	5	3	6	5	1	7	6
negative changes	4	2	2	0	1	1	1	5	0	1	0
total changes	10	6	5	6	6	4	7	10	1	8	6

3.5.3. Differences between the Image and Definition groups – reaction time

Differences between the Image and Definition groups in terms of the reaction time can be found in Table 3.

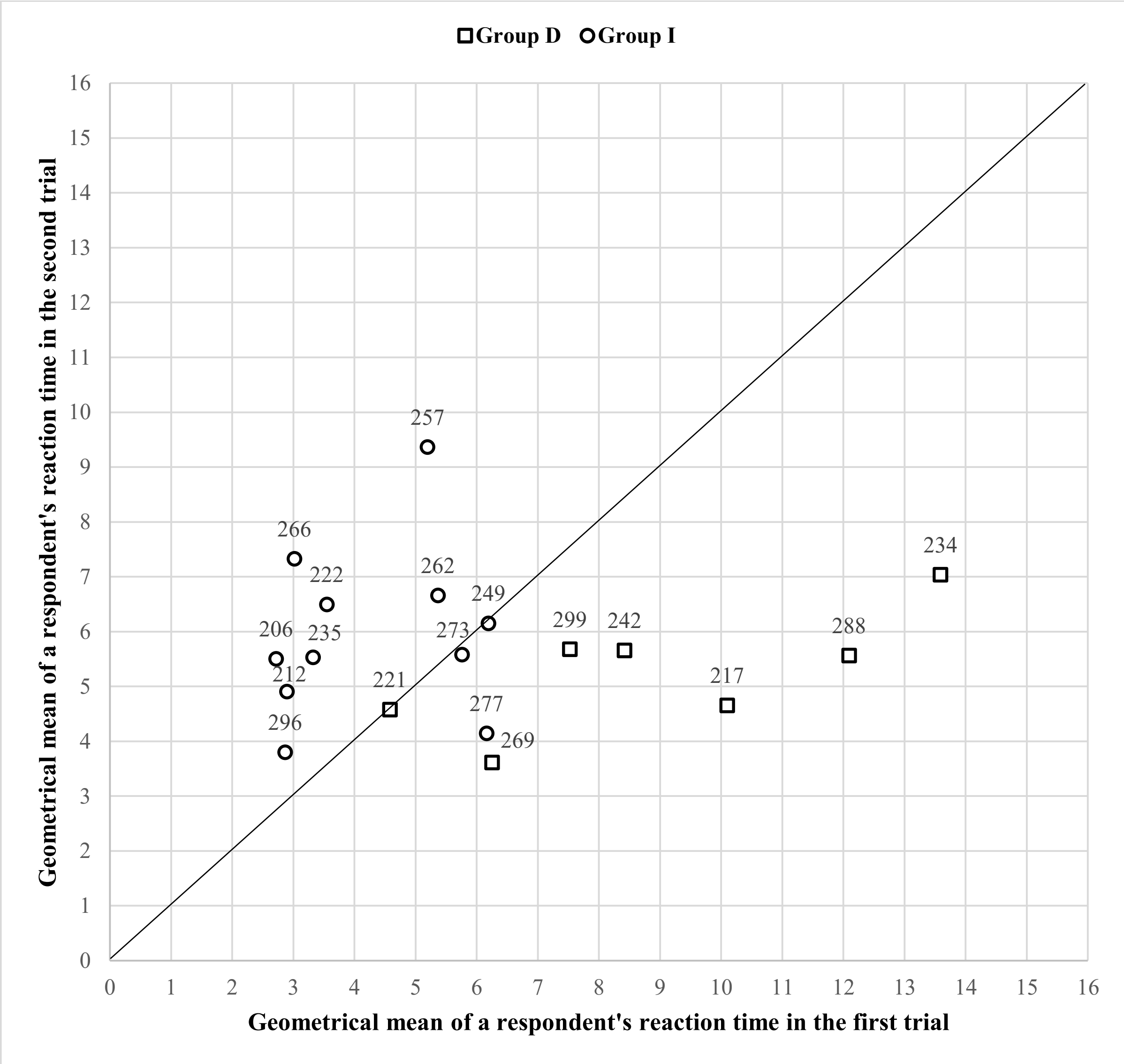
Table 3: Comparison of average reaction times[36] between Groups D and I

		Mean	**MAX**	**MIN**	**MAX – MIN**
Group D	Trial 1	8.42	59.97	1.66	58.30
	Trial 2	5.16	48.03	0.93	47.11
Group I	Trial 1	4.06	30.47	0.86	29.61
	Trial 2	5.78	31.27	1.20	30.07
Both	Trial 1	5.39	59.97	0.86	59.11
	Trial 2	5.53	48.03	0.93	47.11

First, the geometrical mean of reaction times of the Definition group respondents was more than two times greater than the geometrical mean of reaction times of the Image group respondents in the first trial, which means that the respondents of the Image group categorized graphs rather faster than the respondents of the Definition group. A similar conclusion can also be observed from the differences between the longest and shortest reactions (MAX − MIN). Connecting this to the results from Section 3.5.2, the Definition group respondents answered more correctly (by 10% approximately) but it took them longer to decide. The results are not surprising, taking into account the fact that the respondents from the Image group had only the intuitive example-based personal concept image as developed during the image activity. The results support the core theory of this study that an approach to one's decisions differs between both groups.

Second, the geometrical means of the respondents' reaction times were compared after the second trial, showing that the reaction times converged approximately to their mean value. This can also be seen in Graph 11 (the values on the horizontal axis are more extended than those on the vertical axis).

[36] Values were found as a geometrical mean of all reaction times of all respondents in the appropriate groups.



Graph 11: Comparison of geometrical means of respondents' reaction times in the first and second trial

Together with the cross-trial increased accuracy of almost all the respondents, this seems to imply that a smaller variation of the reaction times might mean a well-developed personal concept image. In a group of respondents with an undeveloped personal concept image, the reaction times vary more. Thus, further analyses were made. The variation of the geometrical means of every respondent's reaction times was counted and compared to his/her accuracy. The correlation[37] between these two values is 0.41 in the first and 0.18 in the second trial, respectively. Both correlations are positive which does not support the conclusion above (for higher accuracies, the variation should be smaller, meaning, in an ideal case, a correlation closer to -1). However, the values in the second trial vary less which shows a slight tendency towards the conclusion (in the second trial, the personal concept image of the respondents was better and variation smaller). Nevertheless, a bigger sample would be needed to test this further.

[37] Pearson's correlation coefficient was used taking into account the fact that geometrical means should be distributed approximately normally, see also (Pilous & Janda, 2017).

Third, the results presented in Graph 11 show that the geometrical mean of some respondents' reaction times remained almost unchanged (respondent 221 from Group D and respondents 273 and 249 from Group I) and the geometrical means of the others differed greatly. For instance, respondents 234, 288 and 217 from Group D were much faster in the second trial, whereas respondents 266 and 257 from Group I were much slower in the second trial.

Thus, it is natural to compare how a respondent's reaction times changed within both trials with the two parameters discussed above – how a respondent's accuracy and decisions changed between the trials. This comparison is expressed using the correlation coefficient as follows.

- Correlation between the cross-trial change of the respondent's reaction time and accuracy was 0.29.
- Correlation between the cross-trial change of the respondent's reaction time and number of changed decisions was 0.58.
- Correlation between the cross-trial change of the respondent's accuracy and number of changed decisions was 0.29.

The second correlation is the strongest, meaning that the respondents who tended to be slower in the second trial than in the first also tended to change their decision in more cases.

3.5.4. Learning effect

In both groups, it can be seen that the 'learning effect' was manifested. In the second trial, the respondents reacted in a more correct way and it is noteworthy that the geometrical means of reaction times were closer to each other – the reaction times of particular respondents slightly converged to one another.

The presence of the learning effect raises the question to what extent the improvement in the students' responses can be attributed to the experience with the figures in the first trial and to what extent it can be attributed to the second trial itself. The influence of passing the test does not seem to be high. First, the respondents only had a little time to think about the figures used in the test – each activity required their attention, there were no breaks during the session. Second, they did not know that they would be tested for the second time and finally, they did not know that the same figures would be used in the second trial.

These arguments give an impression that the respondents between both trials improved their personal concept image of a Tall function (not only their conviction about categorizing a few figures), during the follow up interview, that is, at the beginning of the second activity (when the interviewer presented them with examples and non-examples in the image activity or with the definition of a Tall function in the definition activity) or during the test itself. Moreover, almost the same average accuracy and reaction time in the second trial also indicate that the order of definition-image or image-definition activities have no substantial influence on the results (for the concept of a Tall function) but the data does not allow us to make any definitive conclusions.

3.5.5. Differences between examples and non-examples

The findings of the second Pilot study showed that it is reasonable to observe the difference in reaction times between examples and non-examples. In Section 2.6.4, a tentative conclusion was presented that a significant difference in reaction times for examples and non-examples might point to a well-developed personal concept image while an undeveloped personal concept image leads to a relatively small difference in reaction times of examples and non-examples. This conclusion was investigated in several ways. First, the correlation of differences of the respondents' geometrical means of examples and non-examples and the number of correct answers were considered. The correlations of these values was -0.25 in the second trial. The negative correlation goes against the above tentative conclusion.

The comparison of geometrical means of the respondents' reaction times in all three categorization tests (the first and second trials and injective function) are presented in Table 4. The difference of reaction times between examples and non-examples was smaller in the first trial than in the second one. In the first trial, almost the same results were provided by the respondents of the Image group, the difference is greater for the respondents of the Definition group. Moreover, the difference between the average reaction times of examples and non-examples increased in the second trial when the respondents' personal concept image of a Tall function was better on average, as discussed in the previous section. The difference between geometrical means of reaction times for examples and non-examples can also be seen in the case of the injective function.

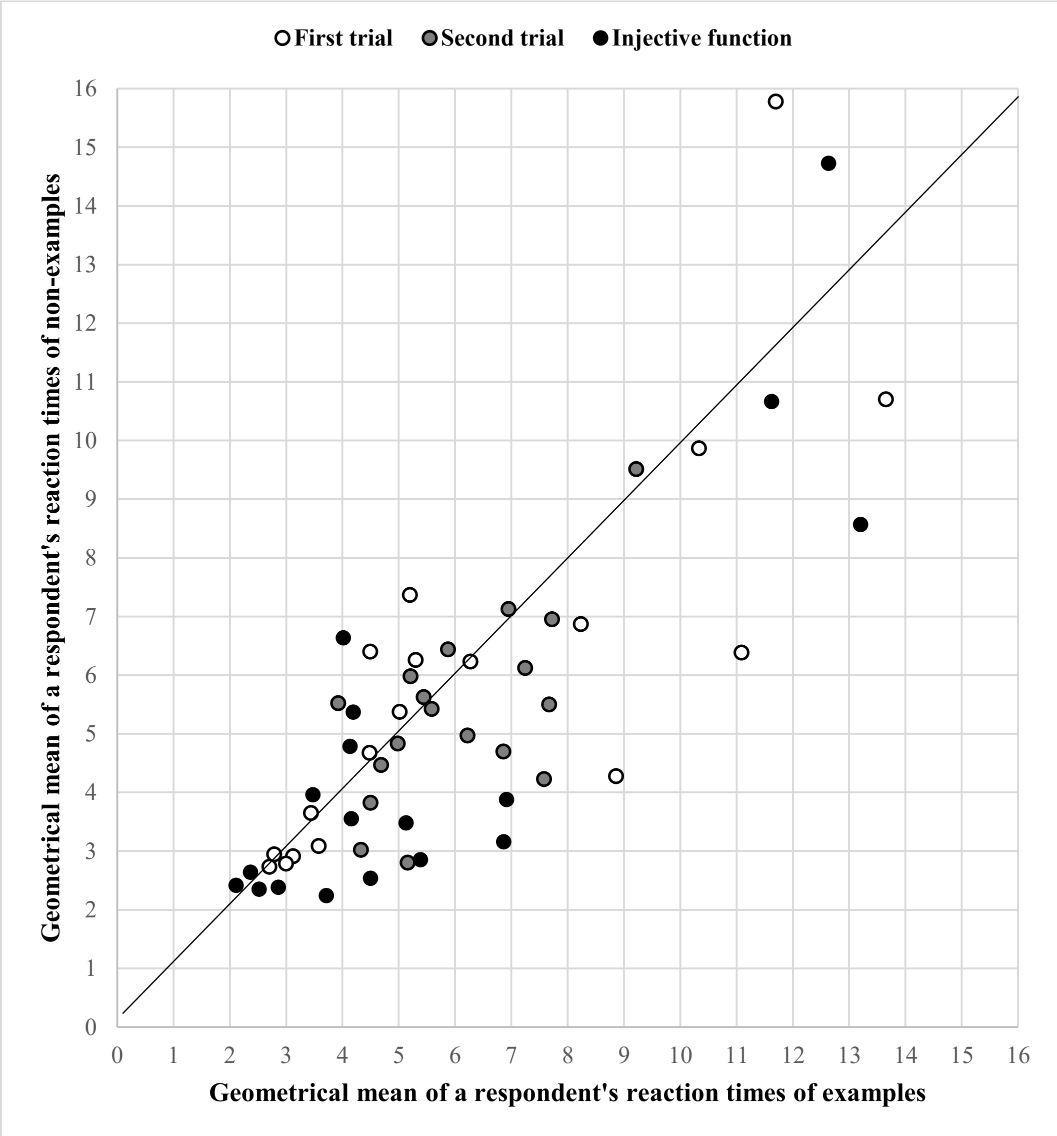
Table 4: Comparison of reaction times and accuracy of examples and non-examples

		Geometrical means of reaction times[38]			**Accuracy [average]**		
		1st trial	2nd trial	Injective function	1st trial	2nd trial	Injective function
Both groups	Examples	5.48	5.90	4.75	0.66	0.81	0.77
	Non-examples	5.30	5.18	4.04	0.72	0.82	0.82
Group I	Examples	4.05	6.09		0.61	0.81	
	Non-examples	4.07	5.49		0.68	0.80	
Group D	Examples	8.84	5.63		0.73	0.81	
	Non-examples	8.02	4.73		0.77	0.86	

The statistical analysis of the significance of this difference provides only limited support for the sample of this size. Thus, additional observations were made. Graph 12 shows the respondents' geometrical means of reaction times of examples and non-examples in all the tests. The tests are denoted by different symbols which means that the reaction times of a particular respondent is behind three different symbols.

