Visual Structure in Mathematical Expressions

Andrea Kohlhase[1], Michael Kohlhase[2], and Michael Fürsich[1]

[1] Information Management, University of Applied Sciences Neu-Ulm
[2] Computer Science, FAU Erlangen-Nürnberg

Abstract. Mathematics uses formulae to express knowledge about objects concisely and economically. Mathematical formulae are at the same time an indispensable tool for the initiated and a formidable barrier to novices. Surprisingly little is known about the cognitive basis of this practice. In this paper we start to rectify this situation with an investigation of how humans read (and understand) mathematical expressions.
A previous exploratory study suggested the interplay of visual patterns and content structure as a key ingredient of decoding and understanding math expressions that differentiates between math-literate and -illiterate subjects. The main contribution of this paper is an eye-tracking study on mathematically trained researchers conducted to verify the mathematical practices suggested by the first study and refine our understanding of the mechanisms.

1 Introduction

The art of expressing mathematical knowledge in math expressions evolved over the last three centuries and has revolutionized the way this is created, stored, and communicated. Given their importance for mathematical practice, surprisingly little is known about the cognitive basis of reading, understanding, and creating formulae.

As math expressions can neither be considered text nor image, it was previously suggested that they form a separate category (see e.g., [Fre]) which humans perceive differently. This perception was coined by Arcavi "symbol sense", i.e., a "*complex and multifaceted 'feel' for symbols [...] a quick or accurate appreciation, understanding, or instinct regarding symbols*" [Arc03, p. 31]. In 2005 W. Schnotz presented a study proving that "*comprehension is highly dependent on what kind of information is presented and how it is presented*" [Sch05, p. 73]. Even though he used only text and image information chunks, it is suspected that math expressions build a category that enables math-oriented persons to understand math in their own way. The "formula shock" [SGM10] effect is well-known, but how exactly do math-oriented people read math expressions differently? If we can get a deeper understanding, then we might get new insights for the design of math software or information systems. Let us summarize:

In [BR14] an eye-tracking study showed that people with a high mathematical expertise related proof items with an according supportive image. They found that the participants tend to jump between text and image and it was suggested that relating the different representations enables the relation in different memory stores. Thus, more effective and efficient retrieval and creation of math knowledge is enabled by the use of math expressions. In particular, with their visual and textual aspects they present a sophisticated cognitive tool for mathematicians.

In [KT13] Kamali and Tompa studied the retrieval for content in mathematical documents. Their empirical research also indicates that math expressions are special. They conclude that math expressions should neither be used as conventional document fragments nor with a too exact retrieval algorithm as both result in very poor search results. Instead they showed that algorithms that are based on making use of the Content MathML representation of the document corpus fare much better.

In this paper we report on two eye-tracking studies we have conducted to better understand the reading and understanding of formulae. Eye-tracking is an interesting angle of attack, as there is a demonstrable correlation between what a participant attends to and where she is looking at – see for example [Ray98] for an overview. The "eye-mind hypothesis" [HWH99] even claims a correlation between the cognitive processing of information and the person's gaze at the specific location of the information.

The paper starts off by briefly reporting on our exploratory eye-tracking pilot study in Section 2. That identified some potential *mathematical* practices when decoding and understanding math expressions by comparing students with different affinities towards math. For this paper we have conducted a new study with trained math researchers, which we report on in Section 3. Section 4 reports on how the new data affects the hypotheses developed in the pilot study. The new data suggests more math practices, which we explore in Section 5. In particular, we refine the "Operator Tree Practice" into the "Gestalt Tree Hypothesis", which combines the visual and conceptual sides of formulae. Section 6 concludes the paper.

2 The Math Expressions Pilot Study

To explore and identify idiosyncratic practices with math expressions, we invited 23 participants to look at concrete math expressions, e.g., the expression in the back of Fig. 1 in an eye-tracking study. [KF16] gives a detailed description. The goal of this exploratory study consisted in finding relevant discrepancies within various user groups. The only difference we could make out though was the one between math-oriented and non-math-oriented subjects.

Group	Female	Male
Math	5	5
¬Math	7	6

The math expressions were shown to the participants as images on a Tobii t60 Eye Tracking Screen (17" and 4:3 ratio with 60Hz). They were asked to think aloud while reading/understanding expressions; audio/video recordings were collected together with the eye-tracking data.

This data was analyzed to establish which practices are particular to mathematically inclined participants. On this basis [KF16] postulated a set of "math

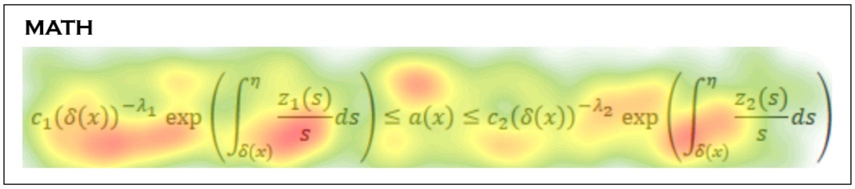


Fig. 1. Exemplary Math Expression with Heatmap of Math-oriented Subjects

practices" which were to be confirmed or rejected by a bigger study focused on mathematicians and more diverse mathematical expressions – the original goal for our main experiment.

