THE THREE WORLDS AND TWO SIDES OF MATHEMATICS AND A VISUAL CONSTRUCTION FOR A CONTINUOUS NOWHERE DIFFERENTIABLE FUNCTION

Juha Oikkonen

Department of Mathematics and Statistics

University of Helsinki

Abstract

We present a visual construction of a nowhere differentiable function in connection to a discussion on mathematical thinking and especially the three worlds of mathematics of David Tall. A special feature of the construction is that the main properties of the function (continuity and nowhere differentiability) can be proved by discussing properties of pictures used to illustrate the definition.

0 Introduction

Continuous nowhere differentiable functions have an important role in the development of mathematics in the 19'th century. Nowadays a great variety of constructions leading to such functions are known. See e.g. Thim [2003]. Being extremely counterintuitive such functions and their existence present also an interesting challenge for learning of the basic concepts of analysis and in mathematical thinking in general. For example, David Tall has been using a construction which he calls the Blancmange function in many of his writings beginning from Tall [1982].

This paper consists of two kinds of material. First we present a construction for a continuous nowhere differentiable function. The construction and the way in which we present it does not appear in e.g. Thim [2003], though many constructions of continuous nowhere differential functions have common features.

We shall actually present the construction twice: in section 1 by means of a visual approach and in section using more formal style.

Later we use this construction and the two ways to approach it as an example for discussing mathematical thinking. We shall be especially interested in the connection between our own ideas of the human (social) and objective sides of mathematics [Oikkonen "Oulu"] and the famous three worlds of mathematics of David Tall.

1 The function f

We shall give in this Section a construction of a continuous nowhere differentiable function by visual means. The construction of our continuous nowhere differentiable function f and the discussion of its properties are written below so that the presentation suits for a group of students in a university course of analysis. Especially it is assumed that the students know in advance the basic properties of the real line including completeness and 'epsilon-delta'-definitions for continuity and differentiability.

What is done below can, however, be easily modified for an audience with no deep familiarity with continuity and derivatives (like a group of upper secondary school students) much in the same way as David Tall describes in many of his papers and for example in the book Tall [2013]. Actually the author has presented a construction of a Peano-curve in a similar way to audiences consisting of upper secondary school students.

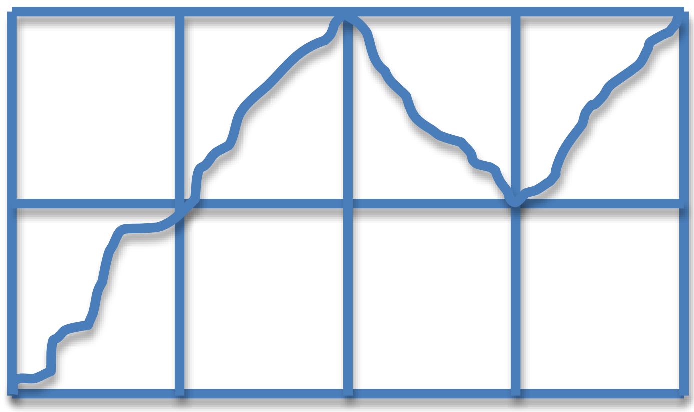
We assume below some familiarity with nested closed intervals. If this would be new to the audience we should spend some time on describing the basic properties of nested open and closed intervals and on the completeness ('there are no holes') of the real line. To be more precise, if $[a_1, b_1]$, $[a_2, b_2]$, … are closed intervals and if $a_1 \leq a_2 \leq$ … and $b_1 \geq b_2 \geq$ ..., then there are real numbers which belong to all of these intervals. If moreover the lengths $b_n - a_n$ tend to 0 as n tends to infinity, then there is a unique common element in these intervals.

This property is actually very interesting for several reasons. It is rather obvious when we 'look at' the real line. Hence it is very close to our 'visual image' of the real line. It is also rather easy to prove this property in an introductory course in analysis. Moreover, this property is a nice version of the compactness of closed intervals and it can be used to give a uniform way of proving the main consequences of compactness in an analysis course by 'cutting closed intervals in halves'.

We define our function f first on [0, 1] so that $f(0) = 0$, $f(1) = 1$, and $0 \leq f(x) \leq 1$ will hold for all x in [0, 1]. After this we shall 'copy' the original f on every [n, n+1] so that $f(n) = n$ and $f(n+1) = n+1$. For example, for $1 \leq x \leq 2$, we put $f(x) = f(x - 1) + 1$.

The process of defining f on [0, 1] goes in steps. The initial step is described above. At every step we get more accurate information about the value f(x). This information tells us that f(x) lies in a certain closed interval $[a_n(x), b_n(x)]$. (We accept here a singleton as a closed interval.) Finally, f(x) will be the unique common element of all these intervals.

The main idea is presented in this simple picture. There is a big rectangle cut to smaller dividing it horizontally in four pieces and vertically in two.



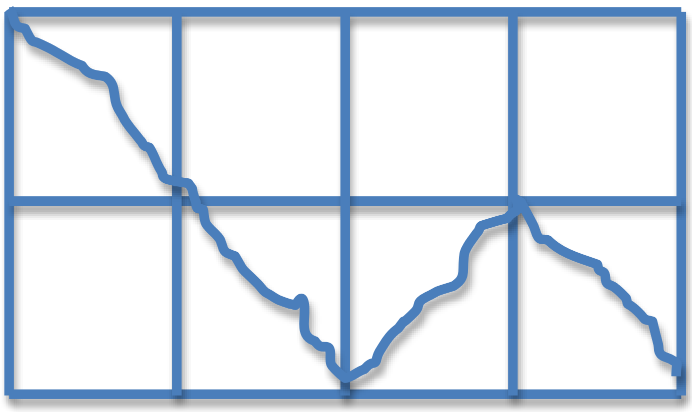
(COMMENT: THESE SKETCHES COULD PERHAPS BE REPLACED BY PHOTOS OF ACTUAL HAND DRAWN PICTURES.)

Initially, the big rectangle is the square [0,1] x [0,1]. This means that $a_0(0) = b_0(0)$, $a_0(1) = b_0(1) = 1$ and for $0 < x < 1$, $a_0(x) = 0$ and $b_0(x) = 1$.

To go to the next step, the big rectangle is cut as in the picture. The 'graph' of the function f is meant in the picture to give an impression of what the function shall roughly look like. The picture is of course 'wrong' in the sense that it does not carry much exact information about f. Indeed, it only indicates in which smaller rectangles the graph of the function f appears. But we can use the picture in order to get some idea of our construction. It does help us to visualize the function we are constructing without being a portrait of the function. At least the author found the sketch very helpful.

The values of $a_1(x)$ and $b_1(x)$ are indicated in the picture above. More precisely, we have for instance for values $\frac{1}{4} < x < \frac{1}{2}$ the values $a_1(x) = \frac{1}{2}$ and $b_1(x) = 1$. At the end points of this subinterval we have $a_1(\frac{1}{4}) = b_1(\frac{1}{4}) = \frac{1}{2}$ and $a_1(\frac{1}{2}) = b_1(\frac{1}{2}) = 1$. Especially, the graph of the function f will go through the corners as indicated.

This process is repeated infinitely many times. In cases where the function "goes from the upper left corner to the lower right corner" (which is above the case on the subinterval [½ , ¾]), the picture is used "upside down" like this.