It can be seen that the respondents categorized examples rather slower than non-examples for a Tall function. Thus, the difference in the categorization of examples and non-examples seems to be a promising result of the study that deserves further research on a larger number of respondents.

[38] The values represent the arithmetical mean of geometrical means.



Graph 12: Comparison of geometrical means of the respondents' reaction times of examples and non-examples in all the tests

3.5.6. Accuracy of the categorization of figures

Next, the accuracy will be considered from the perspective of particular figures. Table 5 presents the number of correct categorizations of particular figures and the number of cross-trial changes in decisions.[39] The number of correct answers is presented relatively due to the different numbers of respondents in the Definition group (7) and the Image group (11).

The figures for which the results differ between the Definition and Image groups are noteworthy. For instance, figures e4 and e5 represent graphs where the Image group respondents had substantially higher accuracy than those from the Definition group even when the latter had generally higher accuracy. On the other hand, figure e6 was categorized by the Definition group respondents more correctly while only one respondent from the Image group

[39] For instance, four respondents changed their decision about figure e1 in the second trial compared to the first trial. Three of them categorized the figure incorrectly at first and after that correctly, one categorized the figure correctly at first and after that incorrectly.

categorized the figure well. An additional analysis of these phenomena will be presented in relation to the qualitative data in Section 3.8.

Table 5: Comparison of accuracies of particular figures in both trials and comparison of cross-trial changes of decision

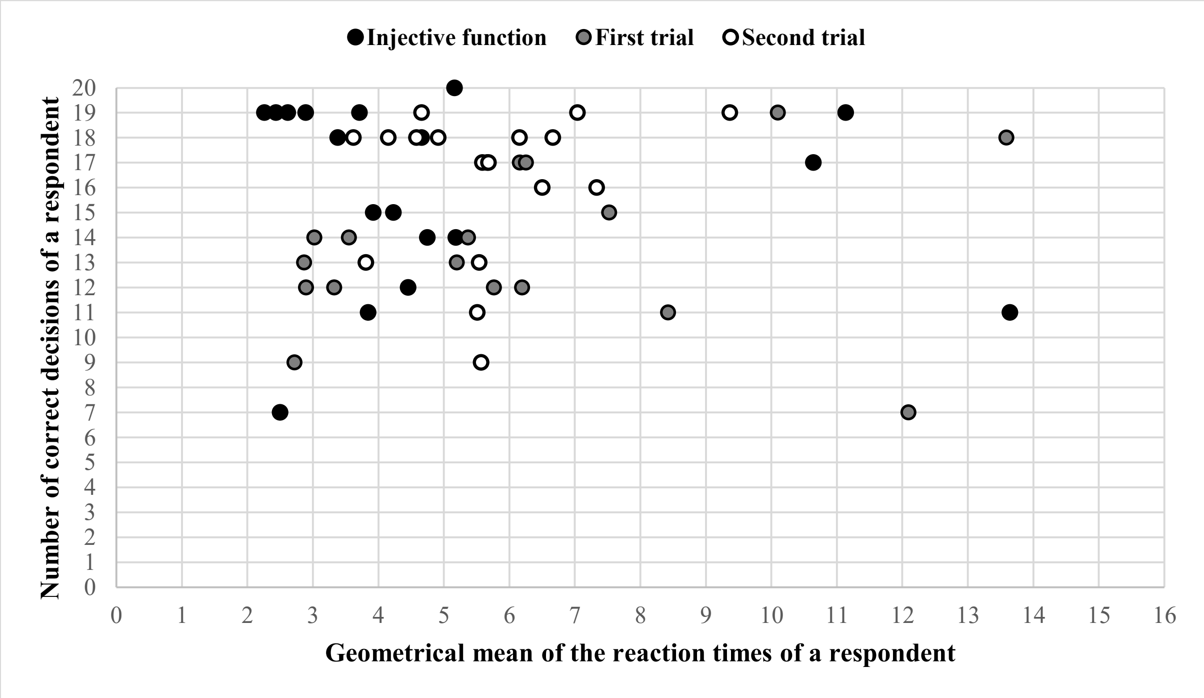
	Number of correct answers				**Number of cross-trial changes**		
	Trial 1		**Trial 2**				
	D	I	D	I	Positive	Negative	All
e1	86%	73%	86%	91%	3	1	4
e2	86%	82%	86%	82%	2	2	4
e3	71%	55%	86%	100%	6	0	6
e4	57%	91%	86%	73%	3	3	6
e5	57%	82%	100%	91%	5	1	6
e6	71%	9%	71%	45%	4	0	4
e7	57%	73%	57%	73%	2	2	4
e8	86%	27%	71%	82%	7	2	9
e9	86%	64%	86%	82%	3	1	4
e10	71%	55%	86%	91%	5	0	5
n1	86%	100%	86%	91%	0	1	1
n2	86%	73%	86%	73%	1	1	2
n3	86%	64%	100%	91%	4	0	4
n4	100%	91%	100%	73%	0	2	2
n5	86%	45%	100%	82%	6	1	7
n6	86%	82%	100%	91%	3	1	4
n7	43%	45%	57%	73%	4	0	4
n8	29%	0%	29%	45%	5	0	5
n9	100%	91%	100%	82%	1	2	3
n10	71%	91%	100%	100%	3	0	3

3.5.7. Comparison of a Tall function and an injective function categorization

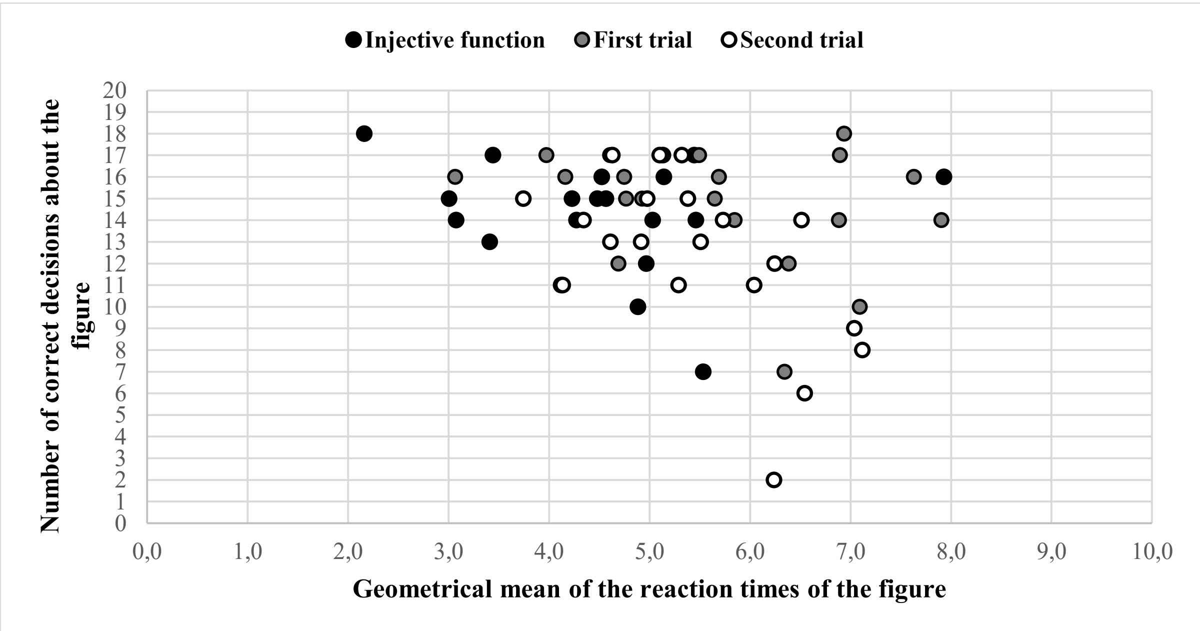
Next, the results of the third simple categorization test focused on the categorization of injective functions will be presented. Graph 13 represents the comparison of reaction times and accuracy of particular respondents in all three categorization tests. As can be seen, the results show signs of similarity. First, the geometrical means of the reaction times are relatively comparable, taking values from 2s to 8s for all the respondents in all the tests. Three groups of respondents can be seen in Graph 13. The group on the top-left of the graph represents the respondents with accurate and fast answers, the group on the bottom-left of the graph represents the respondents with relatively fast reactions whose accuracy varies and the group on the right of the graph represents the respondents who reacted relatively slowly and their accuracy differed a lot.

Graph 14 depicts the comparison with the results of both categorization tests of a Tall function and categorization test of the injective function in the perspective of figures. It can be seen that the figures categorized faster were categorized rather accurately; the accuracy of figures categorized slower varies more. This pattern seems to be the same for both concepts in question and provides promising results for further investigation.

These findings seem to be promising since they represent a possible connection between two different (but still very similar) concepts. However, for deeper analysis, the normalization of the data eliminating general reaction time of a respondent (for instance, the system of ranking described in Section 2.6.3 and applied in Section 3.8) would be needed. I decided not to make this analysis on this sample size without additional information from the respondents since I believe the results would be too distorted.



Graph 13: Comparison of geometrical means of reaction times of respondents and the number of their correct answers



Graph 14: Comparison of geometrical means of reaction times of figures and appropriate number of correct answers

3.6. Qualitative results

In this section, qualitative results related to the interviews with the respondents will be discussed. First, the work of two respondents who were able to get to almost the precise definition of a Tall function based on the image activity is explored. Second, aspects guiding the respondents' decision which appeared dominant in the interviews are described. Finally, two general characteristics of these aspects were observed – whether they lead to a respondent's decision and whether they are objectively measurable.

Note: The following paragraphs contain quotations from the interviews which were translated from Czech. Some corrections of the spoken language, not changing the meaning, were made to make the utterances intelligible. I stands for the interviewer, R stands for the respondent. Graphs used in the simple categorization test (e1,… n1,…) are in Fig. 1 and Fig. 2 on p. 30–31.

3.6.1. Cases of respondents 266 and 277 from Group I: Reaching the definition based on the image activity

After the first simple categorization test, respondent 266 formulated the definition of a Tall function[40] during the interview.

```
I: So, how would you define a Janda function?
R: Hmm, the border points have to be included every time. When I take the range of the function, it has to…
I: Has the mathematical definition occurred to you? How to define it? It does not have to be necessarily the exact definition… use mathematical terminology to describe it.
R: I don't know… let's say the absolute value of the minimum is lower than the absolute value of the maximum.
```

The statement is nearly the same as the original definition of a Tall function. The situation that somebody from the Image group would be able to state this definition was unpredicted. However, the following dialogue shows that the personal concept image of respondent 266 was not accurate.

```
I: Could you describe the process of your decision? When you remember the test, how did you decide?
R: So… [hesitates]
I: The definition… we know it now… more the process… how it worked.
R: So, first, I look if both border points are included, eventually, if it goes to infinity… If this function goes to infinity [points to Fig. 10] then, ehm, then it would not be included [in the category of Tall functions] because it is… when I say its… minimum and maximum in the absolute value have the same value.
```

The respondent considered whether the border points are included in or excluded from the graph of function in his own definition but he did not realize that it did not satisfy the definition every time (even when a maximum and a minimum exists, the border points can be excluded).

[40] It should be noted again that to prevent any preconceptions based on the name 'Tall function', I used the name 'Janda function' during all activities of the Main study.

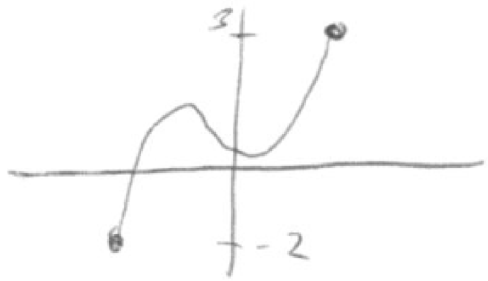


Figure 10: Respondent 266's drawing of a graph to explain his decision process

Similarly, respondent 277 was able to develop an even more precise personal concept image when he described the concept.

```
R: I think, that the [Janda] function has a minimum, has a maximum and the value of the maximum above the axis is greater than the distance of the minimum from zero value.
I: OK, how would you explain this to me if I were someone who had never heard about it? How would you explain to me how to decide whether something is or is not a Janda function?
R: So, first, I look if the function has a minimum and a maximum and in the case that it does not, then I immediately know that it is not [a Janda function]. Then I look at the values of the minimum and maximum and compare them.
```

The cases of respondents 266 and 277 show that it is possible to discover the full definition based on the image activity. This had not been considered probable when designing the study; however, it makes the results from both groups of respondents more comparable (showing that the personal concept images of respondents of both groups do not have to differ much).

3.6.2. Aspects of graphs guiding the decision process

Particular aspects of graphs considered by respondents when categorizing figures in the simple categorization test were observed. The descriptions of aspects are simplified and shortened. Individual aspects and their concrete representatives often do not reflect their exact mathematical meaning, they rather serve as unifying elements that were observed by the respondents.

'Max[41] is greater than |min|', 'Max is smaller than |min|'

These aspects represent the nature of the definition of a Tall function that the maximum of the function is greater than the absolute value of the minimum. A respondent determines the values of maximum and minimum (at least visually) and compares them.

```
R: When I consider this function [points to Fig. 11], here I decided almost instantly because minimum… the absolute value of minimum is smaller than the maximum.
```

[41] In the thesis, 'max', 'min', 'sup', 'inf' mean maximum, minimum, supremum and infimum, respectively. Supremum and infimum are defined on the extended real numbers. Thus, for instance, function $f\colon y = x^3$ has the supremum ∞, infimum $-\infty$ and |inf| = ∞. It also implies that the aspect "sup is greater than |inf|" (see below) is a 'stronger' claim than the aspect "max is greater than |min|".