We will confirm the following math practices from this pilot study [KF16]:

P-MP1: "*Math-oriented people use <u>visual</u> patterns for math detection*".

P-MP2: "*Simple math expressions can be treated as placeholders for argument positions and therefore as neglectable in the math expression decomposition process*".

P-MP3: "*The decomposition of a math expression is organized along its procedural character*".

P-MP4: "*In the decomposition of a math expression, some symbols carry structural information, which is read independently from its functional information*".

and discuss

P-MP5: "*The decoding of a math expression starts from the left until a first meaningful sub-expression is grasped. Further comprehension is chunked into understanding sub-expressions and their relations*".

Before we venture there, we report yet another conjectured math practice based on the pilot study data: We observed that non-math-oriented tended to look over the integral in Fig. 1 in a left-to-right fashion like reading text, while math-oriented subjects seemed to follow on the one hand **P-MP3** and on the other **P-MP5**. A more precise account of the latter is given in Fig. 2, which superimposes the sequence of fixations by an exemplary math-oriented subject over the (first subexpression of) the Integral represented as an operator tree in content MathML. The first two steps (1,2) segment the formula: the proband looked at large left bracket for orientation, identified the first factor $c_1(\delta(x))^{-\lambda_1}$ to its left, and then found the matching right bracket. Steps 3, moves to the integral symbol and step 4 fixes the lower bound – the upper limit does not seem to receive much by the MATH group. Finally, step 5 passes to the body of the integral before step 6 discovers that the integral has the exponential function applied to it. Then the attention shifts to the first factor again and explores its base and exponent (steps 7 and 8). This exploration of the left-hand-side of the inequality is followed by an orientation towards the right-hand-side via the two "$\leq$" symbols, and then an exploration of the right hand side that is very similar to the one detailed in

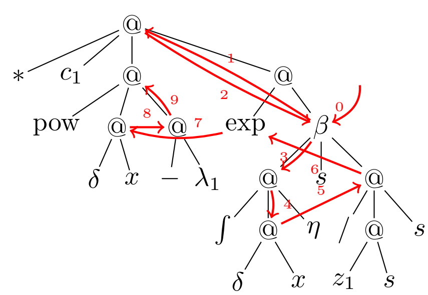


Fig. 2. Math-Exploration of Expression 3

Fig. 2. Thus, we suggest the combination of **P-MP3** and **P-MP5** as a stronger and better framed mathematical practice:

> **MP1:** "*Math expressions are decoded in the order of a depth-first traversal of the corresponding operator tree*".

This is somewhat surprising, since the gaze plot is induced by the proband's *process of decoding the mathematical expression*, whereas the operator tree – represented as a content MathML expression in [KF16] – represents the *result* of the decoding process. In particular, one would think that the decoding process would follow "visual patterns" and be driven by "visual cues".

To verify the established set of mathematical practices we conducted the experiment, we report on in the following.

3 The Math Expressions Study with Mathematicians

Group	Female	Male
MATH	1	28
¬MATH	0	0

This eye-tracking study was carried out at the CICM 2016 conference in Bialystok in July 2017. We were able to recruit 29 participants – all not only math-oriented but except one even mathematically highly trained scientists. The set-up was very similar to the one in our pilot study: we used the same Tobii t60 Eye Tracking Screen (17" and 4:3 ratio with 60Hz) but in a mobile setup.

We selected three math expressions (see Fig. 3) from CICM talks to use for the study, so that we could assume participants were relatively familiar with them. The study contrasted them with various manually constructed (variations

(1)	$\int_m^{4m} \left(e^{3x} + e^x\right) dx = 100$
(2)	$N(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
(3)	$\varepsilon\left(F_{\alpha\to\beta}A_\alpha\right) = app_{\varepsilon\to\varepsilon\to\varepsilon}\varepsilon\left(F_{\alpha\to\beta}\right)\varepsilon\left(A_\alpha\right)$

Fig. 3. The Math Expressions of our Eye-Tracking Study

of) "visual patterns" consisting of a subset of the expression structure replaced by blank space or empty boxes like e.g. the extreme case $\square = \square$. Fig. 7 and 8 show more examples. Probands were asked to carefully look at the equations from Fig. 3 and then for each of a group of visual patterns to determine if these were "*representations*"[3] of the respective expression. Subjects were encouraged to think aloud during the process but to be as fast as possible. The eye-movements, the questionnaire data, and the think-aloud protocol were recorded.