Pictures of this kind can be used also for indicating proofs for the continuity and nowhere differentiability of f – or at least for indicating the thinking behind the formal proofs.

To do this, some notation will help. Notice that at each step n of the construction we use rectangles with certain width w_n and height h_n. Indeed,

$w_1 = 1$ and $w_{n+1} = ¼\, w_n$;

$h_1 = 1$ and $h_{n+1} = ½\, h_n$.

Especially, the form of these rectangles is characterized by the ratio

$h_{n+1} / w_{n+1} = 2^n$.

The first immediate consequence of the construction is that whenever

$|x - t| < w_n$,

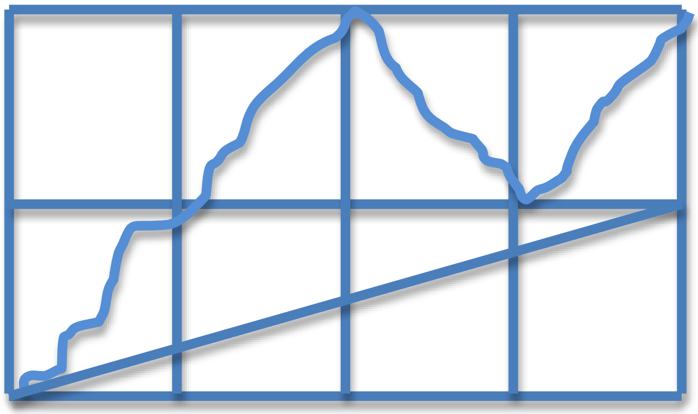
the points (x, f(x)) and (t, f(t)) of the graph of f must lie in the same or consecutive rectangles. (If we were discussing the pictures, It would be natural to show with ones finger the points discussed)

Thus

$|f(x) – f(t)| < 2\, h_n$.

It follows from this observation that f is uniformly continuous.

To prove the nowhere differentiability of f we take a new look at the picture



(The picture is of course too wide here, but it is meant to express the idea.)

Let us show that f is not differentiable at a certain point x_0. We shall consider the difference quotients

$(f(x) – f(x_0)) / (x – x_0)$

for certain other values x.

In every stage n of the construction, we can locate x_0 in a picture like this. We can assume that f 'goes' from the bottom left corner to the top right corner. (The other case where f 'goes' from the top left corner to the bottom right corner is quite similar.)

Assume first that x_0 is 'in' the rightmost quarter. Let the other value x in the difference quotient correspond to the left bottom corner. For geometric reasons we see that the absolute value

$|(f(x) – f(x_0)) / (x – x_0)|$

is at least the slope for the rising line drawn in the picture. Thus

$|(f(x) – f(x_0)) / (x – x_0)| \geq ½\ h_n / w_n$.

But this ratio can be made as big as we like by choosing n big enough! Notice that if x_0 is 'in' any other part of the picture, we have the same estimate. (If x_0 is 'in' the leftmost part, then we take x to 'correspond to' the top right corner of the picture.)

This observation gives us the following result: For every x_0, every $\varepsilon > 0$ and every M > 0, there is x for which $|x – x_0| < \varepsilon$ and

$|(f(x) – f(x_0)) / (x – x_0)| > M$.

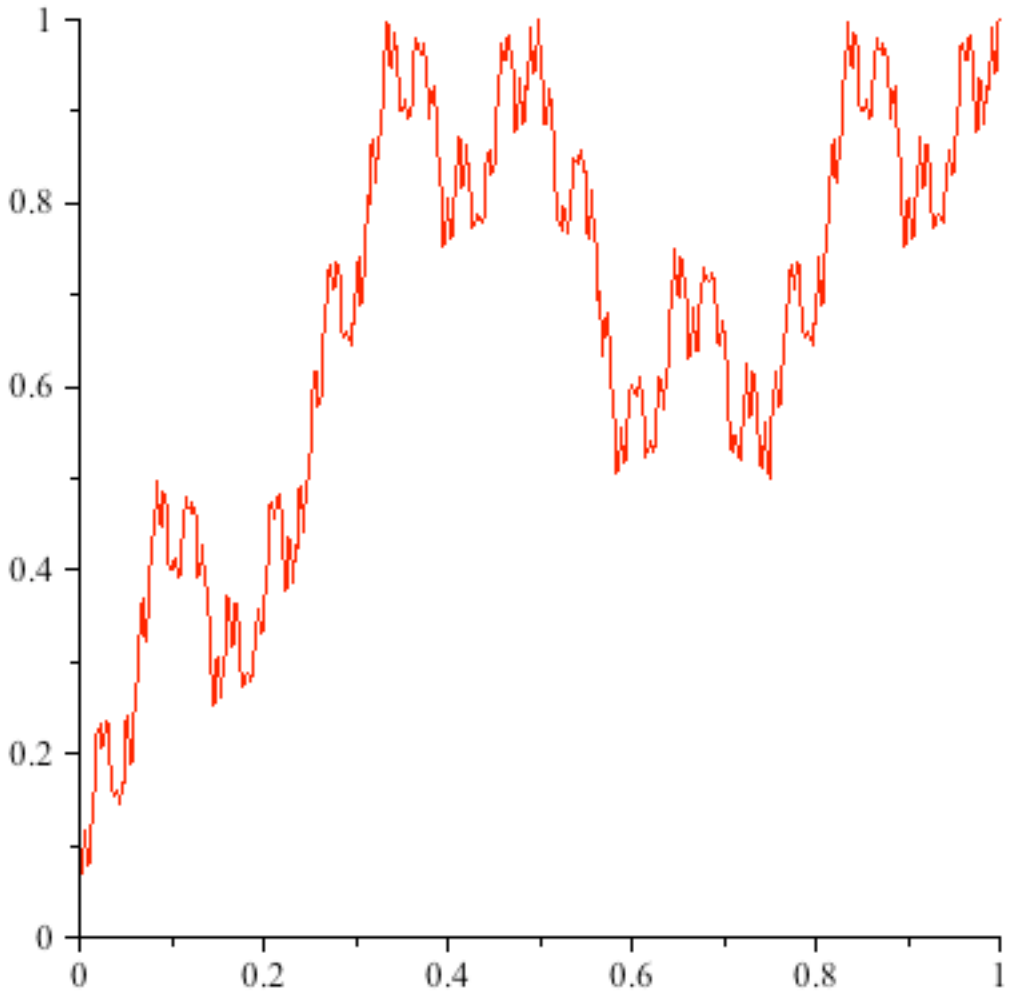
Especially, f is nowhere differentiable since the difference quotient corresponding to any x_0 cannot have a limit.

In this construction and in the arguments above the author likes especially the feature that all the thinking is completely visual (or embodied in the pictures, as will be said later in this paper). It would be wonderful to present this at a backboard!

More exactly, the visual proof consists of the above pictures and a discussion while observing the pictures. This will suffice to convince most novices and experts. In case we would like to write a formalized proof, we could use such a discussion as a receipt. We shall come to this theme in section 3 of this paper.

Since f is continuous and nowhere differentiable, also that the function $g(x) = f(x) - x$ is continuous and nowhere differentiable. This function has the additional property that $g(0) = g(1)$. So extending g on the whole real line is especially simple: just put $g(x) = g(x - n)$ when $n \leq x < n + 1$.

Finally, here is a 'realistic' picture of the function f.



It is produced by Maple using a code kindly written for us by Antti Rasila (in 2007.)

2 Variants of the function

To understand the function defined above we only needed the 'boxes'. There was no reference to any actual approximating functions living in the boxes.

But at each step n of the construction we can in many ways consider a function f_n satisfying the requirements given in terms of the 'boxes'. This can be done in many ways.

It follows easily from the construction that every choice of such functions f_n leads to a sequence of functions uniformly converging to f.

If we choose the functions to be piecewise linear, the situation is in this respect very much like with the blancmange function discussed by Tall in [1982].

We can choose also the functions f_n to be continuously differentiable by using eg., suitable trigonometric expressions. For example,

$$c+\frac{d-c}{2}(1-\cos(\frac{\pi(x-a)}{b-a}))$$

goes from the lower left corner to the upper right corner of [a, b] x [c, d] and has derivative = 0 at the end points.

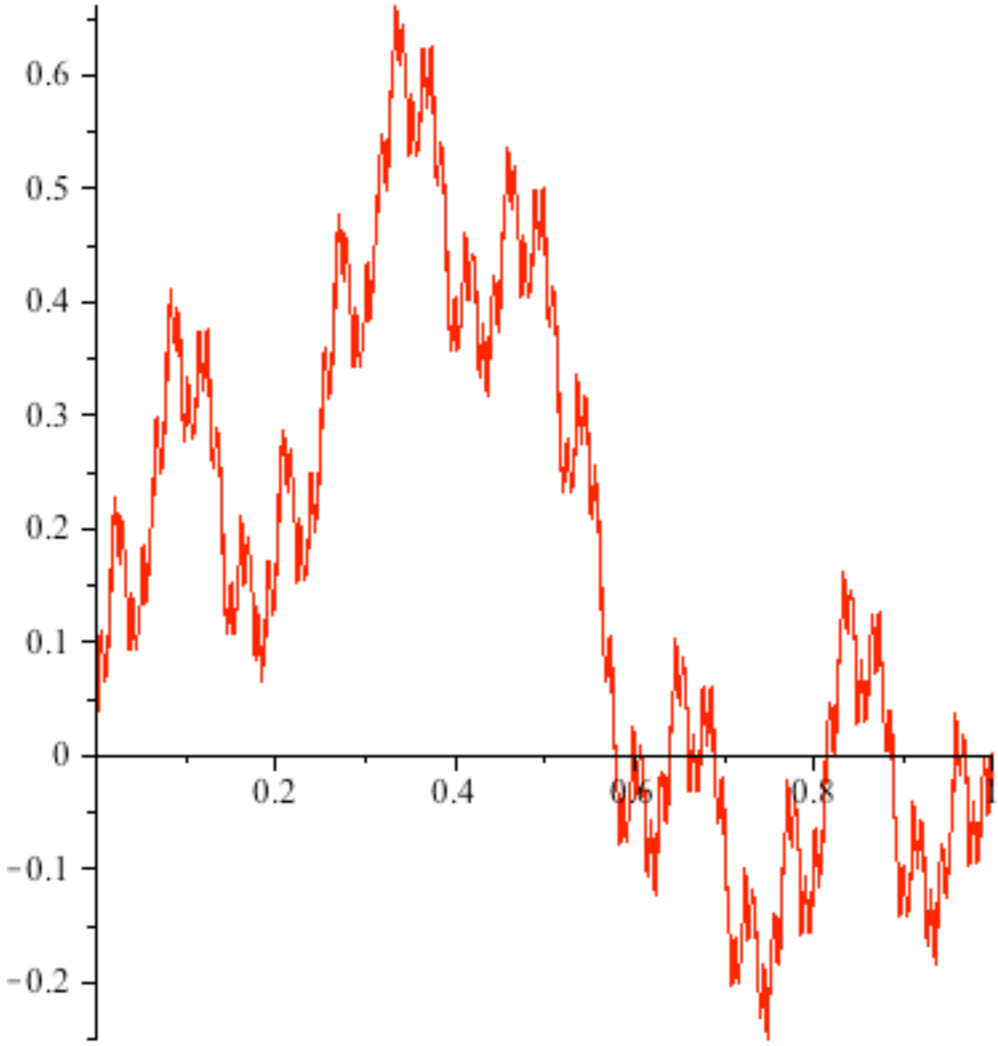
By use of well-known tricks in analysis, the functions f_n can also be made C^{∞}. So our construction has a version which has properties much like the well-known trigonometric sums of the form

$$\sum_{n=1}^{\infty} a_n \sin(b_n x)$$

as in Weierstrass' original construction.

Another interesting point about the function f compared to the blancmange function or the trigonometric series above is that our definition in terms of 'boxes' shows from the beginning of the iteration information about the values of the function: we go to smaller and smaller details inside previous rectangles. This information simply gets more and more accurate. But at every step we are able actually to draw some kind of a picture of the function f by hand. The pictures of the blancmange function and the trigonometric series seem to be more difficult to anticipate.

The function g(x) = f(x) – x mentioned before has also some interest since it satisfies g(n) = 0 for every integer n. Here is a picture of g produced by modifying the code yielding the Maple code that was used in the earlier 'realistic' picture.



There are also other ways to extend f from [0, 1] to the whole real line so that we have a periodic continuous nowhere differentiable function. Here is an example. We denote this other function by f, too.

Define first for x in [0, 1], f*(x) = f(x). Define next for x in]1, 2], f*(x) = f(2-x). Thus we have defined f* on the interval [0, 2] so that f(0) = f(2) = 0 and the graph is symmetric relative to the line x = 1. After this we define for all n and x in [2n,2(n+1)] the value f*(x) by f*(x) = f*(x-2n). Notice that here x – 2n lies in [0, 2].

3 A 'mathematical' version of the construction

Our aim in this section is to describe how the construction presented in the previous section can be given a more exact form.

We use the compactness of closed intervals in the form of the following well-known fact.

Lemma. If $[a_1, b_1], [a_2, b_2], \ldots$ are closed intervals and if $a_1 \leq a_2 \leq \ldots$ and $b_1 \geq b_2 \geq \ldots$, then there are real numbers which belong to all these intervals; and if the length $b_n - a_n$ tends to 0 as n tends to infinity, then there is a unique common element in these intervals.

In the construction of our function we split at every step the square [0, 1] x [0, 1] to smaller and smaller rectangles. This can be expressed in terms of two sequences $(\mathcal{I}_n)$ and $(\mathcal{J}_n)$ of sets of closed intervals so that the rectangles discussed on step n of our construction will be of the form I x J where I is in $\mathcal{I}_n$ and J is in $\mathcal{J}_n$.

We shall need one piece of notation. Let I = [a,b] be a closed interval. Denote

$I_1 = [a, a + ¼ (b - a)]$,

$I_2 = [a + ¼ (b - a), a + ½ (b - a)]$,

$I_3 = [a + ½ (b - a), a + ¾ (b - a)]$,

and

$I_4 = [a + ¾ (b - a), b]$.

Similarly we denote for a closed interval J = [c, d]

$J^1 = [c, c + ½ (d - c)]$,

and

$J^2 = [c + ½ (d - c), d]$.

Let $\mathcal{I}_0$ be {[0, 1]} and $\mathcal{J}_0$ be {[0, 1]} and assume that $\mathcal{I}_n$ and $\mathcal{J}_n$ have been defined. Then $\mathcal{I}_{n+1}$ consists of all intervals of one of the forms I_1, I_2, I_3 and I_4 where the interval I is in $\mathcal{I}_n$. Similarly, $\mathcal{J}_{n+1}$ consists of all intervals of one of the forms J^1 and J^2 where the interval J is in $\mathcal{J}_n$.

Next we shall express the construction described in section 1 in terms of the present notation. We have to define $a_n(x) \leq b_n(x)$ for all x in [0, 1]. Then we shall define f(x) to be the unique common element of all the sets $[a_n(x), b_n(x)]$, i.e.

$f(x) = \lim_{n\to\infty} a_n(x) = \lim_{n\to\infty} b_n(x)$.

As in section 1, we begin with

$a_0(x) = 0$ when $0 \leq x < 1$ and $a_0(1) = 1$.

Similarly

$b_0(x) = 1$ when $0 < x \leq 1$ and $b_0(0) = 0$.

Thus the set

$\{y \mid a_0(x) \leq y \leq b_0(x)\}$

is a singleton if x is 0 or 1 and it is a closed interval if $0 < x < 1$.

Next we shall assume that $a_n(x)$ and $b_n(x)$ have been defined so that for every I = [a, b] in $\mathcal{I}_n$ there is such a J = [c, d] in $\mathcal{J}_n$ that for all x in I both $a_n(x)$ and $b_n(x)$ lie in J, and that (i) or (ii) below hold.

(i) $a_n(x) = c$ when $a \leq x < b$ and $a_n(b) = d$

and

$b_n(x) = d$ when $a < x \leq b$ and $b_n(a) = c$.

(ii) $a_n(x) = c$ when $a < x \leq b$ and $a_n(b) = d$

and

$b_n(x) = d$ when $a \leq x < b$ and $b_n(a) = c$.

The interpretation of (i) is that f shall 'go' from the left bottom corner of I x J to the right top corner of I x J. Likewise (ii) means that f shall 'go' from the left top corner of I x J to the right bottom corner of I x J.

We shall discuss case (i); case (ii) is analogous. The intervals I_1, I_2, I_3 and I_4 and their end-points will be considered separately.

$a_{n+1}(a) = b_{n+1}(a) = c$,

if $a < x < a + ¼ (b - a)$, we put $a_{n+1}(x) = c$ and $b_{n+1}(x) = c + ½ (d - c)$,

$a_{n+1}(a + ¼ (b - a)) = b_{n+1}(a + ¼ (b - a)) = c + ½ (d - c)$,

if $a + ¼ (b - a) < x < a + ½ (b - a)$, we put $a_{n+1}(x) = c + ½ (d - c)$, and $b_{n+1}(x) = d$,

$a_{n+1}(a + ½ (b - a)) = b_{n+1}(a + ½ (b - a)) = d$,

if $a + ½ (b - a) < x < a + ¾ (b - a)$, we put $a_{n+1}(x) = c + ½ (d - c)$, and $b_{n+1}(x) = d$,

$a_{n+1}(a + ¾ (b - a)) = b_{n+1}(a + ¾ (b - a)) = c + ½ (d - c)$, and

if $a + ¾ (b - a) < x < b$, we put $a_{n+1}(x) = c + ½ (d - c)$, and $b_{n+1}(x) = d$.

Assume that $a_{n+1}(x)$ and $b_{n+1}(x)$ are defined in this way. One can easily check that for every I = [a', b'] in $\mathcal{I}_{n+1}$ there is such a J = [c', d'] in $\mathcal{J}_{n+1}$ that for all x in I both $a_{n+1}(x)$ and $b_{n+1}(x)$ lie in J, and that (i) or (ii) below hold.

(i) $a_{n+1}(x) = c$ when $a' \leq x < b'$ and $a_{n+1}(b) = d'$

and

$b_{n+1}(x) = d'$ when $a' < x \leq b'$ and $b_{n+1}(a') = c'$.

(ii) $a_{n+1}(x) = c'$ when $a' < x \leq b'$ and $a_{n+1}(b') = d'$

and

$b_{n+1}(x) = d'$ when $a' \leq x < b'$ and $b_{n+1}(a') = c'$.

This means that the recursive construction shall go on. It follows that we obtain a well defined function f: [0, 1] -> [0, 1] since the intersection of the descending sequence of sets $[a_n(x), b_n(x)]$, n = 0, 1, 2, ... is a singleton. Next we shall take a new look at how to prove that f is continuous and nowhere differentiable.

It follows directly from the definition of the sequences $(\mathcal{I}_n)$ and $(\mathcal{J}_n)$ of families of sets that the length of every I in $\mathcal{I}_n$ is 4^{-n} and that of every J in $\mathcal{J}_n$ is 2^{-n}.

In our construction we use rectangles I x J where I is in $\mathcal{I}_n$ and J is in $\mathcal{J}_n$. The width of such is thus 4^{-n} and the height of such is 2^{-n}. The ratio of height and width of such a rectangle is

$2^{-n} / 4^{-n} = 2^n$.

Proposition: (i) If x and y are in [0, 1] and $|x - y| < 4^{-n-1}$, then $|f(x) – f(y)| < 2^{-n}$.

(ii) The function f is continuous.

Indeed, if $|x - y| < 4^{-n-1}$, then x and y lie in the same or in adjacent members of the family $\mathcal{I}_{n+1}$. Thus the assertion follows directly from the above construction.

It follows from (i) that f is (uniformly) continuous on [0, 1].

Proposition: The function f is nowhere differentiable.

The same rectangles as above can be used to show that f cannot be differentiable anywhere. Consider a point x_0 in [0, 1] and the difference quotients

$f(x) – f(x_0) / x – x_0$.

Fix n for a moment. There is I in $\mathcal{I}_n$ for which x_0 lies in I = [a, b]. Consider the following four cases separately.

(1) x_0 is in I_1. Choose x = a + ½ (b - a). Then $|x – x_0| < ½\, 4^{-n}$ and

$|f(x) – f(x_0) / x – x_0| \geq 2^n$.

(2) x_0 is in I_2. Choose x = a. Then $|x – x_0| < ½\, 4^{-n}$ and

$|f(x) – f(x_0) / x – x_0| \geq 2^n$.

(3) x_0 is in I_3. Choose x = b. Then $|x – x_0| < ½\, 4^{-n}$ and

$|f(x) – f(x_0) / x – x_0| \geq 2^n$.

(4) x_0 is in I_4. Choose x = a. Then $|x – x_0| < 4^{-n}$ and

$|f(x) – f(x_0) / x – x_0| \geq ½\, 2^n$.

In every case there is x satisfying $|x – x_0| < 4^{-n}$ and

$|f(x) – f(x_0) / x – x_0| \geq ½\, 2^n$.

This holds for all n = 0, 1, 2, 3, ... Thus the quotient difference

$f(x) – f(x_0) / x – x_0$

cannot have a limit as x -> x_0. So f cannot be differentiable at x_0.

4 Interlude: dogs wearing socks

Sections 1 and 3 differ in style and emphasis. In Section 1 we emphasized the 'idea' of the construction and in in Section 3 the emphasis was on an attempt to give a rigorous expression of the construction. Such a difference will be the main theme of the rest of this paper. Before that we shall consider a simple example.

Consider the following two problems.

Problem 1. The dog Snoopy wants to wear socks on his paws. Moreover, he wants to have socks of the same colour on all his paws. Snoopy has a box that contains blue, red and yellow socks. The box is in a totally dark storage. There is no difference between the socks besides the colour that Snoopy can use to find out the colour. So Snoopy has to take some socks out of the storage to choose socks of the same colour. What is the smallest number of socks that suffice for Snoopy to take out?

Problem 2. Assume that a set X is the union of three subsets A, B and C,

$X = A \cup B \cup C.$

Find the smallest number k such that for every subset Y of X containing ≥ k elements, one of the intersections

$Y \cap A$, $Y \cap B$ or $Y \cap C$

contains at least 4 elements.

It is perhaps not too risky to guess that many people would find problem 1 much easier than problem 2. But from the point of view of mathematical content, the problems are exactly the same.

The difference is in presentation. Problem 1 suggests strongly mental images to use in working with the problem. Moreover our emotions get easily involved in a positive way: many of us want to help Snoopy. Problem 2 does not raise similar emotions and remains more at a distance. Moreover the mathematical language and notation does not strongly suggest mental images or such to use in analysing the problem.

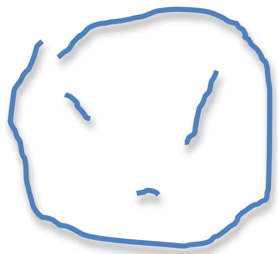
The power of mental images appears sometimes even in the terminology of mathematical research literature. Ramsey theory is a branch of combinatorics where the well-known Pigeon-Hole-Principe is elaborated in a nontrivial way. In Ramsey-theory one divides the family of subsets of n elements (for a fixed n) of a given set in subsets (families) and studies subsets of the given set that are homogeneous for this division in a certain sense.

For example, sets of two people can be dived between those where the people know each other and those where they do not know each other. In this case a homogenous set will be a set of people who all know each other or none of whom knows the others. (‘Knowing’ is here understood to be symmetric.)

The object of Ramsey-theory sounds (and it is!) very technical and abstract. Therefore it is interesting to observe that the divisions studied are often called colourings of sets of n elements in the research literature of the branch of mathematics.

The difference between sections 1 and 3 of this paper are to some extent analogous to the difference between problems 1 and 2 of this section. In Section 1 we have an example of an attempt to (create,) express and communicate mathematical ideas by means of intuitive mental images. In Section 3 we have a more technical presentation and the reader has to construct meaning to it.

Before going to a more careful analysis of this distinction, one more final example of the use of pictures to communicate mathematical information. Here is what one might even call the ‘universal picture’ in mathematics: some kind of an oval and some dots etc. in it.



Many of us have met uses of analogous images in attempts to solve some problem or in situations where somebody is attempting to communicate an idea. As such, there is really ‘nothing’ in the picture. But in connection to the attempts to think or communicate mathematics, such pictures may be of great value.

The pictures in Section 1 have a rather similar role. Next we shall take a deeper look at the distinction discussed in this interlude.

5 The “objective – formal” and “human – social” sides of mathematics

Our interest in this latter part of the present paper lies in the relation between mathematics and us. It turns out that there are (at least) two ‘dimensions’ in which one can make divisions in it. Firstly, there are aspects that are objective and others that are subjective or social. To the first belong printed formulas and pictures that one can find in earlier parts of this paper. To the latter belong my mental images that I as the author had in my mind while writing those formulas or making those pictures, and your mental images that you as a reader had in your mind while reading the paper. This distinction is the subject of the present section. We refer to this by speaking about the two sides of mathematics.

Another distinction separates for example pictures and formulas from each other. This is related to the famous three worlds of mathematics of David Tall and it will be discussed in later sections.

The two sides of mathematics appear in literature in many ways. One of the most important is the distinction between *concept image* and *concept definition* in Tall and Vinner [1981]. (It seems that the origin of David Tall's three worlds of mathematics lies there.) *Concept image* refers roughly to one's understanding of a mathematical concept and *concept definition* to the official definition of the concept.

Sfard (Sfard 1991) considers the dualism between the *operational* and *structural* sides of mathematics. This duality is related to a view of three steps: *interiorization*, *condensation* and *reification*. Here seems to be a vague analogy to the three worlds of Tall to be discussed in a later section.

There are also some other interesting features in Sfard's paper. She remarks about mental images that '... mental images can be manipulated almost like real objects'. She also points out that higher mathematics is considered rather little in literature and goes on to stress the importance of higher mathematics in this respect.

These have some resemblance to the three worlds of Tall, but the ordering makes a big difference. The doctoral thesis of Hähkiöniemi (2006) is interesting in this respect. Tall (2008) remarks that Hähkiöniemi (2006) considered the routes of students towards learning the derivative. Tall say that he 'found that the embodied world offers powerful thinking tools for students' who 'consider the derivative as an object at an early stage'. According to Tall this questions Sfrad's suggestion that operational thinking precedes structural. We shall consider David Tall's views in more detail later.

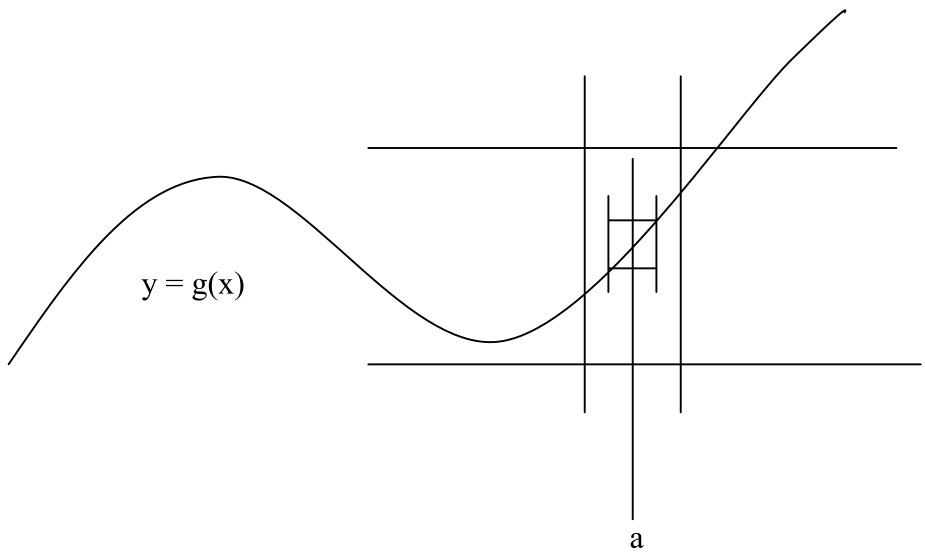
Also the paper Fishbein (1994) has some connections to the present paper.

In our own work as a university teacher of mathematics and a research mathematician, a division between two sides of mathematics has become important. But the emphasis seems to be somewhat different from those approaches referred to above. For us the division concerns what one does 'here and now' while e.g. teaching a mathematical concept: does one in the next moment speak about the ideas behind a mathematical concept or does one work with the formal definition. Earlier versions of our approach are discussed in Oikkonen (2009). (See also Oikkonen (2004) and Oikkonen (2008). The latter is a preliminary version of Oikkonen (2009).)

In Oikkonen (2009) and Oikkonen (2004]) one of the main ideas is to consider the human and formal sides of mathematics in teaching and in 'doing' mathematics. It has turned out fruitful in the author's own work as a university teacher to look at mathematics as something having a formal and human side.

This is very much connected to the author's attempt to share with his students the expert's way to think and to help them to step towards becoming an expert in this way. To concentrate in teaching on the interplay between these two sides, has become especially important in the way in which the author has developed his own teaching. We shall develop below these ideas further and relate them to the famous three worlds of mathematics of David Tall (see Tall (2013).) The construction of our function f at the beginning of this paper will act as an interesting example in this discussion.

Let us consider as an example of the two sides the notion of continuity of a function g at a point a. The idea is simple: g(x) should be near g(a) when x is near a. More exactly, if we draw two horizontal lines, one above and one below y = g(a), then there is are two such vertical lines, one left and one right of x = a that the graph of g(x) does not ‘cut’ the ‘floor’ or ‘roof’ of the rectangle formed by the four lines. And we can make these rectangles lower and lower (smaller and smaller) as long as we like. Two such rectangles are shown in the following picture



Pictures of this kind are related to a rich variety of mental images that are helpful in thinking about continuous functions. Such mental images are examples of what we mean by the human side of mathematics.

As such, the pictures drawn on the blackboard or printed above are objective in the sense that anybody can observe them. We shall come back to this aspect later.

Continuity has also a formal aspect. The continuity of g at a has the well-known definition: g is continuous at a, if (and only if) for every epsilon > 0 there is such a delta > 0 that |g(x) - g(a)| < epsilon for all x satisfying |x - a| < delta.

One of the main purposes of an introductory course in analysis is to teach this kind of definitions and proofs of the main theorems of analysis based on such definitions. But it is not an easy task. This is not helped by the way how we too often begin solutions of examples or proofs of theorems: 'Assume that epsilon > 0. Let delta = 3/7 times epsilon...'

Our experience in teaching analysis supports the idea that it is helpful to change the viewing point from which we look at mathematics. This takes place by combining the formal definitions with an active use of mental images like the one described above. By doing this it is also possible to reveal in teaching the way in which an expert mathematician thinks.

In our experience this kind of an approach helps in making the content of a mathematics course meaningful and understandable to students. Thus a course in mathematics is not only the polished formal content of the course but also – and to the author essentially – the thinking and culture that lies behind the text. We believe that this approach explains partially the success shown in Oikkonen (2009). (There are also other pedagogical ideas involved in this paper.)

Our own path to this kind of an approach results from the striking similarity between two seemingly quite different types of discussion on mathematics in which the we have taken part: those taking place in math days in elementary schools and those taking place when experts discuss some problem in research mathematics. The 'here and now' choice between different kinds of action that was mentioned above seems to be characteristic to such discourses.

So we have two sides of mathematics. But which of them is the correct one? Let us go back to continuity: which side is the correct one, the human (mental) images or the formal *epsilon-delta* definition? Our own answer is that neither of them is the correct one. The concept of continuity depends on both of its sides and it is to us really a kind of interplay between these two sides.

Let us consider our distinction more closely. It seems to us that it may be fruitful to adjust our terminology concerning the two sides of mathematics slightly. Below we shall call the two sides *subjective-social* and *formal-objective*. The first term corresponds to what was called the human side above but seems actually to be more varied. It contains individual aspects like one person's mental images, thinking and emotions, and it also includes social aspects like discussions or gestures. The latter corresponds to what was called formal above but is actually richer. It refers to those aspects of mathematics that are independent of the observer. Such include written mathematics (formulas, text and drawings on a blackboard, books etc.).

The author's personal view is that mathematics is interplay between its *formal-objective* and its *subjective-social* sides. This view comes from experiences in making and sharing mathematical research; and from experiences in teaching. This view seems to be a fruitful basis – at least for the author – for work to improve teaching and learning in mathematics. One aspect in this kind of division between the two sides is that whenever in contact to mathematics like in teaching, one has 'here and now' the choice between these. Think about a lecture about continuity. At every moment the lecturer either stimulates mental images and ideas or handles exact mathematical calculations etc.

Finally, how does the real line fit to this division between the two sides of mathematics? In a sense, the real line lives (in different ways) of both sides. We have and share a rich variety of mental images of the real line on the *subjective-social* side. We have various formal approaches to the real line on the *objective-formal* side. These include the axioms of the reals and constructions of the reals by means of e.g. Dedekind cuts or equivalence classes of Cauche sequences of the rationals.

As has been indicated earlier, this kind of a distinction appears between our Sections 1 and 3 as described in (i) below, and in a sense inside Section 1 as described in (ii) below.

(i) In Section 1 we approach the construction from the point of view of strong reference to mental images and in Section 3 the emphasis is on more conventional mathematical text.

(ii) I as the author had my own mental images of the construction and the function to be constructed. This lead to something objective, observable to everybody like the pictures used in Section 1. And while reading, the readers produce their own mental images.

Next we shall consider a different way of looking at the distinction between sections 1 and 3.

6 David Tall's three worlds

In Section 1 we discussed a construction of the function f by means of certain pictures in which the construction was embodied in a certain sense. At the end of the section there was some symbolic manipulation in connection to discussing the proofs of the continuity and nowhere differentiability of f.

In Section 3 we considered the function f from a more theoretical point of view. The presentation was symbolic of course. But it belongs actually to theoretical mathematics usually characterized by being formal and often referring to an axiomatic treatment of mathematics.

In this sense our construction of the function f and interest in it are related to David Tall's famous "three worlds of mathematics" (see e.g. Tall (2013) or (2008)) in an interesting way. It is also interesting to see the paper by Tall and Di Giacomo (2000) which studies the role of pictures in mathematics and especially in connection to a construction for a everywhere continuous nowhere differentiable function called there the Blancmange-function. See also Tall (1983).

Let us recall Tall's worlds briefly. Tall calls these by Tall the names *conceptual – embodied, proceptual – symbolic*, and *axiomatic – formal*. Tall has discussed his worlds in a great number of writings and the concepts have developed somewhat over the years.

In (Tall 2008) the worlds are described like this.

'The *conceptual-embodied world*, base on perception of and reflection on properties of objects, initially seen and sensed in the real world but then imagined in the mind;

the *perceptual-symbolic world* the grows out of the embodied world through action (such as counting) and is symbolised as thinkable concepts (such as number) the function both as processes to do and concepts to think about (precepts);

the *axiomatic-formal world* (based on formal definitions and proof), which reverses the sequence of construction of meaning from definitions based on known objects to formal concepts based on set-theoretic definitions.'

On page 133 of (Tall 2013) the worlds are described as follows:

'A world of (conceptual) *embodiment* building on human perceptions and actions develping mental images verbalized in increasingly sophisticated ways to become perfect mental entities in our imagination;

A world of (operational) *symbolism* developing from embodied human actions into symbolic procedures of calculation and manipulation that may be compressed into procepts to enable flexible operational thinking;

A world of (axiomatic) *formalism* building formal knowledge in axiomatic systems specified by set-theoretic definition, whose properties are deduced by mathematical proof.'

For our purposes, a good quick feeling of the distinction is given by how Tall in (2013, p. 25) relates the system of the real numbers to these worlds. The real numbers have *embodiment* as a number line, *symbolism* as (infinite) decimals etc. number notation, and *formalism* as a complete ordered field.

Tall's 'three worlds of mathematics' have an interesting relation to our view of the two sides of mathematics as will be discussed in the next section.

7 There are six = 2 x 3 parts of mathematics

In this section we argue that a combination of the two sides discussed in Section 5 and David Tall's three worlds discussed briefly in the previous section may lead to new insight in mathematics. Especially it may help in better understanding of Tall's three worlds.

It seems to us that our distinction between *subjective-social* and *objective-formal* and Tall's division between *conceptual – embodied, proceptual – symbolic*, and *axiomatic – formal* look at similar features in mathematical thinking from two different standpoints. Moreover the resulting 2 x 3 = 6 parts of mathematics help to see easier some aspects. So we shall consider some examples that show how each of Tall's three worlds seems to divide into two sides.

The case of the *conceptual – embodied* world seems especially natural. Our own mental images of mathematical objects or situations are subjective embodiment. It becomes social when a group of people shares such images while working on a problem. Various objects like number sticks etc. made for teaching mathematics are examples of objective embodiment.

A number line was mentioned above as an embodied version of the system of the real numbers. It can belong to either side depending on what we actually mean. The idea of a line of numbers belongs to the *subjective-social* side whereas an actual line drawn eg. on a blackboard belongs to the *objective-formal* side.

But is the real line itself an objective 'mathematical object' belonging to the *objective-formal* side of mathematics? What do we think about it and its existence? In a sense this is not an important question here. On the *subjective-social* side most mathematicians seem to behave as if the real line would actually "be there". But to us, it seems that we cannot distinguish those mathematicians who really believe that the real line "is there in a Platonic universe" from those who only behave as if it existed. The theorems concerning the reals are proved using the axioms of the reals in the *objective-formal* side of Tall's *axiomatic-formal* world and they make no direct reference to the truth or meaning of the actual statement the 'reals exist'. In this sense formalism and platonism are not very far from each other.

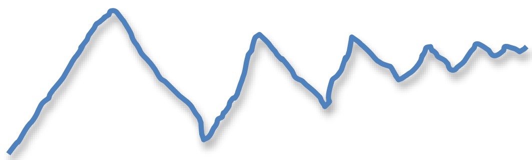
Moreover, it is not clear how to reply from a set-theoretic point of view to the question what the real line really is. Namely, there are different constructions (Dedekind-cuts of the rationals, certain equivalence classes of Cauchy sequences of the rationals) leading to different sets.

The case of the *proceptual – symbolic* world is more interesting. Rules for manipulating symbols and correct application of such rules belong to the *objective-formal* side. These include long divisions in elementary school or solving equations or doing differentiation of expressions for functions in upper secondary school. Students' own minitheories and systematic errors seem to belong to the subjective-social side of the the *proceptual – symbolic* world.

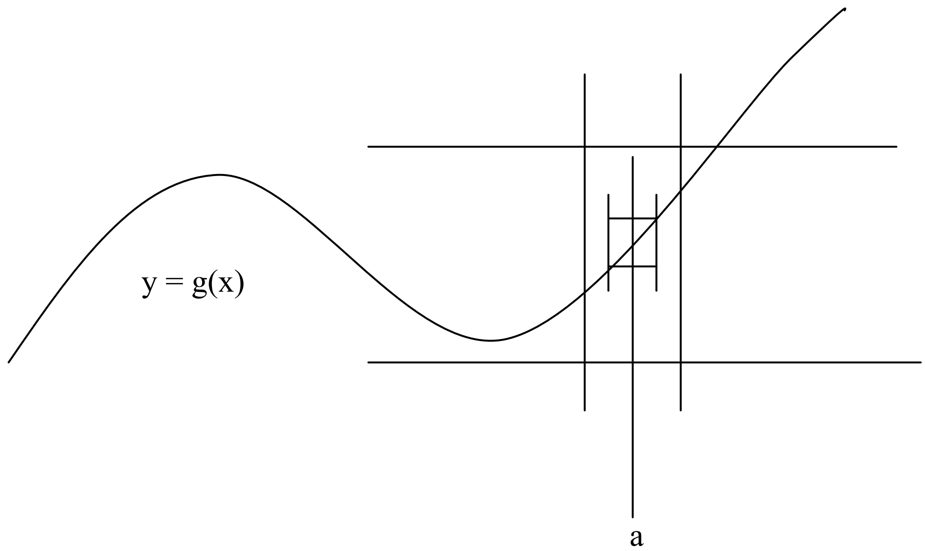
Perhaps also various routines applied in the so called *street mathematics* in basic calculations are also examples of this side of the *proceptual – symbolic* world.

Written university level mathematics with its axioms, definitions and theorems is an example of *objective-formal* side of the *axiomatic – formal* world of mathematics. Higher level strategic discussion on research mathematics belongs to the *subjective-social* side or the *axiomatic – formal* world. An example of this represents the statement "she mixed ideas from physics to analysis to solve the problem".

The step from the *subjective-social* side or the *axiomatic – formal* world to the *subjective-social* side of embodied mathematics with its mental images and gestures is very short. A nice example of this is in the Introduction of W. Hogdes' book (1985) where he tells about a difficulty with his dissertation. The supervisor C. C. Chang made a back and forth movement with his hand and said ‘this should help’. According to Hodges, it helped.



Before leaving this section we shall have closer look at the concept of continuity discussed above in connection to our two sides of mathematics. There we considered the following kind of drawings.



The notion of continuity and the function studied is clearly embodied in such a drawing. (Of course it is possible that there is no specific function that is considered and that the whole discussion concerns the concept of continuity.) This drawing is clearly objective in the sense that everybody can observe it. So the drawing belongs to the *objective-formal* / *conceptual-embodied* part of mathematics.

But these drawings are used either by oneself to think about continuity or by a group of people to discuss continuity. Such actions belong to the *subjective-social* / *conceptual-embodied* part of mathematics.

When one works with examples of assertions concerning continuity, one usually has to manipulate mathematical formulas. As long one thinks or discusses about how to proceed, one acts in the *subjective-social* / *proceptual-symbolic* part of mathematics When these formulas are actually written they become observable and thus objective and so one acts in the *objective-formal* / *proceptual-symbolic* part of mathematics.

But usually the real interest lies in understanding, teaching or using the 'epsilon-delta' -definition of continuity, and so the *subjective-social* or *objective-formal* side of Tall's *axiomatic-formal* world is involved.

8 Looking at the function f from the point of view of the six "places of math"

Now we shall consider the function f discussed at the beginning of this paper form the point of view of our two sides, Tall's three worlds and the six parts of the previous section.

Let us first consider our distinction between the two sides of mathematics. The hand-drawn pictures used in section 1 of this paper are *formal-objective* in the sense that anybody can observe them. In this sense they are like written mathematics. But to use them in connection with our function f as above, several *subjective-social* activities are essential.

The pictures get meaning in connection to thinking and discussing the construction and properties of the function f. The author and hopefully the reader constructed several mental images of their own related to them. If the construction is to be discussed between several people like in a seminar, then some kind of social sharing mental images will take place.

On the other hand such thinking and discussion would be very hard in absence of the pictures (drawn or imagined). In an important sense, the definition and proofs related to f are not merely in the pictures, nor in the thinking and discussions. What seems to be the heart of the matter is the interplay between the discussion and the objective pictures.

In this way the construction of our function f seems to be a good example of the interplay between the two sides of mathematics.

The construction of our function f seems to be a good example of the interplay between the two sides of mathematics. The discussion of the construction of f had strong use of pictures and mental images. But these were a way of communicating the formal construction of the function and formal proofs for the continuity and nowhere differentiability of the function f.

Our construction of the function f seems to relate to Tall's three worlds in the following way. The hand drawn pictures and discussions around them are *conceptual – embodied*. Thus Tall's *conceptual-embodied* seems to contain aspects that belong to our *formal-objective* side (the pictures) and aspects that belong to our *subjective-social* side (mental images and discussions).

Toying with our hand drawn pictures to get a concrete feeling of the function f had an important role in the beginning of this paper. This takes place when one by pen and pencil tries to get a feeling of the function f. (This can be used also when introducing the function f to students – in a university class or even in a secondary school.) Doing these calculations and making these pictures belong to the *human-social* side. The pictures and calculations as such belong to the *formal-objective* side. In Tall's terminology making these drawings and calculations with f belong to the *proceptual-symbolic* world.

The exact formulation of the definition of f and the exact proofs for the main properties of f are in Tall's terminology *axiomatic–formal*. In our terminology these aspects belong to the *objective-formal* side. But working towards these proofs thinking about them belongs to our *subjective-social* side.

This supports a view that our two sides of mathematics and Tall's three worlds of mathematics fit nicely together in a sense that they look at the same mathematical scenery from two "orthogonal" directions. Both of our two sides correspond to aspects of most of Tall's three worlds and each of Tall's three worlds has aspects of both of our two sides. This holds even for the *formal-axiomatic* world for example in the sense that reading and making proofs belong to our *subjective-social* world.

One of the most interesting features of the construction and argumentation concerning the function f at the beginning of this paper is that it serves as an example of an unusual route through the six parts of mathematics to present a piece of higher mathematics. Especially many details were left for the audience. But such omission of detail is typical for most published research mathematics: many details are left for the expert reader.

9 Epilogue

Both Tall's three worlds and our two sides of mathematics are closely related to attempts to understand how mathematics can be made meaningful to people.

At the website http://luma.fi/1324 there is a video where the author plays in this kind of a way with secondary school students with a visual description of a construction of a Peano curve. (The location of this video may change in future; the author shall help the reader.) This construction has very much in common with a the construction of our f in that they both are based on a similar use of simple pictures.

In Tall [2008] David Tall uses the blancmange function to illustrate his three worlds. In his paper, images of the function have a central role as an embodiment of the function. Also the main properties of the function are made obvious by means of these pictures. Indeed, when one zooms towards smaller and smaller details, one becomes convinced that the graph never seems to approach locally a straight line. So the function obviously cannot be differentiable anywhere.

But these pictures are the product of a computer program, not one's own thinking. Hence to believe in the nowhere differentiability of the blancmange function, one has to rely on the output of a machine – i.e. in a kind of a black box. This is however only a very superficial impression. Tall describes in private communication how he lets his audience to get a feeling of what happens in the construction: *"...that the pictures are absolutely intended to make embodied sense and to be interpreted in terms of 'ordinary thinking'. I do it with my hands in talks, imagining successive sawteeth. I use two hands to show a single sawtooth, then bend my fingers over to show two half-sawteeth. I even bend my hands and fingers to suggest four half sawteeth (try it!) but there the physical sensations need to move to imagination. I even get members of the audience to hold imaginary pictures, say with successive sawteeth to imagine the graphs added together. The idea is simple, the computations (which the computer performs) give precise pictures however, the ideas give an embodied proof that the function is nowhere differentiable, which has the power to be translated into a formal proof."* (Tall has had on his website photos of such an occasion in connection to a visit in Australia.)

It is interesting to take one more look at how Tall's three worlds and our two sides are related to learning and doing mathematics. Tall's worlds can easily be seen as three steps of growth towards deeper and more abstract (expertise in) mathematics. But a more correct view seems to be that more than one of them are present in school level mathematics and that all the three worlds are present in an expert's relation to mathematics. They have an established status in scientific literature on how we learn mathematic. Our two worlds come originally (in the author's own development) from observing the similarity between discussing mathematics in a math day with pupils in an elementary school to discussing mathematical research problems with other research mathematicians. This observation has been very helpful in the author's own teaching. At least to the author there is no controversy between Tall's three worlds and our two sides. They are simply two ways to look at learning and doing mathematics from different viewpoints.

References

Fischbein, Efraim. (1994). The interaction between the formal, the algorithmic and the intuitive components in a mathematical activity. In R. Biehler, R. W. Scholz, R. Strässer & B. Winkelmann (Eds.), (pp. 231-245). Dordrecht: Kluwer.

Hodges, Wilfrid. Building Models by Games. London Mathematical Society Student Texts. Cambridge . (1985)

Hähköniemi, Markku (2006). Tools for studying the derivative. Doctoral Thesis, Jyväskylä.

Oikkonen, Juha (2004). Mathematics between its two faces, in Matemaattisten aineiden opettajan taitotieto – haste vai mahdollisuus, L. Jalonen, T. Keranto and K. Kaila eds, University of Oulu, Finland, pp. 23-30, ISBN 951-42-7886-0.

Oikkonen, Juha (2008) Good results in teaching beginning math students in Helsinki, ICMI bulletin. 62, p. 74-80.

Oikkonen, Juha (2009). Ideas and results in teaching beginning maths students, International Journal of Mathematical Education in Science and Technology, 40:1,127 — 138

Sfard Anna (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, 1–36.

Tall, David. (1982). The blancmange function, continuous everywhere but differentiable nowhere, Mathematical Gazette, 66, 11-22.

Tall, David. The Transition to Formal Thinking in Mathematics. *Mathematics Education Research Journal,* 2008, 20 (2), 5-24 [A summary of the framework of three worlds of mathematics as applied to the shift to formal thinking] (2008)

Tall, David. How Humans Learn to Think Mathematically - Exploring the Three Worlds of Mathematics. Cambridge. (2013)

Tall, David (with Silvia Di Giacomo). Cosa vediamo nei disegni geometrici? (il caso della funzione blancmange), *Progetto Alice* 1 (2), 321–336). [English version: What do we "see" in geometric pictures? (the case of the blancmange function)]. (2000)

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

Thim, Johan. Continuous nowhere differentiable functions. Master's thesis, University of Luleå, 2003:320 CIV • ISSN: 1402 - 1617 • ISRN: LTU - EX - - 03/320 - - SE

Watson, Anna, Panayotis Spirou, David Tall. The Relationship between Physical Embodiment and Mathematical Symbolism: The Concept of Vector. *The Mediterranean Journal of Mathematics Education*. 1 2, 73-97. (2003).

A reference in Ramsey-theory