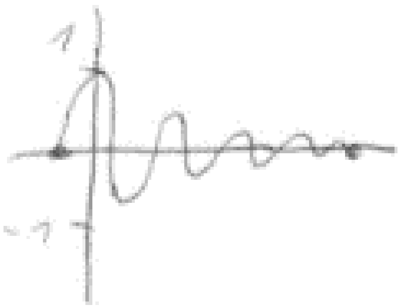


Fig. 11: Reconstruction of figure e7 by respondent 266

'Sup is greater than |inf|', 'Sup is smaller than |inf|'

When considering these aspects, the respondents are actually considering the previous ones. However, in this case, they do not take into account whether the maximum and minimum exist. Consider, for instance, the function corresponding to figure n7 whose supremum equals to $+\infty$, whereas its infimum is zero. In that case, the aspects in question can lead the respondent to categorize the figure as an example. Respondent 221's description of his process of decision illustrates it.

```
R: The maximum of this function [drawing the graph in Fig. 12] would be in infinity and the absolute value of the minimum would be one, which fits the definition.
```

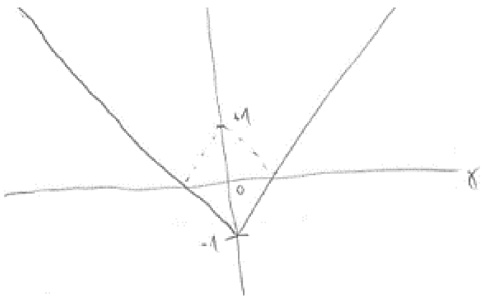


Fig. 12: Respondent 221's drawing as an example of a Tall function

Respondent 221 accepted the $+\infty$ as a value of the maximum. It is probable that he would also accept the $-\infty$ as a possible value of the minimum, which would represent the aspect '*sup is smaller than |inf|*'.

'Sup is equal to |inf|'

In the simple categorization test, several figures represented the functions where the supremum is equal to the infimum (n1, n3, n10 and for some respondents also e7 or e8). For instance, respondent 269 compared the supremum and infimum in the following way.

```
I: How did you decide for each of these graphs [points at the figures that the respondent remembered from the test and drew them]?
R: You mean if I was pressed yes or no and why?
I: Yes, yes or no and why.
R: I think [Fig. 13] both [parts of the graph] end in the infinity so I pressed "no" because it's equal. The maximum is not greater than the absolute value of minimum.
```

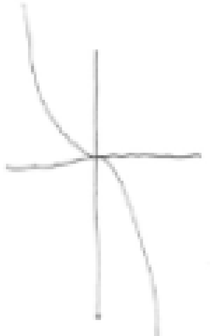


Fig. 13: Reconstruction of figure n1 by respondent 269

The respondent reconstructed and described his decision process about figure n1. This approach is applicable also to figures n3 and n10 where the maximum and minimum exist.

'Maximum or minimum does not exist'

Some respondents realized that the maximum or minimum did not exist for some of the presented graphs and in that case, they decided not to compare the values of their supremum and infimum as in the previous aspects. An example is respondent 234.

```
R: When it was going to the infinity, I got stuck there because I told myself
that it has no maximum, actually. So we cannot say that it is a concrete
value.
```

Note that figures n7, n8 and partly n9 represent graphs which cannot be decided unambiguously without context of how the tending to the axis should be understood. The experience achieved in the pilot studies showed that the discussion about the context of the graph could evoke other issues (for instance, that the respondents would focus on the discussed aspects more than the others). The respondents naturally observed this ambiguity, too. Thus, this case was described separately as "not well observable max or min" (see below).

'All points are above the x-axis', 'All points are below the x-axis'

These aspects differ from all the previous ones in their complexity. The criterion behind them was simple – the respondents observed whether the whole graph is above or below the axis. For instance, respondent 299 drew the parabola (see Fig. 14) as an example.

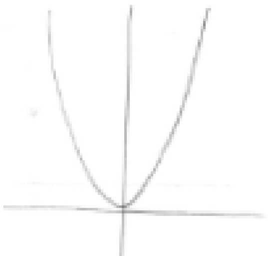


Fig. 14: Respondent 299's drawing of a parabola as an example of a Tall function

```
R: [Drawing one of the three examples of a Janda function]
I: Why the parabola?
R: Because it is whole above the axis so it is… easy.
```

For the respondent, the sufficient condition for categorizing the figure as an example is that all the points of the graph are above the *x*-axis or at least on the axis (cannot be said precisely based on the drawing), even if the maximum of the function does not exist. The second illustration of these aspects comes from the interview with respondent 212.

```
I: Here you decided automatically [pointing at the graph in Fig. 15]? Or…
R: Yes.
I: And you marked it as…
R: It is [an example].
I: And why?
R: Because I had a theory that if the whole function is in the positive [part
of the coordinate system with respect to y-axis], then it is a Janda function.
```

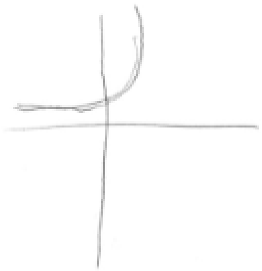


Fig. 15: Reconstruction of figure n7 by respondent 212

The respondent expressed the following process: first, if all the points are above the *x*-axis, the decision is determined by this and leads to its categorization as an example. Second, if not all the points are above the *x*-axis, then the respondent seeks another aspect to decide upon.

An instance of the aspect '*all points are below the x-axis*' is observable in respondent 299's work. She reconstructed figure n4 as shown in Fig. 16 and described his decision. She categorized the figure as a non-example.

```
R: Here was minus 0.5 [points to the intersection with the y-axis]… and it
was whole below the x-axis, so it was easy to decide.
```

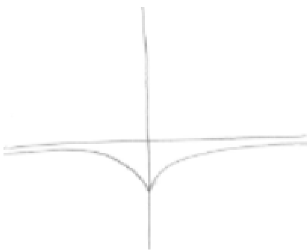


Fig. 16: Reconstruction of figure n4 by respondent 299

'Longer part of the graph is above the x-axis', 'Longer part of the graph is below the x-axis'

While the previous aspects were described by the respondents using relatively precise terms,[42] some respondents worked intuitively with the location of the graph in the coordinate system. This led to formulations such as "Bigger part of the function has to be positive" (resp. 222),

[42] With the exception of aspects '*All points are above the x-axis*' and *'All points are below the x-axis*', regarding possible discussion about cases where some of the points appear *on* the *x*-axis as observed in case of respondent 299 who drew the parabola as an example.

"The graph of the function is slightly above the x-axis" (resp. 296). Another example comes from respondent 212's interview.

I: Is there anything else, do you think?
R: Then, I think… if the bigger part of the graph is in the positive y-axis or negative y-axis.

Respondent 212 expressed the described proportionality of the graphs as a criterion to be considered. Consider also the case of respondent 222.

I: If I was someone who had never heard about a Janda function, and I needed to learn it… How would you explain it to me? What should I do when I get a graph…?
R: When I have a graph [Fig. 17]… I look which part… divided by x… if this part [above the x-axis] is bigger than this one [below the x-axis]… and if the points are included… then it will be a Janda function.

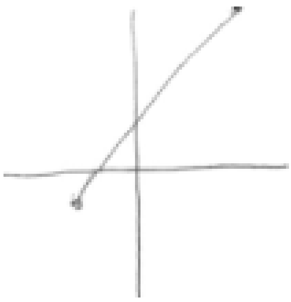


Fig. 17: Respondent 222's drawing of a graph to explain his decision process

The difference outlined above is possible to see in the respondents' formulations. It is not clear which property of the graph[43] the respondents paid attention to. For instance, it could be the real length of the curve, the area between the curve and the x-axis, the values of the maximum, minimum, supremum and infimum discussed above, etc. Concrete terms are missing in the respondents' descriptions and thus, there are no concrete values which can be compared by them. The whole process is rather intuitive.

'Open domain', 'Closed domain', 'Unbounded domain'

The respondents often analysed the domain of a function's graph, considering whether the points on the border of the domain are included in or excluded from the graph. Some respondents also considered whether the domain is bounded or unbounded. For instance, respondent 212 partly described the concept of a Tall function based on the image activity.

R: Sure, the range is important. The domain will not play a big role, in my opinion. […] Then, whether all the points are included in or excluded from the graph will play a role.
I: So, the range and whether the point is included or excluded…
R: It is the domain – if it is open or closed.

Immediately after these statements, respondent 212 drew the examples and non-examples in Fig. 18.

[43] It is not even clear whether they had any concrete property in mind or rather expressed their indefinite "feeling".

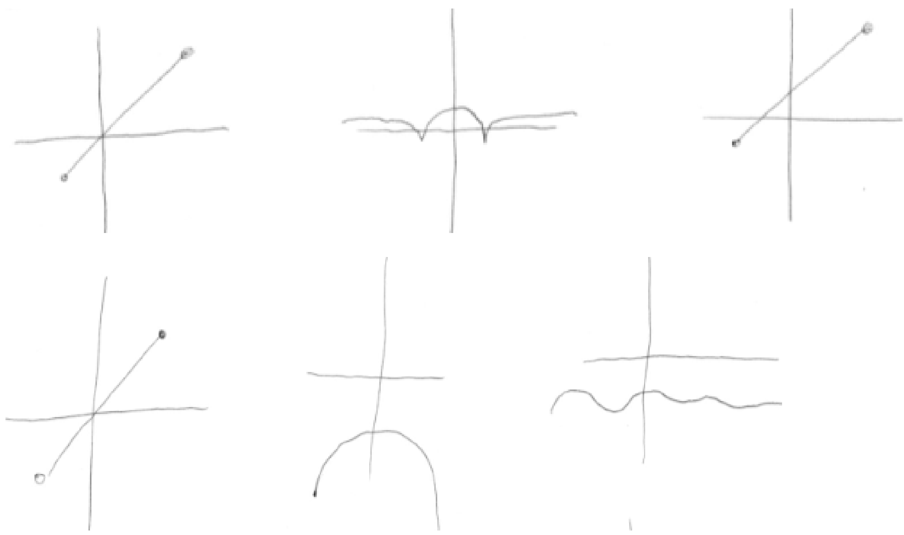


Fig. 18: Examples (top) and non-examples (bottom) drawn by respondent 212

The second example drawn by respondent 212 shows that the unbounded domain of the function did not actually determine her categorization and the appropriate figure was considered to be an example. However, this aspect was a part of her decision making process. This can be seen from her statements even if they did not provide an explicit reason for categorizing the figure as an example or a non-example.

Similarly, respondent 222 described his decision as follows.

R: If there was a function and its bigger part was positive and it had all border points included, then I decided that it was [an example]. Even if the bigger part was positive, but there was one [border] point excluded, then I decided it was not [an example]. And when it was half-to-half, I decided it was not.

This respondent used a combination of two aspects. However, mostly the aspect whether the border points are included in or (at least one of them is) excluded from the graph determines her decision. Nevertheless, it is not clear whether he managed to express her thoughts precisely. For instance, it is possible that she unified border points of the domain with the points where the function takes the values of a maximum and minimum. However, in both cases, the observation of a function's domain was included in her decision process.

'Range of function', 'Bounded function', 'Unbounded function'

The range of the function was considered by some respondents, too, as can be seen in the interview with respondent 212. Her decision is based on the combination of the observation of the domain and range of the function.

R: I think… It is a group of functions which have their domains from minus infinity to infinity and the range is bounded - it [the function] is not from minus infinity to infinity.

Respondent 262 expressed a similar consideration about the range of a Tall function.

R: It has a concrete range which means that it is not going to infinity on the top nor the bottom.

These two aspects are two opposing principles of the same criterion.

'Shape of the graph'

Respondent 221 described his decision about figure n3 as follows.

```
R: This one [reconstruction of figure n3] I remember because I know that here
[the endings of the curve], it was plus infinity and minus infinity. There
[around the origin] was something which confused me.
```

Similar experience was described by respondent 234.

```
R: This [pointing at Fig. 19] I remember because it was something… it was
untypical, how to put it… Such a function, we don't see often.
```

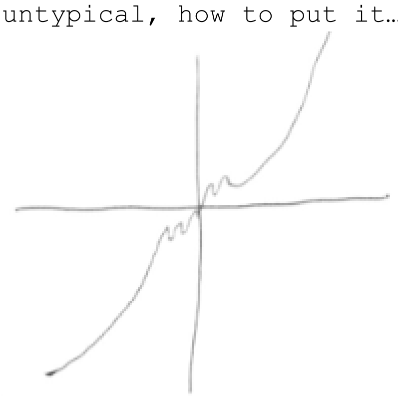


Fig 19: Reconstruction of figure n3 by respondent 234

Figure n3 obviously influenced the decisions of both respondents because of its shape. Similar statements about the shapes of graphs were observed when respondents described their decisions about figures e7 (respondent 217) and e10 (respondents 217, 242). This aspect was considered mostly intuitively and did not determine decisions of the respondents in their words. Thus, it is probable that the shape of a graph generally has an influence on the reaction time of the decision, as it is natural to evaluate the shape of the figure unconsciously.

'More than one max or min'

During his interview, respondent 217 mentioned that he "was not sure if it has a maximum" evaluating those figures which had more than one maximum or minimum (for instance, when he was describing his decision process about figure e5 – see Fig. 20).

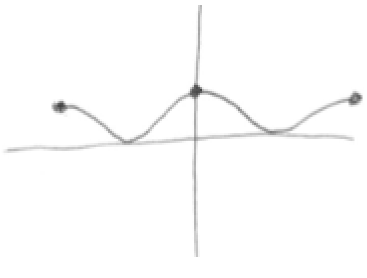


Fig. 20: Reconstruction of figure e5 by respondent 217

Respondent 242 described the same aspect in a general way.

```
I: Did you decide automatically?
R: If I know that I can't find a maximum or a minimum, one of them, then I
automatically pressed N [non-example]. When I saw, for instance, a function
where, at a glance, it could have a maximum but… if it reached that [maximal]
y-value in two points, then I had to think about it.
```

```
I: Yes.
R: Because the maximum does not exist. If I'm correct.
```

While respondent 217 expressed his uncertainty when deciding about a function which has more than one maximum, respondent 242 directly expressed his opinion that the maximum did not exist for such functions. These quotations show that the number of maxima and minima of a function can influence respondents' decisions in both ways.