In the study we have gathered data concerning the order of fixations, the specific areas of the fixations, their length and amount of occurrences, and the questionnaire data. As the questions in the questionnaire were all of the same type – a 3-point Likert scale with options "yes", "no", and "I'm not sure" – we use (if at all) the standard mean as simple quantitative measure for the analysis of the questionnaire data.The eye-tracking specific data like data about fixations were analyzed via visual tools.

In particular, to get a better understanding of the order and intensity in which objects in math expressions are looked at, we used the visualization in form of a **gaze plot**, in which eye fixations are represented by dots that are connected, numbered, and accumulated according to their occurrence in a given time frame and whose size indicates the length of the gaze. **Heatmap**s as the one in Fig. 1 are another visual analysis tool. In an eye-tracking study the longest and most fixated areas are the hottest (red), the rarely fixated ones the coldest (green).

The general approach of our analysis of the new data was to inspect gaze plots and heat maps of randomly chosen probands for conspicuous patterns, interpret these, and check the remaining ones for typicality. We integrate the patterns found in [KF16] into this discussion.

4 Result: Math Practices wrt. Math Expression Decoding

In the following we concisely explain the conjectured math practices and argue confirmation or rejection.

Confirming MP1 and P-MP3; Discussing P-MP5: In Section 2 we have used an operator tree analysis for the left branch of the expression in Fig. 1 to formulate **MP1** ("*Math expressions are decoded in the order of a depth-first traversal of the corresponding operator tree*").

We studied all gaze plots of the math expressions in Fig. 3 to gain more insigt on this. Fig. 4 shows the (sequence of) fixations by proband P04 when reading expression (1) in Fig. 3 as a typical example – the corresponding operator tree on the left of Fig. 11. We distributed all 28 fixations over images a.) – d.) to conserve readability. Generally, the last fixation of the previous box is taken up as the first one ("1") in the following. In the top box we see the very first scan of a formula presented to P04. In particular, the subject doesn't know yet of the tasks to come wrt. this formula. The scan starts roughly in the middle, then fixates all the major components in a left-right-left sweep (fixations "1"-"10")

[3] As expected the term "representation" triggered various philosophical comments concerning its interpretation space and the resulting potentially different correct answers in the questionnaire.

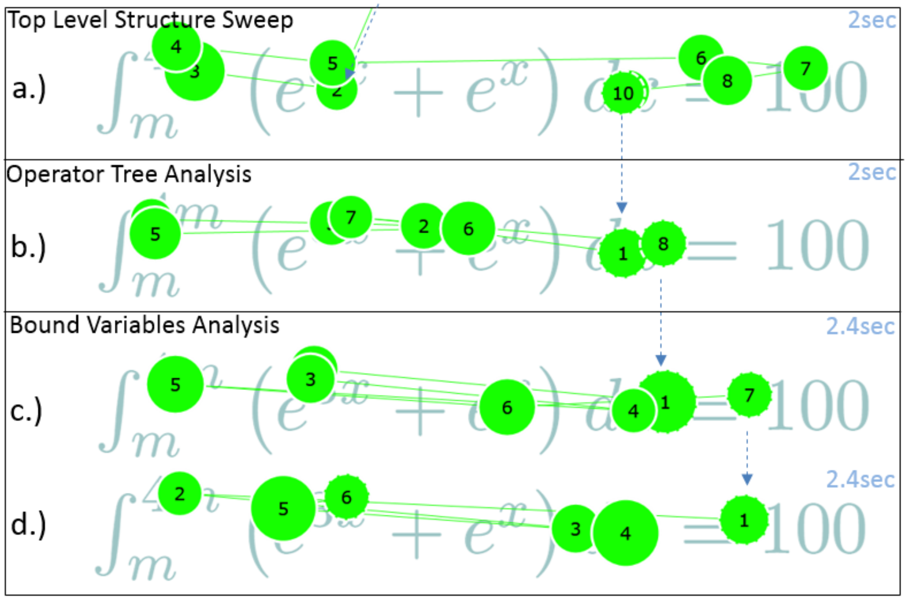


Fig. 4. The Gaze Plot for P04 on expression (1)

ending on the "dx" followed by an operator analysis according to the operator tree and finishing with some cross-checking with respect to the bound variable.

Our data fully confirms **MP1** and thus the weaker **P-MP3** ("*The decomposition of a math expression is organized along its procedural character*"): we consistently see fixations follow the operator tree in the expressions in Fig. 3. In particular, we verified that 21 out of the 29 participants followed the operator tree for expression (1), 18 for (2) and 21 for (3). Fig. 5 shows three data lines for tracing the operator tree which superpose if the values coincide. Note that the conspicuous absence of the blue and red line indicates that most subjects answered consistently.