'Finding of max or min'

The common feature of this and the following aspect is that the respondents had a problem working with points representing the values of maximum and minimum. However, there is an important difference between these two aspects. In some cases, the respondents were not immediately able to identify the points representing maximum or minimum in the presented figure. Respondent 234 described this experience when reconstructing figure e4.

```
R: I got stuck here [pointing at Fig. 21] because I compared these two points
[pointing at the border points]. I missed that the maximum is actually here
[pointing to the correct maximum], I focused on these two endpoints if they
were the same… not the same… just the… central symmetrical.
```

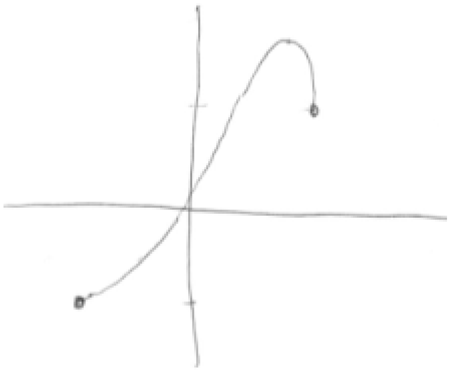


Fig. 21: Reconstruction of figure e4 by respondent 234

This aspect was observed for figures e3 and e4. Some respondents observed the border points of the domains before the points which represent the values of maximum and minimum. It complicated the finding of the points representing the minimum and maximum at the first glance.[44]

'Not well observable max or min'

Unlike the previous aspect, in this case, respondents had problems identifying the precise value of maximum, minimum, or both. The fault is not necessarily on their side. The maximum and the absolute value of the minimum are close in graphs e7 and e8, making them difficult to distinguish.[45] Similarly, graphs n7 and n8 are not clearly interpretable without additional discussion about the respondent's understanding of the graph.

A situation representing this aspect occurred when respondent 277 reconstructed figure e8 (see Fig. 22) and described his decision.

[44] I believe that this is caused by the fact that some parts of the decision making process is realized unconsciously and as such, it cannot be influenced by a respondent consciously.
[45] This was done on purpose, see Section 2.2.2.

R: Here, I wasn’t sure if the minimum is really… The difference was so small here that I was not sure.

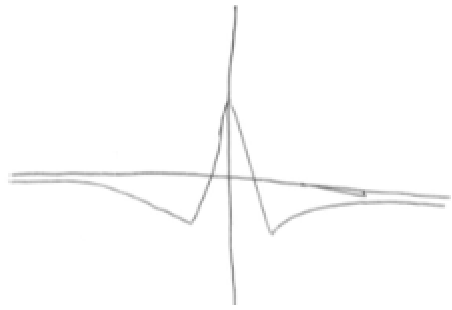


Fig 22: Reconstruction of figure e8 by respondent 277

On the other hand, respondent 299 described her decision about figure e7 as she reconstructed it (see Fig. 23) as follows:

R: This one looks nice… I saw that the absolute value of the minimum is equal to the maximum, so it was easy.

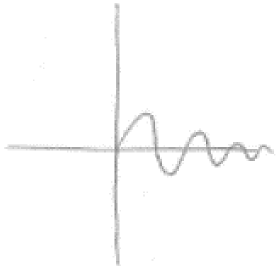


Fig. 23: Reconstruction of figure e7 by respondent 299

As can be seen from the original figure e7, the maximum has actually a greater value than the absolute value of the minimum, and thus, should be categorized as an example. It is noteworthy that the respondent did not think about the possibility that the values differ, but she directly categorized the figure as a non-example.

The difference between the aspects ‘*Finding of max or min’* and *‘Not well observable max or min’* is that they are reactions to different phenomena. When using the latter aspect, a respondent is looking for the concrete values of the minimum and maximum to compare them.

‘Principle of discontinuity’

Figures e10, n9 and n10 are the only functions in the test which are not continuous. Discontinuous functions (and concrete points of discontinuity) were, for some respondents, in their focus. For instance, respondent 262 described the concept of a Tall function.

R: I think that [a Tall function] is continuous which means that it is not interrupted anywhere.

The respondent considered the continuity as a necessary condition for examples of a Tall function which led her to categorize discontinuous functions as non-examples. Some respondents had problems with the concept of continuity itself. For instance, the discontinuity of the graph was confusing for respondent 266.

R: This one, I didn’t consider a Janda function. Because one of these points… The zero is not included…

```
I: I don’t understand. That the zero is not included in the graph.
R: That it is not in the range of the function. No, that it is not included
in the graph [points at the graph in Fig. 24a].
```

Fig. 24a and 24b: Reconstruction of figures n10 (left) and e6 (right) by respondent 266

```
R: Similarly here [points at the graph in Fig. 24b]. Here is also a point
which is not included in the graph.
```

Even though respondent 266 had a very good personal concept image of a Tall function (see Section 3.6.1), the presence of the point of discontinuity seems to be in the centre of his description and thus, his decision. However, it confused him and he reconstructed the graphs which do not represent functions at all (the graph on the right takes two different values at zero; both graphs have a point of discontinuity marked by an empty point which does not make sense). Based on his quotation, it is also questionable whether this aspect determined his decision more than the other aspects.

‘Resemblance to known examples and non-examples’

The respondents who underwent the image activity were categorizing some of the figures in the test according to their similarity with the examples and non-examples presented to them during the image activity. Consider, for example, respondent 249.

```
I: What did you focus on when you were deciding about the figures in the
test?
R: First, it was the resemblance for sure. When I saw the figure, I thought
of the figure from the beginning [the image activity], which was the most
similar one…
…
R: These two hooks [pointing to the reconstruction of figure e4 in Fig. 25]…
I remember them because one hook was in the counterexample at the beginning.
And here were two [in the test].
```

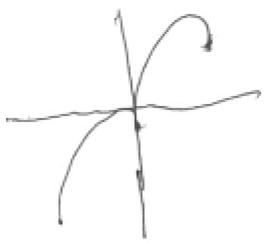


Fig. 25: Reconstruction of figure e4 by respondent 249

The respondent explicitly expressed that he had decided based on the resemblance to the known figures, too. However, it is obvious that the problem is in the understanding of ‘resemblance’. The respondents were looking for common properties of a previously unspecified character. No respondent described his/her concrete criteria of the resemblance but it is probable that they were based on some previously described aspects of graphs. In the case of respondent 249, it would probably be the resemblance with the non-example from the image activity (see Appendix A) based on aspect '*Shape of graph*'.[46]

3.6.3. Other aspects – misty personal concept image

A few respondents apparently had a misty personal concept image of a Tall function. They provided vague descriptions of their decision processes and features of graphs taken into account. For instance, respondent 235 described a Tall function as a combination of "sine and tangent rules", "belonging to three quadrants", "passing through zero" and "belonging to the positive and negative part of the coordinate system". Even if some of these features are intelligible and understandable (especially the last two), it is not clear how they had been combined when the respondent was deciding about the figures in the test.[47] Another problem is that the respondent had not generalized from the examples and non-examples seen during the image activity but only described observed features. Thus, no concrete aspect fits her decisions processes. Such an undeveloped personal concept image will be called *misty personal concept image* in the following sections.

Following the statements of respondent 235, it is probable that there are other aspects of the graphs whose nature was not expressed by the respondents and thus escaped my attention. Moreover, there is no exact way to characterize an aspect used for decision making and what sets it apart from the other decision-makers. Thus, the aspects described above represent those which were possible to interpret.

3.6.4. Types of aspects influencing the respondents' decisions

The aspects of graphs considered by respondents during the categorization were presented above; here we will summarise the ways respondents used them in their decision making process. We will distinguish aspects that are determining and non-determining, approving and disclaiming, and finally, problematic. This distinction will be found useful in the analysis of data in the following sections.

A respondent sometimes described *if* the aspect was used for his/her decision at the end. For instance, the statement "If there was a function and its bigger part was positive and it had all the border points included, then I decided it is…" represents the combination of aspects which determined the respondent's decision – the respondent described one of the possible cases and indicated that the result of his/her decision was unambiguous. On the other hand, the statement about figure n3 "here, it was something which confused me" shows that the respondent considered an aspect (even when it cannot be said which one) but was not able to express the way it influenced his/her decision.

The aspect can be labelled as determining the respondent's decisions if it was formulated in the form of a necessary condition such as, for instance, the claim of respondent 262: "I think that [a Tall function] is continuous, which means that it is not interrupted anywhere."

[46] This case is also interesting because the figure was categorized correctly as an example; however, using an incorrect aspect of 'resemblance'. Thus, it is not possible to reconstruct the process of the respondent's decision satisfyingly.

[47] The respondent demonstrated a very weak personal concept image and after the first activity, he was unable to reproduce any of the examples in the test correctly.

The distinction between *determining* and *non-determining* aspects should be understood as a scale (rather than dichotomy) where on the one side, there is a precise algorithmic decision which yields unambiguous results, and on the other, there are decisions based on intuition where an aspect mostly plays an unpredictable role. Moreover, if an aspect determines the decision of one respondent, it does not have to determine the decision of another – different respondents use aspects differently.

If a respondent described a determining aspect or their combination, it was often possible to distinguish *how* an aspect had influenced the respondent's decision. For instance, respondent 299 followed the aspect '*All points are above the x-axis*' when she stated "it is whole over the axis so it is… easy" and categorized the figure as an example. The presence of the aspect led the respondent to categorize the figure as an example. Such aspects will be called *approving*.

Disclaiming is such an aspect which leads the respondent to categorize the figure as a non-example. For instance, respondent 222 described one of the possible pathways to her decision through the aspect '*Sup is equal to |inf|*' by the words "when it was half-to-half, I decided it is not".

In some cases, a respondent's decision can be labelled as determining; however, he/she does not (and mostly is not able to) express its result. For instance, respondent 277 described her decision using the aspect '*Not well observable max or min*' by words "The difference was so small here that I was not sure" – the aspect itself and its use is clear but the result of the decision is also dependent on the respondent's perception of the graph. In other cases, a respondent considered aspects of graphs which were not related to his/her decision. The aspect '*Shape of graph*' represents this case – a respondent does not determine his/her decision following this aspect; however, the aspect itself is considered. When a respondent described the decision in one of these ways and the result of the decision was unclear, then the aspect was *problematic* for a respondent.

Particular aspects were often formulated in the form of necessary and/or sufficient conditions for both categorization of a figure as an example (representing approving aspects) and as a non-example (representing disclaiming aspects) of a Tall function.

3.6.5. Measurability of the aspects

The aspects can also be described in terms of their strength, that is, a respondent's ability to describe *how exact* their decision based on the aspect can be. It shows whether a respondent used unambiguous terms for the description of their decisions or rather approached their decisions intuitively.

This difference is well observable in the contrast between the aspects *'Longer part of the graph is above/below the x-axis'* and '*Sup is greater/lower than |inf|*' and also represents the difference between the description of typical Image group and Definition group respondents. While respondent 212 claimed that "if the bigger part of the graph is in the positive *y*-axis or negative *y*-axis", there is no clear way how to process the appropriate comparison without an additional specification (for instance, by measuring).[48] Another aspect which is hardly measurable by a respondent is the '*Resemblance to known examples and non-examples*'. None

[48] It is possible to find some exact ways to quantify the aspect, for instance, by measuring the length of the curve or area of the region bounded by a graph of a function, etc. However, none of the respondents specified a measure other than determining the value of maximum and minimum.

of the respondents described the way the resemblance of graphs were considered. On the other hand, for the aspect '*Sup is greater/lower than |inf|*' (or similarly, '*Max is greater than |min|*'), the respondents used concrete notions which they could mostly quantify if asked (find and compare precise values, decide exactly whether a condition is met or not, etc.).

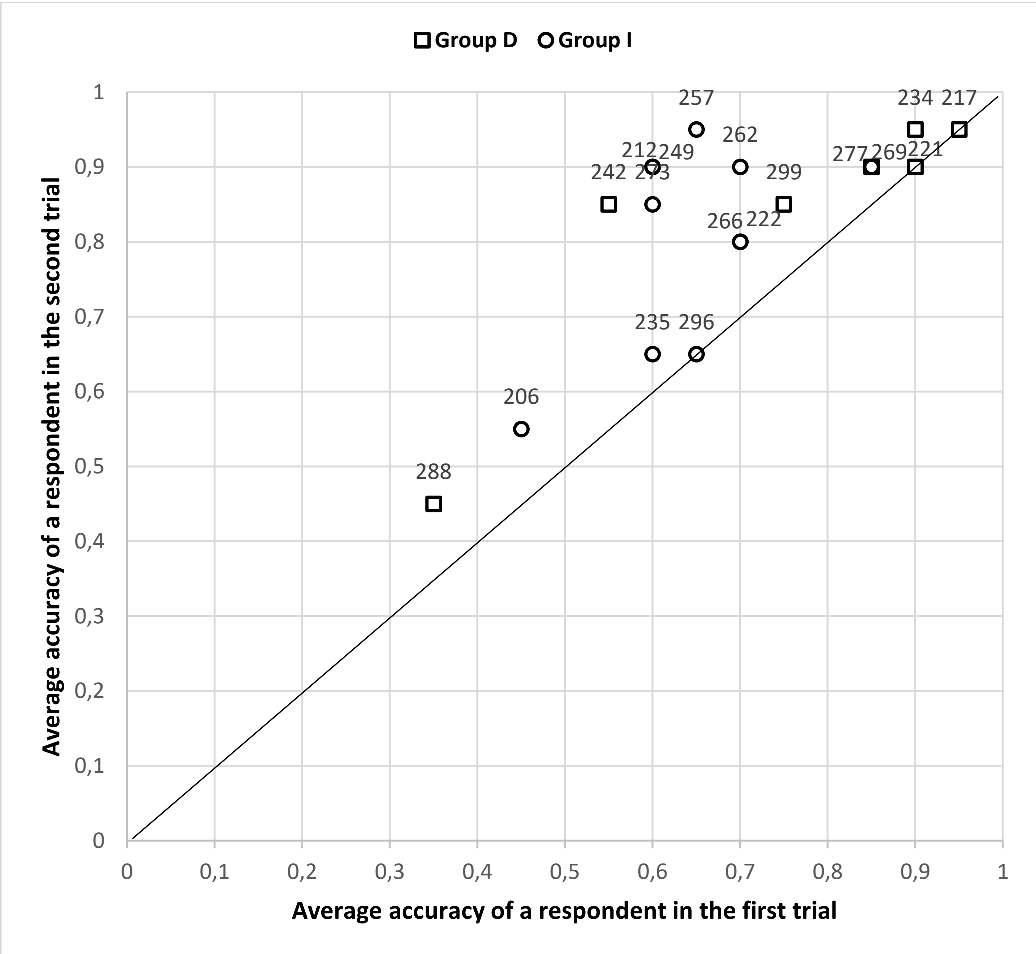
It seems that the measurability of an aspect and whether it is determining or not is the same; however, this is not so. The difference is that while a respondent may not be able to accurately express how exactly he/she decided using a given aspect, he/she can still use the aspect as a determining one (as is the case for aspects *'Longer part of the graph is above/below the x-axis'*).