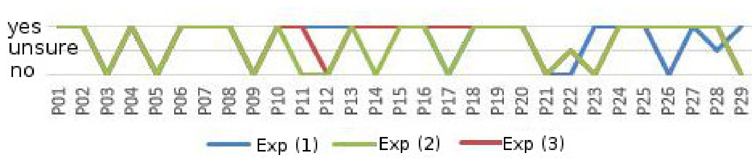


Fig. 5. Operator Tree Tracing

We could not confirm the unconstrained left-to-right aspect in **P-MP5** ("*The decoding of a math expression starts from the left until a first meaningful sub-expression is grasped. Further comprehension is chunked into understanding sub-expressions and* *their relations*"): by and large the first fixations are near the center of the formula, and then tend to move into the leftmost argument of the first operator (as is consistent with **MP1**). We now attribute the left-to-right aspect in **P-MP5** to the fact that the equation in Fig. 1 is an inequality chain, whose visual pattern (cf. **P-MP1**, but also see below) probands directly recognize and move into the left-most argument – i.e. all the way left in the expression. The chunking on the other hand could be observed, see Def. 2 for more discussion.

Confirming P-MP2: For **P-MP2** ("*Simple math expressions can be treated as placeholders for argument positions and therefore as neglectable in the math expression decomposition process*") we have to show more evidence that simple subexpressions in math expressions are ignored by most subjects in our study.

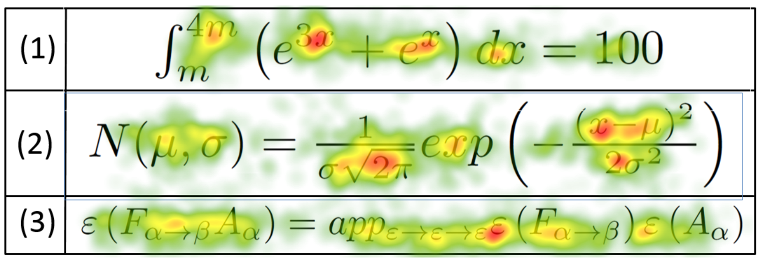


Fig. 6. Detecting Top-Level Structure in Formulae

We observed (see Fig. 6) that probands almost consistently did not fixate the lower limit m – a variable – of expression (1), whereas they overwhelmingly fixated the lower bound $\delta(x)$ and neglected the upper limit η in the integral of the exploratory study in Fig. 1. We conclude that probands see single variables in argument positions as "arbitrary" and only return to them if they also appear in more restricting context. So, we confirm **P-MP2**.

We conjecture that the upper limit in expression (1) and the lower in Fig. 1 are used as visual anchors, since they are larger and more complex.

Confirming P-MP1: We can uphold **P-MP1** ("*Math-oriented people use visual patterns for math detection*") for **detection** – i.e. determining that a text region was "Math": even though we only presented very reduced visual patterns (e.g. $\square = \square$, $\square \leq \square$) to probands, none of them complained that this is not "Math".

Confirming P-MP4: The mathematical practice **P-MP4** ("*In the decomposition of a math expression, some symbols carry structural information, which is read independently from its functional information*") can be confirmed. In Fig. 6, we can observe that the brackets and the integral are almost never fixated. Even the equation sign and the number on the right hand side of the equation are mostly unfocused. We remark that large (stretchy) operators like brackets and displayed integrals are more salient as structuring operators.

In Fig. 4 we have to assume that the overall structure of the formula has been determined in the very first sweep. Thus, together with **P-MP4** the top-level structure of the expression seems to be visual, so that we conjecture

> **MP2:** "*Visual patterns are used for detection of top-level structure in expressions*".

5 Result: The Gestalt Tree Hypothesis

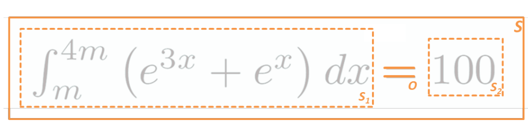
Once more let us have a closer look at Fig. 4. We already observed that in box a.) P04 essentially conducted a sweep of the expression from left to right to grasp the top-level structure. In the next box b.) the head operator of the body of the integral (the operator "+" in fixation 2) is fixated, then the first summand, the upper integral limit, and the second summand to come back to the "dx". In the last sweep depicted in the lowest box containing c.) and d.) of Fig. 4 the fixations jump between subexpressions and integral limits and seem to concentrate on the bound variable x. One is tempted to interpret the gaze plot as containing three phases. The first one establishes the nature as an integral equation, the second one establishes the operator tree of the integrand, whereas the third pass checks the occurrences of the bound variables.

5.1 Visual Structure and Gestalt in Math Expressions

The above gives us the leading intuition and further evidence strongly suggests that an even stronger version of **MP2** and **MP1** may hold. Before we can analyze and state this, we will have to invest in some terminology that combines concepts from computer vision and formula structure.

Definition 1. The **top-level (visual) structure** (TLVS) of an expression consists of a **segment** S – a set of pixels that is considered a meaningful part of the image – that encompasses the expression, an **operator** O (also called the **head symbol**), and **sub-segments** S_i of S that mark arguments of O.

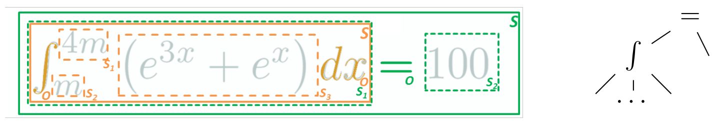
Example 1. The top-level structure of expression (1) is an "equation", where S is the whole image, O is the symbol "=", and the sub-segments S_i correspond to the left/right hand sides of the equation:



Note that TLVSes can be nested; e.g. the left hand side of (1) is an integral, so here O is the integral operator together with dx. We have three segments for the limits and the body of the integral. This nesting structure induces a tree structure, where the nodes are operators: the operator tree of the expression.

Definition 2. The **visual structure** (VS) of a mathematical expression consists of a hierarchical segmentation S of the expression together with an operator tree T, such that S – seen as a tree given by the nesting structure of segments- and T are isomorphic, i.e. tree structure and operators coincide.

Example 2. We obtain the following (two-level to ensure legibility) visual structure for (1), it yields the operator tree on the right:



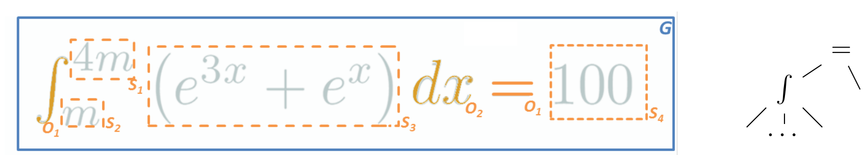
The top-level heat map of expression (1) in Fig. 6 and the initial "sweep phase" of the gaze-plot in Fig. 4 show us an interesting feature wrt. our data: It seems that all of the segmentation of (1) in Example 2 is something the experienced readers can take in at one go. Arguably (1) is an "integral equation", which is a – two-deep – visual structure in the sense of Definition 2.

This 'holistic' phenomenon seems to be closely related with the notion of "**Gestalt**" as defined in [Wag+12, p. 1218]: "*A Gestalt is an integrated, coherent structure or form, a whole that is different from the sum of the parts. Gestalts emerge spontaneously from self-organizational processes in the brain. Gestalts result from global field forces that lead to the simplest possible organization, or minimum solution, given the available stimulation.*". Even though classical Gestalt theory is controversial modern versions are actively debated today [ibid.]. Following these ideas we define

Definition 3. A **condensed visual structure** of an expression E consists of a segment G that encompasses E, an **operator tree** T, and segments S_i of G, such that the S_i correspond to the leaves of T.

The **Gestalt** of E is a condensed visual structure (G, T, S_i) that can be decoded by human readers holistically.

Example 3. The Gestalt of (1) is that of an "integral equation", i.e.



Note that Gestalts can be nested just like TLVS. We call such a nested tree a **Gestalt tree** (GT), and GT naturally induce operator trees a well (trees of trees can be flattened to trees).

With Definition 3 we can postulate a set of **Gestalt patterns** for mathematicians, i.e. operator tree/segmentation patterns that can be used to decode mathematical expressions – in principle – modulo resource considerations like the size of the argument segments. We conjecture that Gestalt patterns are learned along with the mathematical concepts they correspond to. Thus, the set

$$\int_{\square}^{\square} \left(\qquad\qquad \right) dx = 100$$

Q: representation of (1)? **A**: yes/no/unsure $\rightsquigarrow$ 22/4/3

Fig. 7. Accepting Visual Patterns

of Gestalt patterns is particular to a mathematician or mathematical community of practice and defines what she/they can decode easily. Note also that Gestalt patterns are very close to the visual patterns we use for testing in our study.

Using our new concepts, we can reformulate and strengthen **MP2** to

MP3: "*Gestalt patterns are used for decoding structure in math expressions*".

We remark that **MP3** resolves the apparent contradiction in the observation that the gaze plot seems to follow the operator tree of an expression which should be the result of the decoding process: The (instantaneous) recognition of the Gestalt of a formula establishes the inner segmentation, and so it is not surprising that subsequent recursive exploration respects this.

Supporting MP3 in the Eyetracking Data: We observe that all probands respect the segmentation of the Gestalt tree, e.g. in equations the first couple of fixations are all in the left-hand side, and then switch over to the right-hand side, only occasionally fixating the head symbol. Together with **P-MP4** above this strongly supports **MP3**: as the visual cues of the TLVS are not fixated, but the segmentation is obeyed, it must be taken in precognitively – i.e., in one piece.