3.7. Perspective of respondents

In this and the following section, both the quantitative and qualitative results presented above are combined and re-analysed from the point of view of respondents and figures. Three groups of respondents were identified based on the comparison of accuracies and of reaction times in both categorization tests. Unless stated otherwise, all interview samples are from the interviews after the first trial.

3.7.1. Respondents whose accuracy was poor and did not improve notably

Accuracies of respondents 206, 235, 288 and 296 were very low in the first trial and they were not improved much in the second one (by two correct answers at most, see Graph 15).



Graph 15: Comparison of average accuracies of respondents in both trials

Respondent 288

Respondent 288 is the only one from Group D with an average accuracy below the average of a random attempt. The respondent only used the concepts of maximum and minimum in the interview (probably because of their usage in the definition presented in the definition activity). The only aspect identified during the first interview when presenting the graph (see Fig. 26) as a non-example was '*Sup is equal to |inf|*'. However, her categorization of figures n1 and n3 as examples during the test is partly contradictory to this aspect. Thus, the respondent's personal concept image is hardly describable.

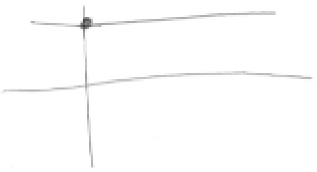


Fig. 26: Respondent 288's drawing of a non-example of a Tall function

Respondents 206, 235 and 296

The rest of the respondents with low accuracy which did not improve much between the trials belonged to the Image group. They share some characteristics. Respondents 206 and 235 are the only ones whose description of the concept and decision during the interview could be classified as *misty* (see Section 3.6.3) which means that they did not describe concrete aspects (only separate observations from the image activity were made with no tendency to generalize) and described their decisions about the figures vaguely. Moreover, both reconstructed almost no figures from the test and if so, these reconstructions differed considerably from the originals. Respondents 206 and 296 had the highest number of changes of their decisions between the trials (namely 10), respondent 235 changed her decision between the trials 5 times (the average in the whole data sample was 4.8).

All three respondents had consistent results from the perspective of reaction times. The geometrical mean of all their reaction times in the first trial was 2.72s, 3.32s and 2.86s for respondents 206, 235 and 296 – these respondents were among the fastest.

3.7.2. Respondents whose accuracies were good in both trials

The accuracy of respondents 217, 221, 234, 269 and 277 was already relatively high in the first trial – they had 17 (respondents 269 and 277), 18 (221 and 234) or 19 (217) correct answers out of 20.

Respondent 277

Respondent 277 was the only one from this group who belonged to the Image group. Her results differ from the rest in two other characteristics. First, as shown in Section 3.6.1, she was able to completely reconstruct the definition of a Tall function based on the image activity. And second, she is the only from the Image group whose average reaction time was considerably faster in the second trial[49] (see Graph 16 below).

[49] The average reaction times of respondents 273 and 249 were shorter in the second trial, but only in the range of tenths of a second.

Respondents 217, 221, 234 and 269

Respondent 221 is the only participant who did not express any doubt or hesitations from the beginning of the session. His description of the concept of a Tall function was precise, only two figures were categorized incorrectly – n8 and n9. His answers between the trials remained unchanged and reaction times ranged from 2.1s to 12.7s in the first trial and from 1.6s to 12.5s in the second one. He is the only representative of the four respondents whose personal concept image was sufficient in the first trial (18 correct answers – it is noteworthy, that even when respondent 217 had 19 correct answers, he showed some doubts and hesitations over certain aspects of the presented graphs).

The other respondents mentioned particular aspects which were problematic from their perspective. For instance, respondent 217 stated: "When it was unbounded, I had to think how I understood it, eventually, when it had more than one max or min, I had to think if it is taken as a maximum." Respondents 234 and 269 were also affected by the aspects '*More than one max or min*' and '*Max or min does not exist*' but mentioned the aspects '*Finding of max or min*' and '*Not well observable max or min*', too.

3.7.3. Respondents who markedly improved their accuracy

As shown in Graph 15, respondents 242, 273, 212, 249, 257 and 262 improved their accuracy in the second trial more than the others – by four correct answers or more. Such an improvement was most probably caused by realizing important issues connected to the concept of a Tall function – the respondents improved their personal concept images. A closer look at their decisions reveal some similarities among them.

Respondent 242

Respondent 242 was the only one from this group who belongs to the Definition group. His decisions were influenced by a number of aspects (for instance, '*Shape of graph*', '*Range of function*', '*Max or min does not exist*'), which led him to categorize the presented figures mostly as non-examples – 19 figures were categorized as non-examples (except figure e1) by him in the first trial. It was observed that the determining aspect was often '*More than one max or min*' when he thought that maximum and minimum did not exist if there were more than one points reaching the value. He mentioned it when he was describing his decision generally and when he was reconstructing figure e6 (see Fig. 26). However, he also categorized figures e3 and e4 as non-examples even if these figures represent functions which take the values of maximum and minimum only at one point. Thus, another uncovered aspect had been probably used by him as well.

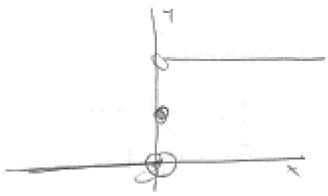


Fig. 26: Reconstruction of figure e6 by respondent 242

The respondent realized his misunderstanding about the aspect '*More than one max or min*' during the image activity (probably based on the second group of figures, see Appendix A). He confirmed this during the second interview. Another finding about him comes from the comparison of his reaction times. His accuracy improved and reaction times became faster in the second trial – it seems that when the 'internal discussion' about the questionable aspect was solved, other aspects were easy to be considered.

Respondents 212, 249, 257, 262 and 273

The following aspects were mentioned or observed during the interviews with the rest of the respondents from the group.

212 – *'Range of function'*, *'Open domain', 'Closed domain', 'Longer part of the graph above/below the x-axis', 'Resemblance to known examples and non-examples'*

249 – *'Longer part of the graph above/below the x-axis', 'Resemblance to known examples and non-examples'*

257 – *'Open domain', 'Closed domain', 'Range of function'*, *'Longer part of the graph above/below the x-axis'*

262 – '*Principle of discontinuity', '*Range of function'*

273 – *'Open domain', 'Closed domain', 'Range of function', 'Principle of discontinuity', 'Resemblance to known examples and non-examples'*

All the respondents in question were able to describe one or more aspects which they used during their decisions and/or for their description of the concept. Each focused on a slightly different combination of these aspects. However, the description of the concept was rather intuitive, the respondents did not express their decisions precisely and this fact was sometimes stressed by them. Consider, for instance, respondent 249.

```
I: How would you define a concept of a Janda function?
R: To define it so that I can say it is [a Janda function] for one hundred percent… It's all based on some assumption that I see a resemblance there… Because I see the pictures [from the image activity] in front of me… comparing those pictures.
```

Some respondents were able to construct their own examples and non-examples but were not able to describe a determining aspect of their decisions. An example is respondent 257 whose examples and non-examples are shown in Fig. 27. Her only statement was that a Tall function has to be "rather positive".

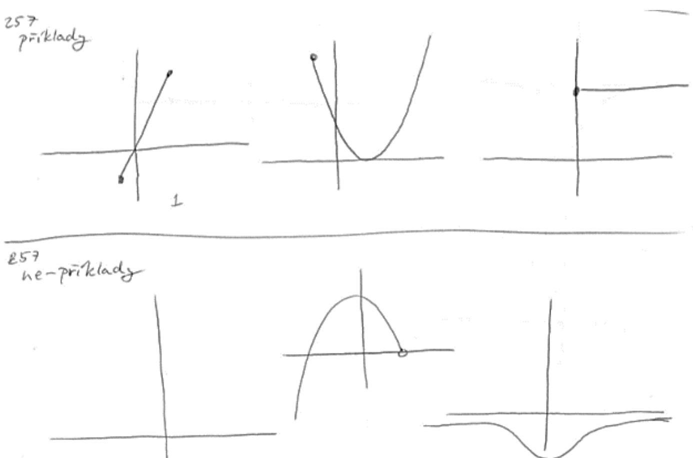
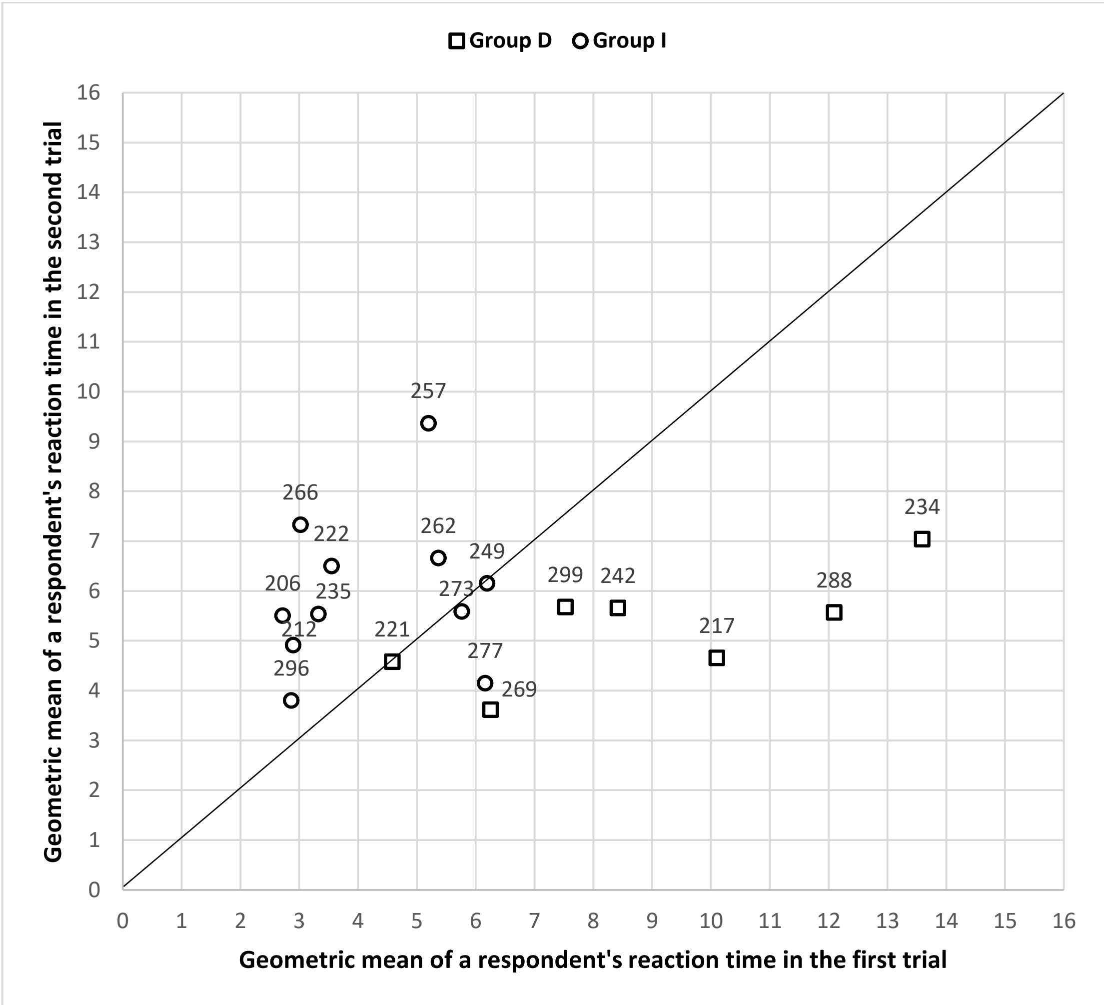


Fig. 27: Examples (top) and non-examples (bottom) provided by respondent 257

Another unifying characteristic of all the respondents in the group except respondent 212 was that their average reaction times in the first trial were very similar (from 5.19s to 6.15s, while for respondent 212, it was 2.89s).

3.7.4. Considering the three groups of respondents in connection with their reaction times

The above categorization of the respondents can also be seen from the perspective of their reaction times. Graph 16 depicts some other findings.



Graph 16: Comparison of geometrical means of respondents' reaction times in the first and second trial

The respondents of the Image group with weak or misty personal concept image and a low ability to explain their decision process who did not improve their results between the trials much (206, 235, and 296) had average reaction times from 2s to 4s. As can be seen from the graph, there are other respondents with a similar average reaction time (212, 222 and 266). The average accuracy of these respondents was 0.6 (respondent 212) and 0.7 (respondents 222 and 266) in the first trial and improved by 6 and 2 correct answers, respectively, in the second trial.

The rest of the respondents of the image group (249, 262, 273 and 257) had considerably higher average reaction time than the previous ones (from 5s to 7s). All these respondents were also in the category of those who improved their cross-trial decisions substantially – for four or more correct answers.

Finally, respondents 217, 221, 234, 269 and 277 had high average accuracy in both trials and were also substantially faster in the second trial in comparison with their first trial (except respondent 221). Two respondents are worthy of special attention here. The first is respondent 221 who had a precise personal concept image of a Tall function even in the first trial and his average reaction time almost did not change. The second is respondent 277 who is the only respondent of the Image group who fully reconstructed the definition of a Tall function and whose average reaction time ranked among other respondents from the Definition group.

3.8. Perspective of figures

In this section, the quantitative and qualitative results will be combined from the point of view of figures. Reaction times corresponding to particular figures were used to identify two groups of figures. Again, all interview samples are from the interviews after the first trial.

Table 6 summarizes general results of reaction times of particular figures. Columns *min RT* and *max RT* describe a spectrum of respondents' reaction times (as can be seen, the longest times vary a lot from 7.55s for e2 to 59.97s for n10 which is not surprising considering that some figures are harder to categorize than others). The rest of the values in the table (in columns 3, 4, 5 and 8, 9, 10) represent an *average rank of reaction times* of all the respondents together and Group I and D separately. The average rank of reaction times is a construct which objectivizes the comparison of reaction times between all the respondents. All reaction times of a respondent are ordered from the fastest to the slowest, each figure is labelled by a number from 1 to 20 (which means that smaller numbers represent a generally faster categorization of the figure). Thus, the average rank of reaction times is an arithmetic mean of all the respondents' ranks in the group (average rank 10 represents the middle value). The grey shading in the table emphasizes examples and non-examples as candidates for further examination which is presented below.

Table 6: Minimal and maximal reaction times and average ranks of particular groups of respondents

Min and max RT and Average relative rank										
	Trial 1					Trial 2				
	Reaction time		Average rank of Group:			Reaction time		Average rank of Group:		
	min	max	I+D	I	D	min	max	I+D	I	D
e1	1.37	19.77	8.1	8.6	7.1	1.65	11.46	7.5	7.5	7.6
e2	1.72	7.55	6.4	7.8	4.3	2.80	19.10	9.1	10.2	7.3
e3	1.82	30.96	11.7	11.5	12.0	3.12	13.92	12.9	11.6	14.9
e4	1.96	55.62	12.1	10.4	14.9	1.31	48.03	14.3	14.3	14.4
e5	1.28	36.53	9.4	8.8	10.3	1.65	8.25	8.4	9.5	6.7
e6	2.07	18.38	13.2	13.5	12.7	1.46	17.78	13.2	14.5	11.0
e7	2.14	24.53	12.7	11.8	14.1	2.38	18.49	12.1	12.8	10.9
e8	1.79	30.47	12.6	12.9	12.1	3.97	12.21	12.7	12.8	12.6
e9	0.86	40.22	9.8	8.8	11.4	2.75	10.58	11.0	10.0	12.6

e10	1.53	20.95	10.7	10.4	11.1	2.33	14.49	10.4	8.3	13.9
n1	2.08	19.00	9.4	9.5	9.3	0.93	8.11	5.7	5.5	6.0
n2	1.35	26.29	10.6	10.6	10.6	1.95	18.46	10.4	11.1	9.4
n3	2.42	24.48	10.1	10.5	9.4	2.04	9.68	7.0	6.9	7.1
n4	1.87	20.52	9.4	9.4	9.4	1.26	14.81	8.9	7.9	10.4
n5	1.79	11.95	8.2	9.9	5.4	1.20	12.17	9.2	10.7	6.9
n6	2.17	13.32	10.2	11.9	7.6	2.23	14.30	10.1	10.5	9.3
n7	1.51	28.96	12.3	12.4	12.1	1.46	21.41	9.4	7.9	11.7
n8	2.45	23.74	12.6	13.4	11.3	1.83	22.57	10.7	10.6	10.9
n9	1.73	53.30	9.4	6.6	13.9	2.73	17.28	13.9	14.2	13.6
n10	2.52	59.97	11.1	11.2	10.9	2.26	31.27	13.1	13.2	13.0

3.8.1. Figures for which the reaction times of Group D respondents were the lowest in the first trial

As can be seen from Table 6, the average ranks of Group D respondents for figures e1, e2, n5 and n6 (see Fig. 28) are the lowest from all. These figures represent graphs where the values of maximum and minimum are clearly observable and the values are easily comparable (the difference of values is obvious). Moreover, functions take the values of maximum and minimum only at one point (except figure e2), they are continuous and with closed domains and ranges. Thus, it is probable that the decision process of a respondent should not be loaded by problematic aspects much. Whereas the reaction times of the Definition group respondents were very low in this case, results of the Image group respondents are not so extreme.

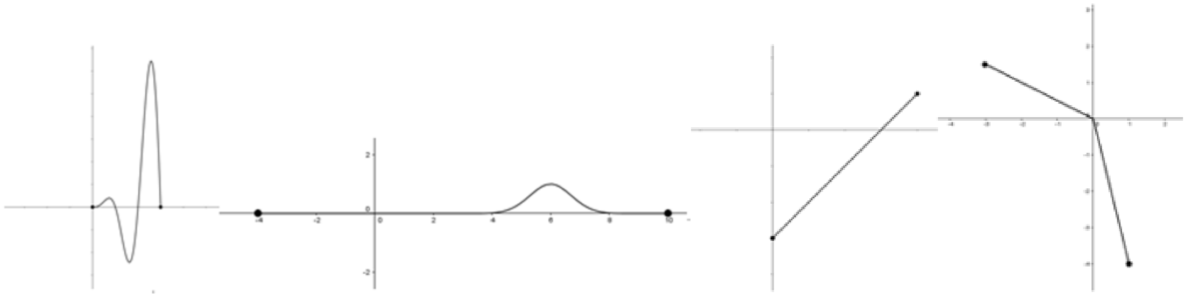


Fig. 28: Figures e1, e2, n5 and n6

3.8.2. Figures n9 and e4: Different decision making processes

The difference between Groups I and D in terms of reaction times in the first trial is the highest in the case of figure n9 (see Fig. 29). Whereas the respondents of the Image group had generally the fastest reaction time for this figure (the geometrical mean was 2.93s, the average rank of the respondents' reaction time was 6.6), the respondents of the Definition group were deciding much slower (the geometrical mean was 13.60s with the average rank of the respondents' reaction times 13.9). The comparison of accuracies (see Table 5 in Section 3.5.6) shows that almost all the respondents marked this figure as a non-example in the first trial (except respondent 273) but the comparison of reaction times indicates that the way of this decision varies a lot between the respondents of both groups.

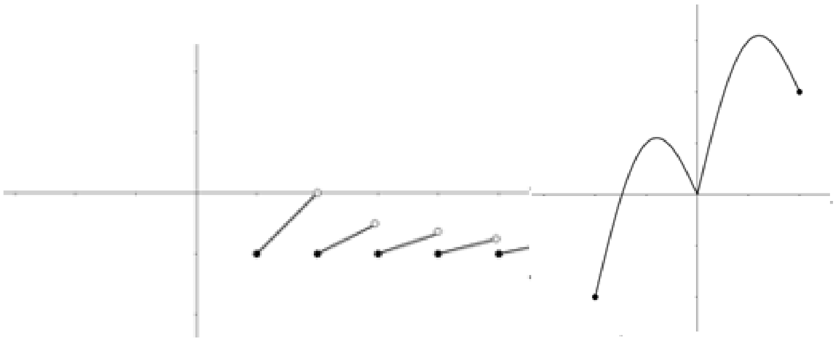


Fig. 29: Figures n9 and e4

Figure n9 is complicated in comparison with the other figures as many aspects play a role here (for instance, '*Sup is smaller than |inf|', 'Maximum or minimum does not exist', 'Principle of discontinuity, 'More than one max or min'*). Moreover, the graph visually differs from the graphs used during the image activity significantly (see Appendix A) and its domain is questionable in the positive part of the real axis because it is not clear how the graph continues for greater values of x-axis.

The analysis of the second trial showed much slower reactions of the Image group respondents when categorizing figure n9 (geometrical mean: 7.55s, average rank of the respondents' reaction times: 14.2) and slightly faster reactions from the respondents of Group D (geometrical mean: 6.07s, average rank of the respondent's reaction time: 13.6). In the second trial, the figure was categorized in terms of reaction times similarly to the respondents of the Definition group in the first trial (the average rank is generally higher for both groups) – the process of decision making is unified among the respondents.

For figure e4, the respondents of the Image group had the average rank of 10.4 vs. 14.9 of the Definition group respondents. The difference is not as high as in the case of n9, however, the results provide reasons for discussion, too. The important aspect for this figure was '*Finding of max or min'* while respondents finding concrete values of maximum and minimum sometimes compared the border points of the domain first while the respondents whose decisions were not based on this aspect did not have to solve this issue.

Both figures whose results were described above seem to be promising candidates which can help to distinguish different processes of decisions in terms of reaction time measurement.

Chapter 4

Discussion and conclusions

In the theoretical part of the thesis, three perspectives of a concept in an individual's mind were discussed. From the static concept representation, we moved to the developmental perspective called concept formation and to the perspective of immediate operation called concept processing. Here, when discussing the results of the Main study, this trajectory will be reverted. Concept processing is what was directly observed and from this perspective some conclusions are made about concept formation and concept representation.

4.1. Processing of a Tall function concept (RQ1, RQ2)

4.1.1. Differences between the Image group and the Definition group

Some differences between the two groups were observed in the Main study. The respondents who started with the image activity (learned about a Tall function through examples and non-examples) had generally faster reaction times on average in the first trial (geometrical means were in intervals from 2s to 7s) and had approximately ten percent smaller average accuracy. Respondents who started with the definition activity (learned about a Tall function through the definition) had generally slower reactions (geometrical means from 4s to 14s) and their decisions were more accurate. Thus, respondents who learned in different ways also decided differently when categorizing objects.

Except for the differentiation discussed above, it is possible to split the respondents of the Image group into two groups with faster responses (from 2s to 4s in average) and slower responses (from 5s to 7s) in the first trial. This difference is also connected to the accuracy in both trials in which the 'faster Group I respondents' were not able to improve their personal concept image based on the definition activity as successfully as the 'slower Group I respondents'. This finding will be discussed further.

A clear difference between the respondents building their knowledge on examples and non-examples and those building knowledge from the definition was found for a complicated graph of function (n9) which combines several problematic aspects but for which the negative decision was easy to make. Both groups also differed in reactions to figure e4 where the finding of the maximum could cause problems.

Following these findings, it is possible to say that the difference in decisions of both groups is not negligible. Moreover, there are differences in decisions even within the Image group itself, that is, the same learning experience does not necessarily lead to the same results in terms of categorization. These findings support the main idea presented in Section 1.5 and Section 2.1.

Specifically, that some respondents prefer intuition and some prefer a conscious deductive approach to decision making (Kahneman, 2011), some of them are logicians and some of them are intuitionists (Poincaré, 1952) and some prefer reasoning from their personal concept image and some from their personal concept definition (Vinner & Hershkowitz, 1980; Tall & Vinner, 1981).

4.1.2. Aspects of graphs and their role in the decision process

The qualitative analysis showed that most respondents of the Image group grounded their decision in their considerations about concrete aspects of presented graphs. Six respondents of this group were able to express particular aspects they considered for their decision. Their decisions were based on what Alcock and Simpson (2002) call *property abstraction*. Three respondents of Group I did not reach this level of their concept image – during the interview, they were not able to describe concrete aspects which determined their decision.

The last two respondents of Group I extracted something that can be understood as a definition based on the image activity and made decisions (with some possible exceptions) using this definition. Personal concept images of these respondents fit most to the form of concept of a Tall function as represented by its mathematical definition and also as presented to respondents of the Definition group. The two respondents of Group I and most respondents of Group D decided in most cases using their definition (Alcock and Simpson call this decision *working from definition*). However, in some cases, exceptions were identified and will be interpreted below.

4.1.3. Determinability and measurability of aspects and their role in the decision process

Even when it is possible to distinguish between two different kinds of decisions among respondents as described above, the interviews with respondents revealed that many of them do not follow them in all cases. A typical example is respondent 266 who was able to derive the formal definition based on the image activity only; however, he incorrectly categorized six figures in the first and four in the second trial. Respondent 266 did not decide in accordance with his personal concept definition. This finding is also reported by other authors (Tall & Vinner, 1981; Vinner & Dreyfus, 1989).

Those respondents who had not yet formed their own personal concept definition (they were deciding using the identified aspects of the figures or even more intuitively) also did not usually decide following any exact criteria. This can be inferred from the interviews which showed that it played a role in to what extent the identified aspects were measurable or to what extent they determined a respondent's decision. Some of the respondents of the Image group expressed that they knew that their image of a concept in question fits some of the figures presented in the image activity, only. However, because they had no better tool for their decision, they still followed this image when deciding even when they were aware that it does not fulfil features of a mathematical concept and an appropriate category.

The determinability and measurability of observed aspects naturally provide a potential source of respondents' incorrect responses. Many respondents also used some kind of optimization in their decision (for instance, from the definition, they drew the claim "if there is an open range, it is not a Tall function" or "if there is more than one maximum or minimum, it is not a Tall function"). Incorrect implications are also potential sources of incorrect responses. Another potential source of errors was intentionally included in the design of the test which contained graphs where some properties were hardly observable without additional context. It is important to note here that these three phenomena have a different place in the decision process. Problematically observable properties of graphs are connected to perception and interpretation of a figure. Implications from a respondent's personal concept image are based on his or her mathematical knowledge connected to this concrete concept. And the determinability and measurability of aspects are rather dependent on a respondent's cognitive structure, specifically on how he/she works with mathematical concepts generally.