MP3 is also borne out by probands' answers in Fig. 7, where a majority accepted the visualization of the TLVS of expression (1) in Fig. 3: when asked whether this Fig. 7 is a representation of (1) 23 of 29 answered positively and only 4 negatively. Finally, 28 of 29 probands rejected out of hand that the TLVS visualization in Fig. 8 could be a representation of expression (3). This shows that the operator O is a constitutive part of the TLVS and thus the Gestalt.

$$\boxed{\qquad\qquad} \leq \boxed{\qquad\qquad\qquad\qquad}$$

Q: representation of (3)? **A**: yes/no/unsure $\rightsquigarrow$ 1/28/0

Fig. 8. Rejecting Visual Patterns

5.2 Gestalt in Visual Imagery

The most baffling result of our study is that when confronted with visual patterns that abstracted the equations from Fig. 3, directly after the (full) equations themselves, probands fixated the locations where salient parts of the hidden subexpressions had been. Fig. 9 shows the situation; note that the fixations are quite similar to those in the second box in Fig. 4, i.e. after the integral equation Gestalt pattern has been decoded. The only real difference are fixations "6", "7", and "9" in Fig. 9, which seem to delimit the body segment of the integral Gestalt.

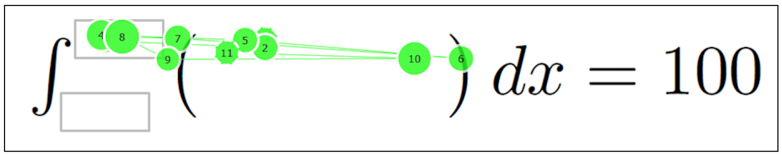


Fig. 9. Visual Imagery (P04) for a Visual Pattern for Expression (1)

We can describe this phenomenon in terms of **visual imagery**. The term describes the experience of "seeing with the minds eye" or in other words the human ability of "*to conjure up a visual experience in the absence of retinal stimulation*" [CHH12]. Whether the mental image is stored in the human brain in a pictoral form, a linguistic descriptive form, or a combination is still a vivid topic of study in cognitive psychology, see [Tye00] for a discussion. Here, we simply use the term for our observation of our subjects' eye movements along mathematical subexpressions that are absent in the image looked at in the study.

Surprisingly, visual imagery was not an isolated phenomenon, but in fact widespread in 28 probands (the eye-tracking recording of one proband was damaged and thus inconclusive) of our study. For expression (1) for instance, 16 clearly used visual imagery, for 9 subjects it was not sure and only 3 probands definitely did not use visual imagery. Fig. 10 shows more gaze plots; here we have superimposed the elided parts of the original formula in light blue to show the precision of the visual imagery.

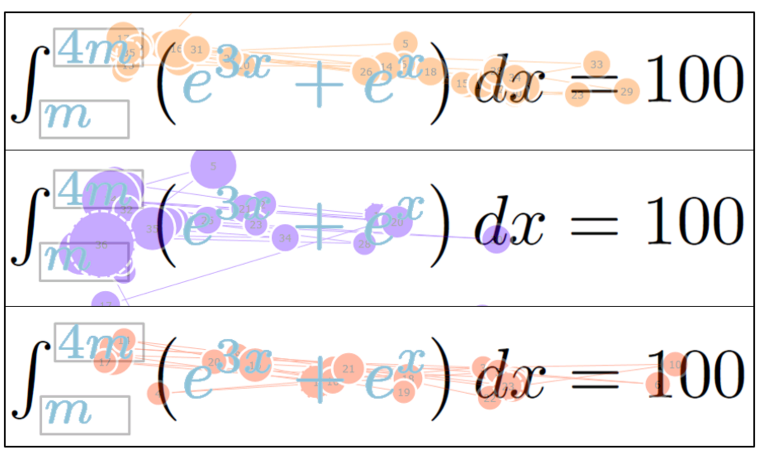


Fig. 10. More Visual Imagery of P02, P14, and P16

On the other hand visual imagery was much less pronounced in the probabilistic distribution function (2) and the type theory formula (3): For expression (2) the decision whether subjects used visual imagery was much more difficult. We identified 6 as clearly exhibiting visual imagery, with 10 we were not sure and 12 were not. For expression (3) we didn't dare a valid assessment. This difference to (1) may be a consequence of the fact that in subsequent visual patterns the 'box abstraction' became expected and thus only box sizes were taken into account: Instead of visual imagery of formula structures we start to see visual scans of box boundaries or dimensions. We will have to conduct further experiments to distinguish between effects here.

Even though the visual/structural imagery effect diminishes in our study, we interpret it as a strong support for the notion that Gestalt patterns drive the decoding process for mathematical expressions (in trained mathematicians), i.e **MP3**.

The Formula Understanding Phase: But **MP3** only covers part of the data we observe when probands study formulae. Concretely, it covers the first two sweeps we saw from P04 in the two top boxes in Fig. 4. But many probands of the main study (and most math-oriented of the exploratory one) followed it up with a "cross-checking phase", which consists of fixations that directly access interesting semantical features – e.g. bound variables in various subexpressions, or other variables, literals, or subtrees that occur more than once. We conjecture that this "direct access" is only possible because the Gestalt tree determines the **loci** (i.e. locations in the presentation that correspond to subtrees in the content/operator tree) for them. The few exceptions did consistently cross-check after each of the last phases, which fits as well.