The above findings are in accordance with some cognitive models of decision making and reasoning (Sternberg & Sternberg, 2012). Thus, it makes sense to suppose that these aspects can play a role in similar categorizations of other mathematical concepts but also in other tasks where the definition describes a category of mathematical objects. Moreover, it appears that personal concept images of respondents are appropriate to general cognitive models of concept representations (Sternberg & Sternberg, 2012) for (most) respondents of the Image group and to the model of concept image and concept definition for (most) respondents of the Definition group.

Another point of view is how a category behind a concept of a Tall function is seen by respondents. Some illustrations of how close to the classical meaning (in the sense used by Alcock and Simpson, 2002, and Lakoff, 1987) of a category the respondents were are presented in the following section.

4.1.4. Representation of a Tall function category

In the perspective of theories of categorization, access to a category of Tall functions of most Group I respondents is close to how category representations are described by Wittgenstein (1958), Lakoff (1987) and Rosch (1973). However, a central role of a concrete prototype was not observed in the presented study. I believe that this is due to the fact that the process of concept formation was relatively short, the number of examples the respondents met was relatively small and also, from the perspective of a respondent dealing with the tasks in the study, it was not necessary to target and take over a specific figure as a prototype. However, more research is needed to develop this assumption further.

On the other hand, the respondents of the Definition group approached categorization more in a classical way (Alcock & Simpson, 2002) and their image of a Tall function category was determined by a definition and it was sharper (see Section 1.3.1) – they identified and tended to use measurable and determining aspects of the graph for their decisions. Some exceptions were observed, however, which seem to support these findings. An example is respondent 277 of the Image group who developed a precise definition based on the image activity and following that, her decision process became very similar to the decision processes of Group D respondents.

4.1.5. Differences in categorization of examples and non-examples (RQ6)

The results of the second Pilot study showed that there is a difference in reaction times of categorization of examples and non-examples. Further discussion of these results can be found in (Pilous & Janda, 2017) where they were explained by the complexity of the two actions. To confirm that the graph of a function fullfils the definition of a Tall function, several conditions must be met concurrently, while to refute, only one unmet condition suffices. Thus, the former action takes longer.

A similar, albeit smaller, difference was observed in the Main study, too. The examples were processed by respondents slower than non-examples on average in both trials, and also in the test focused on the concept of injective function (RQ4, RQ5). However, the sample size does not allow us to provide any statistically based conclusions.

Based on these results, I hypothesize that the difference is in the structure of such a decision. When one presumes that the figure does not fit the definition, it suffices to find one aspect only which will show it. On the other hand, when one presumes that he/she is dealing with an example, it is more complex to prove it. Nevertheless, we should be aware that this difference is dependent on the complexity of tasks in question. The complexity of these tasks for concrete figures should also be considered.

If this hypothesis holds for more mathematical concepts, there are some implications. For instance, the difference between reaction times for examples and non-examples would be a possible indicator of the level of a respondent's personal concept image.

4.1.6. Convergence of the reaction times

Another difference between the two groups can be observed when the results of both trials are compared. It was shown that respondents of the Image group tended to have slower reactions in the second trial in comparison with their first trial, while respondents of the definition group tended to have faster reactions. This is also a potential source of differentiation between the respondents who decided rather intuitively than the ones who were more precise and used a definition.

Moreover, in the second trial, reaction times of both groups were all closer together and also moved closer to their average (which was 5.39s in the first and 5.53s in the second trial). This finding corresponds to the observed difference in the categorization of examples and non-examples and is also a potential indicator of the stage of a respondent's personal concept image formation.

Sometimes, a respondent comes to the stage of thought (conscious or unconscious) that his or her concept image is developed and his or her decisions are stable. It was mostly the members of the Definition group who had better accuracy in both tests. This can best be seen from respondent 221's work. He categorized 18 from 20 figures correctly in both trials and did not change any of his decisions across the trials. Moreover, his average reaction time was almost the same in both trials. A unifying factor was observed in the responses of respondents with high accuracy in that their reaction times were very similar (around 6s on average, with one exception).

The case of respondent 221 shows that in some cases, the concept image can remain stabilized – decisions about the figures are not changing and would probably be the same if the task was repeated. The respondent was not confronted with any conflicting property (Duffin & Simpson, 1993) and, thus, his concept image remained unchanged. Reaction times also do not differ much in this case.

4.1.7. Comparison of categorization different concepts (RQ4, RQ5)

The simple categorization test based on the injective function concept was added into the Main study for comparison with the results of a Tall function categorization. Common patterns in categorization were identified by the qualitative analysis. The most promising one consists of the fact that the figures categorized faster were categorized rather accurately while the accuracy of figures categorized slower varies more. This finding (with the theoretical common pattern of the categorization of examples and non-examples described above) represents a possible source of information about the respondents' categorization. However, it is not possible to formulate definite conclusions about two different approaches to categorization (using concept image or concept definition) due to a small sample.

4.2. Formation of a Tall function concept (RQ3)

Following the above, what can be said about a Tall function concept formation in an individual's mind? The respondents of both groups were able to utilize the definition and examples and non-examples for the improvement of their personal concept image. Examples and non-examples fulfilled their role as elaborated by Watson and Mason (2002). They were shown to serve as tools for clarifications of details from the definition of a concept (respondents of the Definition group were able to improve their personal concept image based on examples and non-examples presented). To some respondents, they served as tools for inductive reasoning (they were able to reconstruct the definition of a concept from them).

Thus, it seems clear that even when the final personal concept images of respondents are very similar (and in most cases, they were), how examples and non-examples were used and processed by respondents of both groups were different.

From the perspective of theories of concept formation, a process-concept paradigm (Sfard, 1989; Dubinsky, 1991; Hejný, 2012) was observed as marginal since the concept of a Tall function only delimits a sub-category of general functions which were relatively well-known by the respondents of the Main study. Moreover, functions were understood almost exclusively as mathematical objects in all studies. On the other hand, some phenomena described above showed the importance of conflicts (and also an absence of conflicts) during concept formation (Tall & Vinner, 1981; Duffin & Simpson, 1993).[50]

4.3. Representation of a Tall function concept

The results of the Main study showed that when respondents learn about a concept in different ways, their responses in a categorization task will differ from the perspective of reaction time. However, the question is what we can infer from these results about a concept representation in respondents' minds. Specifically, if there are some differences in personal concept images of respondents other than their approach to categorization. Moreover, it is desirable to identify more widely applicable differences (having a global character, Alcock & Simpson, 2017) – which would be less dependent on the context (the types of concepts and tasks used).

Three tentative conclusions can be made. First, it was observed how the new concept integrates into an existing cognitive structure of a respondent. The quality of this integration is individual – some respondents were able to develop their own definitions, some respondents were not able

[50] No conclusions about the value of particular interpretative frameworks should be made based on these statements. The reader should be aware that in a different context (with different respondents or concepts), suitable interpretative frameworks would probably differ, too.

to construct anything that could be considered as an image of a concept, etc. This points to the importance of existing structure for an additional learning process.

Second, respondents identify different aspects of a concept and interpret them in different ways. Thus, an individual's category behind the concept has different characteristics where those developed based on the definition fit properties of classical categories better.

Finally, it seems to be clear that if one or the other approach to the learning of mathematical concepts prevails, it will influence the nature of an individual's learning and, thus, theoretically also an individual's resulting personal concept image. This has a practical implication; we should be cautious not to over-prioritize concrete approaches until their effects are sufficiently explored.

4.4. Consequences for teaching and learning

4.4.1. Teacher's influence on a student's personal concept image

One implication of the above results is the realization that a teacher can influence all three cognitive components in connection to a personal concept image – the concept's representation, formation and processing. My research has tentatively shown that the order of image and definition activity caused little or no difference in the final personal concept image of respondents. Both were more or less useful for respondents (while the definition activity was seen as more useful).

Based on my results, it is reasonable to assume that a teacher's emphasis on work with examples and non-examples may have consequences not only for students' specific concept images but also for the way they learn about similar concepts. It was shown in a specific case of a Tall function concept that both learning approaches, examples and non-examples and the definition, can be useful and, more importantly, both can be useless in some cases. However, some importance lies in the ways these approaches influence the general learning process. This remains to be seen.

4.4.2. Precise knowledge of definition does not guarantee correct responses

It was shown that the respondents of the Definition group and also the respondents from the Image group who were able to develop their own definition had better accuracy in the categorization test. However, in accordance with literature and common teacher experience, it does not guarantee correct responses (for instance, Vinner & Dreyfus, 1989). Nevertheless, as was shown in the Theoretical part, the explanatory power of definitions is sometimes overrated and one has to be aware that even when a definition is presented and explained, it does not guarantee that students will be aware of all critical aspects of a concept.

4.5. Methodological findings

The simple categorization test was designed and validated as a research tool for the examination of representation, formation and processing of mathematical concepts. Pilot study 1 showed several issues which can affect the results of the test: 'Blackout effect', 'Previous example effect', 'Effect of awareness of measuring reaction times' and 'Effect of the first figure(s)' (see sections 2.5.3 and 2.6.5). Moreover, it was also noted that the amount of instruction presented to the respondents should be carefully considered in order to achieve the optimum amount of information from the categorization test.

The results of Pilot study 2 revealed an enormous influence of the ‘Effect of measuring reaction times’ on the data obtained. Therefore, the recommendation to exclude this information from the instruction for respondents was made. It is also possible to reduce the influence of some of the above effects (for instance, ‘Effect of the first figure(s)’) by using the categorization test within a one-on-one session between the interviewer and the respondent as in the Main study.

These findings can inform any future research on this topic using similar methodological tools. For example, during the analysis of the data, I recommend not excluding long reaction time outliers (based on my experience and the notes of Ratcliff, 1993) where the short ones should be excluded in accordance with the literature (Thorpe et al., 1996). Several ways of analysing data from categorization tests were presented in (Pilous & Janda, 2017) and the sections presenting results of the second Pilot study (2.6.4) and the Main study (3.5, 3.6, 3.7 and 3.8).

4.6. Limitations of research and further research

From the theoretical perspective, it is worth noting a distinction between three perspectives of an image of a concept in an individual’s mind: concept representation, concept formation and concept processing. All these perspectives were grounded in existing literature. Concept representation was grounded in the theory of concept image and concept definition and theories of categorization. Concept formation was grounded in the theories based on the process-object paradigm (such as the APOS theory) and the theories of historical development of mathematical concepts (Kvasz, 2008). Concept processing was grounded in dual process theories and theories of categorization. Thus, one of the limitations of the study might consist of the selection of particular theories for the interpretation of results. However, these were selected with the purpose of taking into account the perspectives of mathematics education, cognitive psychology and history of mathematics as domains with a deep influence on the topic of the thesis. The results of my research were interpreted within this framework. The use of the three perspectives represents one of the strengths of this research.

The number of respondents participating in the Main study represents the second limitation. However, the exploratory nature of the presented studies required a thorough examination of the respondents’ decision-making process. A higher number of respondents is appropriate as a further step in examining categorization using the proposed methodology as it would allow for the statistical demonstration or refutation of some formulated hypotheses.

The use of respondents’ reaction times in the simple categorization test as an additional source of information on the categorization of mathematical objects can also be a potentially limiting factor. For this reason, and because of the lower number of respondents, I avoided conclusions that would be based on a purely quantitative interpretation of the data obtained. I consider the combination of qualitative and quantitative approach to be another strength of this work.

Another limitation lies in examining only one, newly created, concept of a Tall function, with only additional results from the test focused on one more concept (injective function). Nevertheless, the newly developed concept was important in order to make the input knowledge of the respondents more uniform. Another possible investigation in the field is thus the verification of the results obtained for other mathematical concepts.

Finally, the use of a newly-developed methodological tool – a simple categorization test – represented a challenge. After the first Pilot study, various limitations of the test were described (see the previous section), some of them seemingly critical. However, methodological recommendations were made to reduce or eliminate these negative effects and, hopefully, they could contribute to the harmonization of a possible methodology in case of further investigation in this field.

4.7. Conclusions

To sum up, the focus of the thesis was to examine which information can be abstracted from a relatively simple process, such as the categorization of mathematical objects and what conclusions can be drawn from them about an individual's personal concept image. The studies presented are mostly exploratory; thus, the description of the process of categorization is far from complete. A predominantly qualitative focus of the Main study proved to be reasonable and valid and, in my opinion, provides a good basis for further investigation in this area.

Based on the studies carried out and the analysis of the results obtained, evidence has been presented that different approaches to categorizing mathematical objects can be distinguished and that categorization as such can be a useful tool to describe various properties of mathematical concept images in learners' minds. With this approach, I attempted to link up with other studies that point to the important role of categorization in the domain of mathematics education, whether used as a theoretical framework or as a research tool.

References

Alcock, L. (2001). *Categories, definitions and mathematics: Student reasoning about objects in Analysis* (Doctoral dissertation, University of Warwick).

Alcock, L., & Simpson, A. (2002). Definitions: Dealing with categories mathematically. *For the Learning of Mathematics*, *22*(2), 28–34.

Alcock, L., & Simpson, A. (2017). Interactions between defining, explaining and classifying: the case of increasing and decreasing sequences. *Educational Studies in Mathematics*, *94*(1), 5–19.

Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics*, *55*(1), 11–27.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Babai, R., Levyadun, T., Stavy, R., & Tirosh, D. (2006). Intuitive rules in science and mathematics: A reaction time study. *International Journal of Mathematical Education in Science and Technology*, *37*(8), 913–924.

Brown, J. R. (1998). What is a definition?. *Foundations of Science*, *3*(1), 111–132.

Dubinsky E. (1991). Constructive aspects of reflective abstraction in advanced mathematics. In L. P. Steffe (Ed.), *Epistemological foundations of mathematical experience*, (pp. 160–202). New York: Springer.

Dubinsky, E., & McDonald, M. A. (2001). APOS: A constructivist theory of learning in undergraduate mathematics education research. In D. Holton, M. Artigue, U. Kirchgräber, J. Hillel, M. Niss, & A. Schoenfeld (Eds.), *The teaching and learning of mathematics at university level* (pp. 275–282). Dordrecht: Springer.