> **MP4:** "*Once the Gestalt tree is established, probands cross-check details across loci in the Gestalt tree*".

We tentatively interpret this "cross-checking phase" we see the eye-tracker data as a semantic understanding phase, where formula readers correlate information from the subexpressions to each other. We conjecture that in this phase readers also correlate subexpressions with expressions from the context. To check this, we would have to conduct experiments, where some of the context is explicitly represented in the document presented to the user. We expect to see fixations on the context expressions in the "cross-checking phase".

The Gestalt Tree Hypothesis: We sum up **MP3** and **MP4** as

> **Gestalt Tree Hypothesis (GTH)**:
> The process of decoding mathematical expressions has two phases
> 1. recursively establishing the Gestalt tree by matching the formula presentation against Gestalt patterns (the **reading/parsing phase**).
> 2. cross-checking structural detail across loci in the visual tree (the **understanding phase**).

We formulate this as a hypothesis even though our data supports it quite strongly, since our study was only designed to study the effects of “visual patterns” on the formula decoding process of mathematicians and not particularly the GTH.

Representing the Gestalt Tree: Now that we have identified potential mechanisms used by math-oriented human readers in reading and understanding formulae, let us see how that relates to the representational practices in mathematical knowledge management.

The core intuition behind the Gestalt concept and by homomorphic extension behind Gestalt trees is that they correlate visual patterns and operator structure. This directly maps to the concept of “parallel markup via cross-references” in MathML [Aus+10, section 5.4.2].

Fig. 11 shows the content MathML tree of expression (1) on the left and the presentation tree on the right. The cross-references that mark up corresponding subtrees are shown as green dashed lines. The content tree is made up of operator applications (@ and their children) and bindings (β nodes). In our example the binding is the integral where the first child of the β is the binding operator – the integral with its limits –, the second is the bound variable x, and the third is the integrand, which is an operator application again. In the layout tree we find layout primitives like rows and sub/superscript patterns.

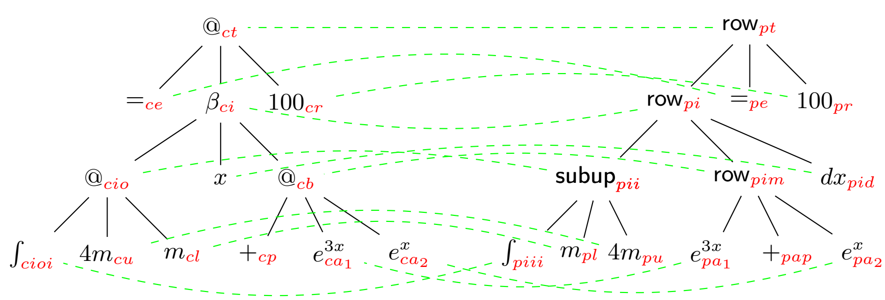
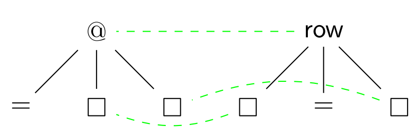


Fig. 11. Content and Presentation Trees of the Integral Expression (1)



If we tease apart layer for layer conserving the cross-references, we get one-level correspondences like the one shown on the right that correspond directly to TLVS or (more-level) VS, which can be condensed into Gestalts by backgrounding inner cross-references. Coincidentally, these structures directly correspond to the notation definitions for OMDoc [KMR08]. The notation-definition-based parsing process we studied in [TK16] can (after the fact) also be seen as an implementation of the first phase in the Gestalt Tree Hypothesis.

6 Conclusion

In this paper we have tried to shed some light on the processes governing how trained mathematicians read and understand mathematical formulae. After an exploratory study with a mixed group of math-oriented and non-math-oriented students, which revealed cognitive practices particular to math-oriented subjects, we conducted a follow-up study with 28 mathematically highly trained scientists to study these mathematical practices in more detail. Our study is a qualitative data analysis – for which 28 subjects is a very respectable number. In particular, we do not claim a quantitative signiticance.

Our scientific contribution consists in the confirmation/refinement of several math practices regarding the decoding of math expressions. In a nutshell, we could show that math expressions are decoded in the order of a depth-first traversal of the operator tree, simple ones often only serve as placeholders for argument positions, and visual patterns are used for top-level structure detection in math expressions.

In the eye-tracking data, it became apparent that the structure detection phase of the parsing process is almost instantaneous and "holistic" which led us to coin the concept of a formula Gestalt and hypothesize that Gestalt patterns govern the way mathematicians parse formulae. While the Gestalt Tree Hypothesis is strongly supported by the eye-tracking data obtained from our study, the study itself was not designed to test this hypothesis per se. Therefore, more careful studies are needed to prove it in the future.

Moreover, we noticed significant differences between the groups of participants that self-identified as computer scientists vs. mathematicians for gazing at expression (3) in Fig. 3, which is in the domain of type-theory that is more tied to CS than math: The CS group paid significantly more attention to the type components, whereas the true mathematicians seemed more interested in the homomorphic structure at the term level. This suggests that familiarity with the domain affects the way probands read formulae. We leave the investigation of this interesting aspect to future work.

Finally, it would be interesting to compile a catalog of Gestalt patterns commonly in use in Mathematics. Actually, given that we can represent Gestalt patterns in notation definitions, we can interpret the notation definitions in the SMGloM terminology base [Koh14] as a start that could be tested for cognitive reality.

References

[Arc03] Abraham Arcavi. "The role of visual representations in the learning of mathematics". In: *Educational Studies in Mathematics* 52.3 (2003), pp. 215–241.

[Aus+10] Ron Ausbrooks et al. *Mathematical Markup Language (MathML) Version 3.0*. W3C Recommendation. World Wide Web Consortium (W3C), 2010. URL: `http://www.w3.org/TR/MathML3`.

[BR14] Jana Beitlich and Kristina Reiss. "Das Lesen mathematischer Beweise - Eine Eye Tracking Studie". In: *Beiträge zum Mathematikunterricht 2014*. Ed. by J. Roth and J. Ames. WTM-Verlag, 2014, pp. 157–160.

[CHH12] Radoslaw M. Cichy, Jakob Heinzle, and John-Dylan Haynes. "Imagery and Perception Share Cortical Representations of Content and Location". In: *Cerebral Cortex* 22.2 (2012), p. 372.

[Fre] Hans Freudenthal. *Didactical Phenomenology of Mathematical Structures (Mathematics Education Library)*. Dordrecht: Reidel. ISBN: 9027715351.

[HWH99] John M. Henderson, Phillip A. Weeks Jr., and Andrew Hollingworth. "The effects of semantic consistency on eye movements during complex scene viewing". In: *Journal of Experimental Psychology: Human Perception and Performance* 25.1 (1999), pp. 210–228.

[KF16] Andrea Kohlhase and Michael Fürsich. "Understanding Mathematical Expressions: An Eye-Tracking Study". In: *Mathematical User Interfaces Workshop*. Ed. by Andrea Kohlhase and Paul Libbrecht. July 2016. URL: `http://ceur-ws.org/Vol-1785/M2.pdf`.

[KL16] Andrea Kohlhase and Paul Libbrecht, eds. *Mathematical User Interfaces Workshop at CICM*. July 2016. URL: `http://ceur-ws.org/Vol-1785/`.

[KMR08] Michael Kohlhase, Christine Müller, and Florian Rabe. "Notations for Living Mathematical Documents". In: *Intelligent Computer Mathematics*. Ed. by Serge Autexier et al. LNAI 5144. Springer Verlag, 2008, pp. 504–519. URL: `http://omdoc.org/pubs/mkm08-notations.pdf`.

[Koh14] Michael Kohlhase. "A Data Model and Encoding for a Semantic, Multilingual Terminology of Mathematics". In: *Intelligent Computer Mathematics 2014*. Ed. by Stephan Watt et al. LNCS 8543. Springer, 2014, pp. 169–183. ISBN: 978-3-319-08433-6. URL: `http://kwarc.info/kohlhase/papers/cicm14-smglom.pdf`.

[KT13] Shahab Kamali and Frank Wm. Tompa. "Retrieving Documents with Mathematical Content". In: *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Ed. by Gareth J. F. Jones et al. Dublin, Ireland: ACM, 2013, pp. 353–362. ISBN: 978-1-4503-2034-4.

[Ray98] Keith Rayner. "Eye Movements in Reading and Information Processing: 20 Years of Research". English. In: 124.3 (1998), pp. 372–422.

[Sch05] Wolfgang Schnotz. "An Integrated Model of Text and Picture Comprehension". In: *The Cambridge Handbook of Multimedia Learning*. Ed. by R. E. Mayer. Cambridge University Press, 2005 (2014), 49–69 (72–103).

[SGM10] Alexander Strahl, Julian Grobe, and Rainer Müller. "Was schreckt bei Formeln ab? - Untersuchung zur Darstellung von Formeln". In: *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung* 0.0 (2010).

[TK16] Ion Toloaca and Michael Kohlhase. "Notation-based Semantification". In: *Mathematical User Interfaces Workshop*. Ed. by Andrea Kohlhase and Paul Libbrecht. July 2016, pp. 73–81. URL: `http://ceur-ws.org/Vol-1785/M6.pdf`.

[Tye00] Michael Tye. *The Imagery Debate)*. Representation and Mind series. MIT Press, 2000.

[Wag+12] Johan Wagemans et al. "A Century of Gestalt Psychology in Visual Perception II. Conceptual and Theoretical Foundations". In: *Psychological Bulletin* 138 (6 Nov. 2012), pp. 1218–1252.