Duffin, J. M., & Simpson, A. P. (1993). Natural, conflicting and alien. *Journal of Mathematical Behavior*, *12*(4), 313–28.

Duffin, J. M., & Simpson, A. P. (2000). A search for understanding. *The Journal of Mathematical Behavior*, *18*(4), 415–427.

Edwards, J. C. H. (1979). *The historical development of the calculus*. New York: Springer-Verlag.

Fischbein, E. (1993). The theory of figural concepts. *Educational Studies in Mathematics*, *24*(2), 139–162.

Freudenthal, H. (1986). *Didactical phenomenology of mathematical structures* (Vol. 1). Dordrecht: Springer Science & Business Media.

Goldenberg, P., & Mason, J. (2008). Shedding light on and with example spaces. *Educational Studies in Mathematics*, *69*(2), 183–194.

Hejný, M. (2012). Exploring the cognitive dimension of teaching mathematics through scheme-oriented approach to education. *Orbis scholae*, *6*(2), 41–55.

Hershkowitz, R. (1989). Visualization in geometry – two sides of the coin. *Focus on Learning Problems in Mathematics*, *11*, 61–76.

Iannone, P., Inglis, M., Mejía-Ramos, J. P., Simpson, A., & Weber, K. (2011). Does generating examples aid proof production?. *Educational Studies in Mathematics*, *77*(1), 1–14.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kleiner, I. (1989). Evolution of the function concept: A brief survey. *The College Mathematics Journal*, *20*(4), 282–300.

Kruschke, J. K. (2005). Category learning. In K. Lamberts, & R. L. Goldstone (Eds.), *The handbook of cognition* (pp. 183–201). London, England: Sage.

Kurtz, K. J. (2015). Human category learning: Toward a broader explanatory account. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 63, pp. 77–114). Amsterdam: Elsevier Academic Press.

Kvasz, L. (2008). *Patterns of change: linguistic innovations in the development of classical mathematics*. Basel: Birkhäuser.

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. Cambridge: Cambridge University Press.

Lakoff, G. (1987), *Women, fire and dangerous things – what categories reveal about the mind,* Chicago: University of Chicago Press.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Mason, J., & Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics*, *15*(3), 277–289.

MacHale, D. (1980). The predictability of counterexamples. *The American Mathematical Monthly*, *87*(9), 752–752.

Michener, E. R. (1978). Understanding understanding mathematics. *Cognitive Science*, *2*(4), 361–383.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, *27*(3), 249–266.

Needham, T. (1998). *Visual complex analysis*. Oxford: Oxford University Press.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.

Nunez, R., & Lakoff, G. (1998). What did Weierstrass really define? The cognitive structure of natural and ε-δ continuity. *Mathematical Cognition*, *4*(2), 85–101.

Piaget, J., Grize, J. B., Szeminska, A., & Bang, V. (1977). *Epistemology and psychology of functions* (Vol. 23). Dordrecht: Springer Science & Business Media.

Pilous, D., & Janda, D. (2017). Reaction time of simple categorization of subclasses of functions. In D. Szarkova et al. (eds.), *16th Conference on Applied Mathematics, APLIMAT 2017 – Proceedings* (pp. 1189–1196). Bratislava: Institute of Mathematics and Physics, Faculty of Mechanical Engineering.

Poincaré, H. (1952). *Science et methode: Science and method*. New York: Dover Pubs.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350.

Rosch, E. (1999). Principles of categorization. In E. Margolis, & S. Laurence (Eds.), *Concepts: Core readings* (pp. 189–206). Cambridge, MA: MIT Press.

Sfard, A. (1989). Transition from operational to structural conception: The notion of function revisited. In G. Vergnaud et al. (Eds.), *Proceedings of PME XIII* (pp. 151–158). Paris: International Group for the Psychology of Mathematics Education.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*(1), 1–36.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*(3), 278–304.

Sternberg, R. J., & Sternberg, K. (2012). *Cognitive psychology*. Wadsworth: Cengage Learning.

Stillwell, J. (2010). *Mathematics and its history*. New York: Springer Science & Business Media.

Tall, D. (2001). Natural and formal infinities. *Educational Studies in Mathematics*, *48*(2–3), 199–238.

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, *12*(2), 151–169.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *NATURE*, *381*(6), 520–522.

Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2012). Naturally biased? In search for reaction time evidence for a natural number bias in adults. *The Journal of Mathematical Behavior*, *31*(3), 344–355.

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*(6), 655–668.

Vinner, S. (1983). Concept definition, concept image and the notion of function. *International Journal of Mathematical Education in Science and Technology*, *14*(3), 293–305.

Vinner, S. (2002). The role of definitions in the teaching and learning of mathematics. In Tall D. (Ed.), *Advanced mathematical thinking* (pp. 65–81). Dordrecht: Springer.

Vinner, S., & Dreyfus, T. (1989). Images and definitions for the concept of function. *Journal for Research in Mathematics Education*, *20*(4), 356–366.

Vinner, S., & Hershkowitz, R. (1980). Concept images and common cognitive paths in the development of some simple geometrical concepts. In R. Karplus (Ed.), *Proceedings of the 4th International Conference for the Psychology of Mathematics Education* (pp. 177–184). Berkley: University of California

Watson, A., & Mason, J. (2002). Student-generated examples in the learning of mathematics. *Canadian Journal of Math, Science & Technology Education*, *2*(2), 237–249.

Watson, A., & Mason, J. (2006). *Mathematics as a constructive activity: Learners generating examples*. London: Routledge.

Wittgenstein, L. (1958). *Philosophical investigations*. Oxford: Basil Blackwell Ltd.

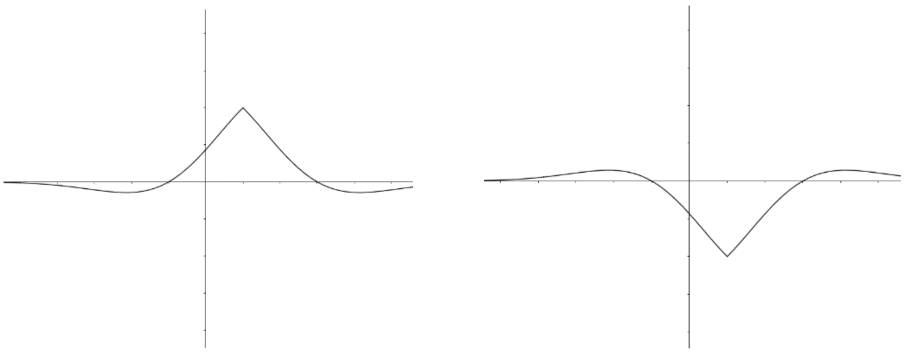
Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.

Zazkis, R., & Leikin, R. (2008). Exemplifying definitions: A case of a square. *Educational Studies in Mathematics*, *69*(2), 131–148.

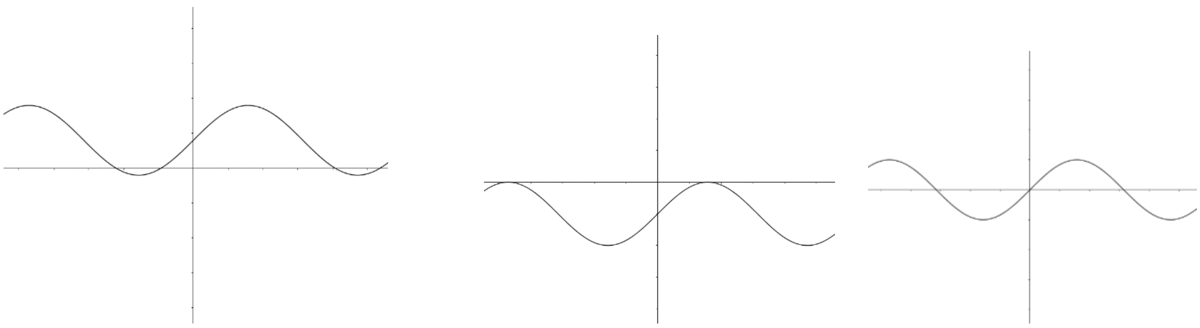
Appendix A

Figures used in the image activity of the Main study

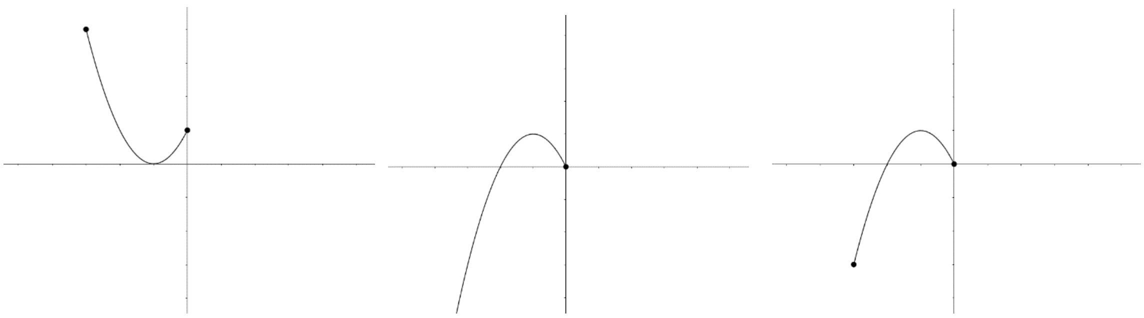
G1 (example and non-example):



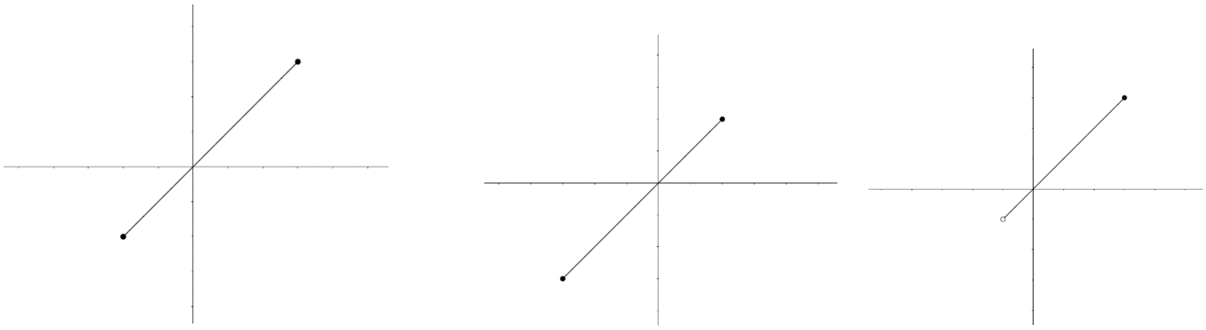
G2 (example and 2 non-examples):



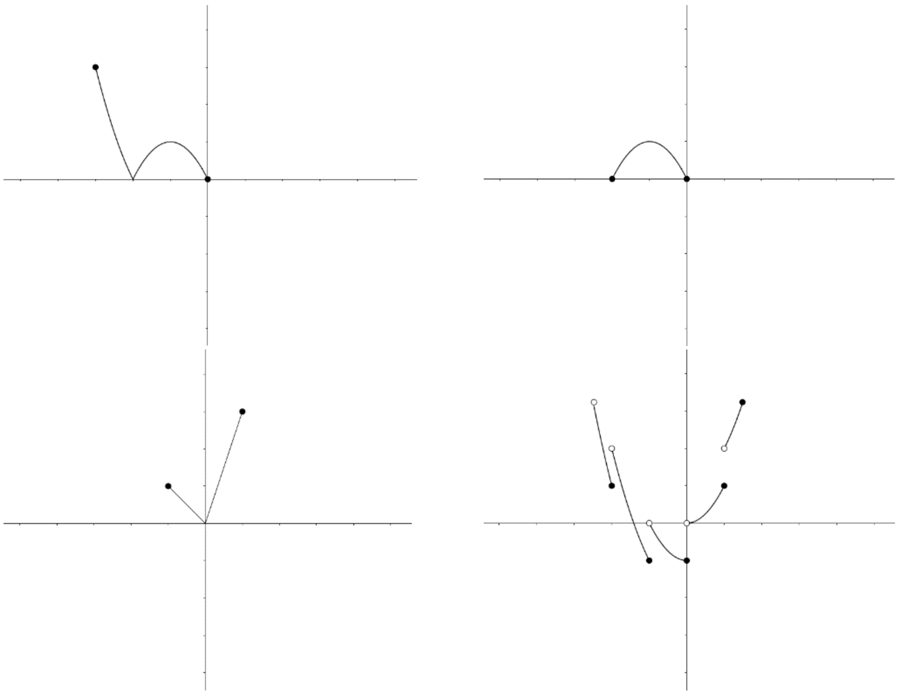
G3 (example and 2 non-examples):



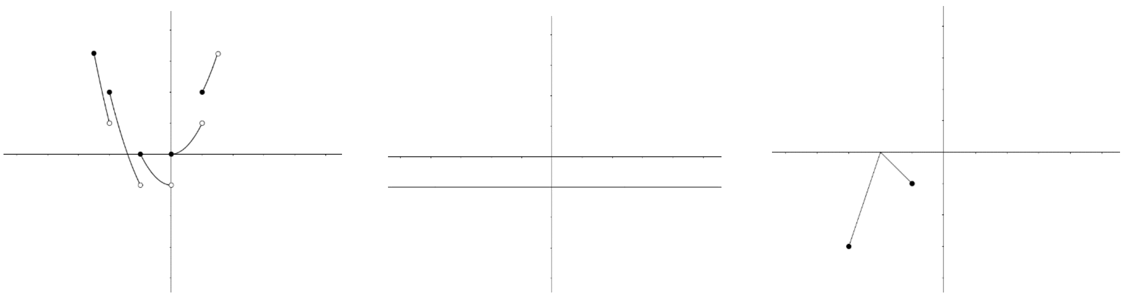
G4 (example and 2 non-examples):



G5 (examples excluded from the image activity after the pilot interviews):



G6 (non-examples excluded from the image activity after the pilot interviews):



Appendix B

Figures used in the injective function simple categorization test

Examples:

Non-examples: