Discourse Phenomena in Mathematical Documents

Andrea Kohlhase[1], Michael Kohlhase[2], and Taweechai Ouypornkochagorn[3]

[1] Hochschule Neu-Ulm, Germany
[2] Friedrich-Alexander University Erlangen-Nürnberg, Germany
[3] Srinakharinwirot University, Thailand

Abstract. Much of the wealth of industrialized societies is based on knowledge that is laid down and communicated in scientific/technical/engineering/mathematical documents: highly structured documents that contain diagrams, images, and – most daunting to many readers – mathematical formulae. It seems clear that digital, interactive documents have the potential to improve reading these kind of documents, and thus learning and applying this kind of knowledge.
To understand how such improvements could be designed, we explore how formula understanding interacts with the surrounding text in mathematical documents. We report on an eye-tracking experiment with 23 engineering students reading a "solved problem" based on a simple differential equation. We observe for instance that – triggered by formulae – readers backjump to previously identified semantic loci and that this behavior is independent of depth of understanding in mathematically trained readers. Based on our observations, we propose novel in-document interactions that could potentially enhance reading efficiency.

1 Introduction

Millions of people engage in reading and understanding scientific/technical/-engineering/mathematics (STEM) documents – Germany alone has about two million scientists and engineers. So even a single-digit improvement of this reading/understanding productivity will translate to considerable societal effects.

To raise the productivity we focus on the mathematical documents themselves. Currently, almost all documents are static – usually printed or in page description formats like PDF. They are highly structured, contain diagrams, images, and mathematical formulae, the last one being most daunting to many readers. As reading behaviour does not depend on the medium according to [ZC12], replacing these static documents by digital, interactive documents has the obvious potential to make reading them, and thus learning and applying STEM knowledge, more personal, efficient, effective, and fun. But first we need a better grasp on how people read and understand STEM documents, especially the mathematical parts.

In this paper we report on a design research study for establishing a nascent theory based on the observed phenomena on the discourse-level of embedded mathematics. That is, we were interested in the cognitive mechanisms of reading STEM documents that allow us to systematically design interactive features to improve the reading experience. **Design Research** is a rather young area in Human Computer Interaction – see [ZSF10; HC10]. It includes the "Research for Design" approach, which typically results in **nascent** theories: "*propose tentative answers to novel questions of how and why, often merely suggesting new connections between phenomena*" [EM07, p. 1158].

Concretely, we report on an eye-tracking study that focuses on the interplay of text and formulae in the written communication of mathematical knowledge. This 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.

Related Work Generally, a lot of studies were conducted to understand in which way eBooks should be designed to improve the reading experience, a summary is given in [Mar09]. When reading academic and scholarly materials, readers triage[4] documents by scanning. Studies like [BL07] have consistently found that titles, section headings, and emphasized text have a high value for document triage and facilitate reading this way. Marshall reports in [Mar09], that when reading more intensely, readers move back and forth through the document. Some authors assume that backjumps are "*an implicit sign that the reader is having difficulty understanding the material*" [Cha+16].

In previous work the first two authors have studied how humans read and understand mathematical formulae from their visual representations via eye-tracking experiments. In [KF16] we show that the level of affinity towards mathematics distinguishes how readers process formulae and in [KKF17] we show that mathematically trained readers process formulae by recursively identifying content-oriented patterns (Gestalts) that build up the formula semantics (Gestalt Tree Hypothesis). This first parsing phase is followed by a "formula understanding phase" in which salient semantic loci are systematically re-visited, e.g. for tracking bound variables in integrals.

Our literature search didn't reveal any results concerning the **discourse level** – i.e., the phenomena above the phrase and formula structures: sentences, paragraphs, dialogue – in mathematics or even STEM documents.

Overview After documenting the experimental setup in Section 2 we observe and discuss conspicuous patterns in subjects' gaze behaviors in Section 3. In Section 4 we proceed with a quantitative analysis of gaze data on specifically selected document fragments. Observations and explanations are summarized in a nascent theory for reading mathematical documents in Section 5. Subsection 5.1 discusses how these findings can lead to better interaction with STEM documents and Section 6 concludes the paper.

[4] **Document triage** refers to readers' practice of rapidly evaluating documents to determine whether they contain wanted information or not.

2 The Eye-Tracking Study

Experimental Setup This eye-tracking study was carried out with the students of a one-week special course "*Content and Form: How one manipulates the other*" at Srinakharinwirot University (SWU), Thailand. Most of the 23 participants were students of the Biomechanical Engineering Program at SWU, the remaining three were students of Electrical Engineering. The primary textbook of these programs is "*Advanced Engineering Mathematics*" by Erwin Kreyszig [Kre06]. Reading Kreyszig's book in English is a well-practiced part of the program, so the English/Thai language barrier should be minimal for reading. We also note that in the sense of [KF16] all participants are math-oriented and mathematically trained.

In our study we used a solved problem called "*Radioactive Decay*" in [Kre06, p. 13], which had been transliterated into HTML5 with MathML by the authors, as mathematical document[5]. Here, Kreyszig uses the example of radiocarbon-dating the Ötzi mummy to present standard methods for solving a boundary value problem induced by the simple ordinary differential equation $y' = ky$. This example comes very early in Kreyszig's book, therefore we assume that all subjects were familiar with the content and able to understand the mathematics.

The 23 students (16 female, 7 male) were presented the example on a Tobii t60 Eye-Tracking Screen (17" and 4:3 ratio with 60Hz) in a mobile setup. They were asked to think aloud while reading/understanding the document[6]; audio/video recordings were collected together with the eye-tracking data.

In a post-test, subjects were asked to write a from-memory summary of the document they read in the study in Thai. This was announced beforehand to subjects and was mainly intended to provide motivation for reading the document in detail. This document was later evaluated to assess the level of understanding of the participants. We believe that the "re-production" of results in Thai helped break down the language barrier.

Classifying the Understanding-Level of our Test Subjects The Thai summary documents were evaluated to classify the individual level of understanding at various levels. In particular, the post-test results were assessed according to whether the participants:

1. addressed the problem objective, i.e., determining the time of death of Ötzi;
2. addressed the physical background and/or solution methodology;
3. cite the eventual answer to the problem; and
4. correctly used formulae in the document.

[5] Figures 12 and 15 together show the full document content, see `https://kwarc.info/people/mkohlhase/data/DuStd-18/radiocarbon.html` for the transcription.

[6] This has been unsuccessful, it seems that the English/Thai language barrier combined with a cultural reluctance to speak without preparation together with the unfamiliar situation induced prohibitive cognitive load which prevented students from speaking. When we realized this, we asked one student to "think aloud in Thai", but this largely only resulted in a translation of the document, in particular not in the desired stream of cognition, so we dropped this idea.

We aggregated these to estimate the *level of understanding* by the students of the problem setting and methodology (U_p from 1. and 2.) and that of formulae in the text (U_f). For the former, we used a true/false scale, for the latter a five-point Likert scale: “yes”, “maybe”, “cannot judge”, “no formulae”, and “formulae catastrophically wrong”. Note that these judgements are not independent: A problem in understanding the methodology renders subsequent aspects not applicable: Indeed we had this in three cases, we found that subjects had misunderstood the problem to be “*Find the half-life of the mummy*”, “*Find the half-life of* ${}_6C^{14}$ *or* ${}_6C^{12}$ *will occur after the death*”, and “*Prove the ratio of Carbons is 52.66%*”.

It turns out that values in the understanding aspects U_p and U_f are largely identical where applicable, if we identify “yes” and “maybe” in U_p with “yes” in U_f and analogously “no formulae” and “formulae catastrophically wrong” in U_p with “no” in U_f. Therefore we used the “general aptitude” (high/low on the combined score) for grouping the participants into the *two groups* LOW and HIGH. See Figure 1 for the distribution, where “na” stands for non-applicable as we couldn’t sort the respective participants into any of the groups.

LOW	7
HIGH	14
na	2
Total	23

Fig. 1. Participants

3 Patterns in the Gaze Plots

We will now analyze the results of the eye-tracking experiment qualitatively. Concretely, we will study gaze plots, i.e., visualizations of fixations[7] over time, generated by the eye-tracker and show typical patterns. In general we distinguish between text, inline-math, and display-math areas and observe discourse phenomena as behavioural gaze patterns on these.

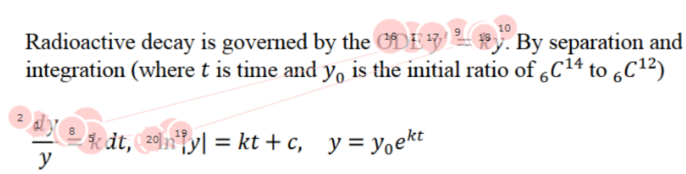


Fig. 2. Backjumping to Equations

Jumps The first pattern, we observed in the majority of participants’ gazes, is a **jump** from a math expression in a display-math area back into a related inline-math expression close-by and above, a “regression”. Figure 2 shows an example: While the subject looks at the equation “$\frac{dy}{y} = k\,dt$”, he/she jumps back to the ordinary differential equation “ODE $y' = ky$” in the text above which it was derived from. Afterwards he/she continues to the ensuing equations in the same display-math line. Interestingly, we did not observe once a forward jump from the display-math to an inline-math expression. That

[7] The eye moves discontinuously, making short stops (called **fixations**; we count any stop that is longer than 60 ms) and separating by rapid jumps (called **saccades**).

is quite natural at the beginning when one doesn't yet know what is about to come, but at the end, when all content is already known, we could have expected a different behaviour.

This backjumping pattern also happens in a more complex way between distinct display-math lines as in Figure 3, where the participant looks at an equation – here $y_0 e^{kt} = 0.5 y_0$ – after fixating on the left hand side (fixations 1,2) he jumps back to the equation $\frac{dy}{y} = k\,dt$ three lines above (fixations 3-5) and returns to the right hand side (fixations 6,7).

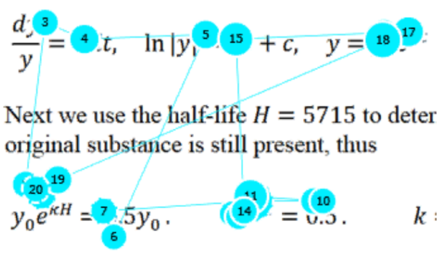


Fig. 3. Complex Backjumping

In the next equation $e^{kt} = 0.5$ something similar happens: after examining the equation, the subject jumps back to the equations $\ln|y| = kt + c$ and $y = y_0 e^k t$, glances at them without fully examining them again, and returns to the first equation on the lower line (glances at that and continues back to equation $e^{kt} = 0.5$ that was the origin of the detour).

Justification Processing In Figure 4 we see a participant reading the equation $\frac{dy}{y} = k\,dt$ and then jumping back to the phrase "By separation" and dwelling on this extensively before returning to the equation and moving on. Here "By separation and integration" is the justification for the equations $\frac{dy}{y} = k\,dt$ (separation) and $\ln|y| = kt + c$ (integration); indeed it is the main method introduced in the same chapter of [Kre06] that the document in our experiment was taken from. So the **Justification Processing** pattern describes backjumps to justification keywords in the text that help to process the reasoning behind the solution.

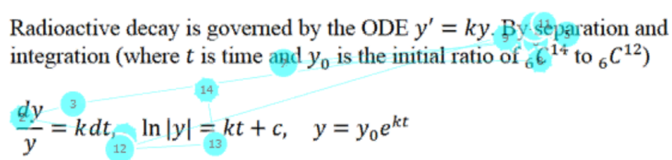


Fig. 4. Processing Justifications

Declaration Lookup The next pattern is somewhat less frequent, but still observed regularly. Figure 5 shows a situation, where the test subject is reading the equation $y = y_0 e^k t$ and starts from the left with y (fixation 10), continues to y_0 (fixation 11), jumps to the phrase "y_0 is the initial ratio ...", and back to the equation, which is examined more closely before moving on. In this **declaration lookup** pattern the backjumps are captured

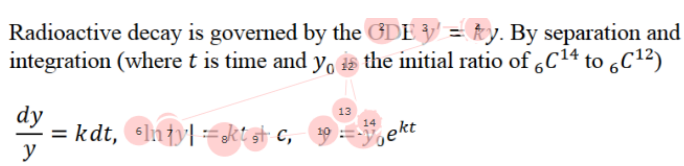
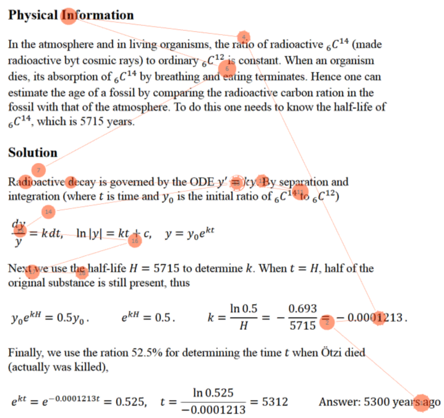
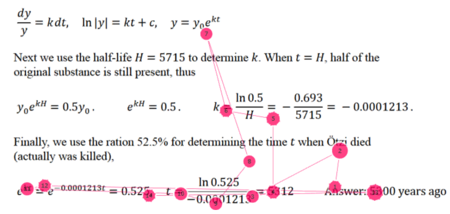


Fig. 5. Regression to a Declaration



(a) Text and Equations



(b) Equations Only

Fig. 8. Final Re-Check after Answer was Read

that are reaffirming information. Humans keep information in short term memory only for 10 - 30 seconds, so such jumps are necessary to keep the content available.

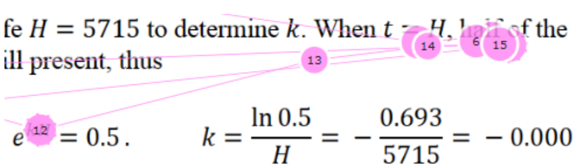


Fig. 6. Multiple Declaration Lookup

Figure 6 shows another instance of this. The participant reads the (exponent of the left-hand side in the) equation $e^{kH} = 0.5$ and then glances at the k in the line above and then reads "$t = H$ half of the", which reminds him of the description of the half-life H.

Re-Check and Re-Orientation A very common pattern is that participants read the explanatory text and the equations essentially sequentially or linewise (see Figure 7), with the latter being subject to the jumps and declaration lookups described above.

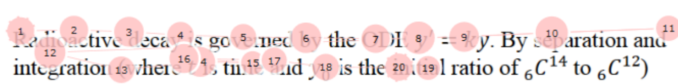


Fig. 7. Reading Linewise

But when they reach the answer at the end, about 2/3 of the subjects went over large parts of the document again in a much more targeted fashion. We call this pattern a **re-check**. Figure 8 shows two typical situations. Extended re-checks also occurred before the final answer was reached, but this was much less frequent. In this situation extended excursions focused on the background, problem, and physical information sections – we call these **re-orientations**.

Solution Pre-Scan Dually, we sometimes see a pre-scan that skips ahead to the eventual solution glancing over salient features of the document until the answer has been reached (see Figure 9).

This is usually followed by a line-wise reading of the solution text, which is supposedly informed by the pre-scanned text.

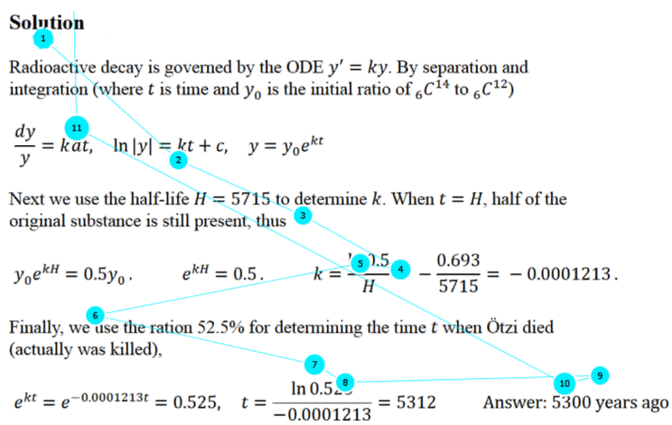
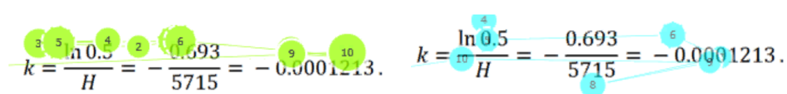


Fig. 9. Solution Pre-Scan

Multiple Ways of Reading Equations We found different patterns of reading the equation $k = \frac{\ln 0.5}{H} = \frac{0.693}{5715} = 0.0001213$: Figure 10(a) e.g. starts out with the middle of the three equality symbols, moved to the first, and then to the value 0.5 and then to the natural logarithm ln. From here the subject focuses on the value 0.693 for three fixations (6-8) before he/she moves on to the final value on the very right of the equation chain. Note that this sequence is consistent with the Gestalt Tree hypothesis from [KKF17] if we assume the middle = to be the main operator and the two others to be the main operators of its arguments.



(a) Equation Chain (b) Value Computation

Fig. 10. Different Ways to Read Equations

The proband of Figure 10(b) comes in from the phrase "determine k" and focuses on the first fraction, directly moves to the final value, and then fixates the middle right fraction, returns to the value, and then passes to the middle left fraction before moving on. We can assume that this participant interpreted the equation chain as an instance of the "value computation frame": an equation chain with a variable on the left and a scalar value on the right. As in [KKF17], we assume that single identifiers can be interpreted without fixation, so the initial parse of the Gestalt identifies the "outer equation" $k = 0.0001213$, which is indeed the relevant information. The middle fractions are checked as a secondary objective.

Whitespace to Think With Finally, we repeatedly found patterns like the ones in Figure 11. We interpret this as participants seeking a place without distracting information to fixate while thinking about what they were reading. We also frequently observed long fixations of the part headings – especially the heading "Solution", which seemed to serve a similar purpose.

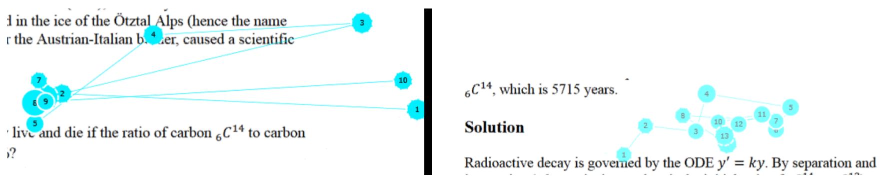


Fig. 11. Times for Reflection: Whitespace to Think With

One could think that the fixations on this area were due to mindless reading times, that is reading at the same time as thinking about something else. But in [RRS10] it was shown that mindless reading was always immediately preceded by especially erratic eye movements, so we can refute this argument.

4 Patterns in the Elicited Data

To make use of the eye-trackers statistical data, one has to define **Areas of Interest** (AOI). For each AOI eye-tracking specific metrics like " Time to First Fixation" or "Visit Duration" are elicited from the gathered data. A **visit** is the period of time from entering an AOI to leaving it, whereas a **fixation** occurs, when the eye focuses on a position for a certain amount of time. A visit contains at least one fixation otherwise it is not counted as a visit as the test subject otherwise just jumped over the AOI without having time to perceive any information.

To get a better understanding of the Gestalt Tree we distinguished between the visual document structure and the visual structure of mathematical expressions, so we defined **structure-driven AOIs** as in Figure 12 and **math-driven AOIs** as shown in Figure 15. The colored areas in those figures represent such areas of interest. For instance, we marked the heading of each subsection in Figure 12 independently from its content, so we have a Background-Title-AOI and a Background-Area-AOI. These were then used for analyzing the eye-tracker's data, predominantly using the mean for comparing the eye-tracking specific metrics.

4.1 Structure-Driven Areas of Interest

Kreyszig's example consists of four top-level text areas introduced by explicit headings: the background information, the problem statement, the solution-relevant physical background, and the solution. All these were given a clearly

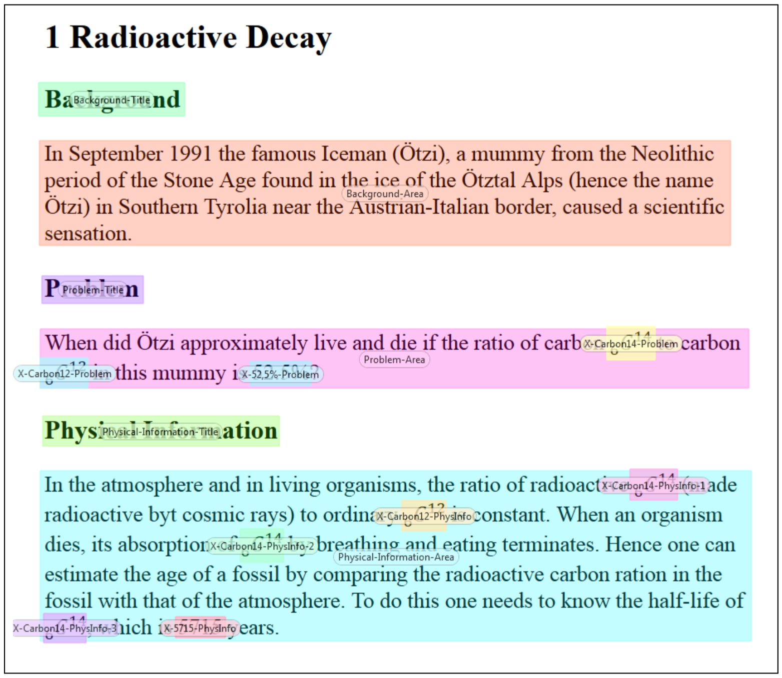


Fig. 12. Structure-Driven AOIs

marked title and provide an obvious visual text structure (see AOIs on x-axis in Figure 13).

First, we were interested how often our subjects visited or fixated the text components, respectively. In Figure 13 we visualized the number of visits and fixations of the structural text components depending on participants' level of understanding.

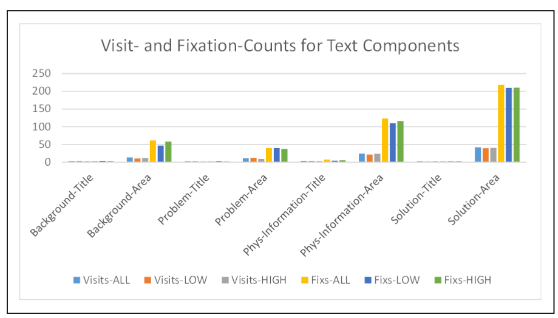


Fig. 13. Visits and Fixations for S-AOIs

An interesting observation consists in the strong difference between fixations and visits. Especially in the solution area the number of fixations is more than quadrupling the number of visits. That means that subjects jumped a lot between sub-elements within the respective area during a visit. So the interactivity rate for the solution area is highest, followed by the relevant physical information area. The background information was more often fixated than the problem description. Looking at the size of the areas, this is not surprising: the problem statement is about half the size of the background area, almost doubled by the physical information, which in turn is almost doubled by the solution. Therefore, we could expect a linear growth in the number of fixations. But if we look closely, we can see that it is more than linear.

At first glance surprisingly, there is neither a difference in terms of number of fixations nor visits between students of high or low level document understanding. [MGA14] already found a similar result with respect to text comprehension. Our analysis suggests that this is true with respect to mathematical text as well.

Next, we looked at other metrics to (nevertheless) found our intuition about existing differences among the participant groups HIGH and LOW.

Figure 14 shows the relative total duration of visiting or fixating specific areas, distinguishing the ones with a high level of understanding from the ones with a low level. Note that the visit duration has always to be higher than the fixation duration, as subects' fixations depend on their visit of this area. Moreover, we used the standard threshold for fixation length of $60ms$, so besides the durations for non-classified participants, all shorter fixations are also summed up in the general "*-ALL" variables.

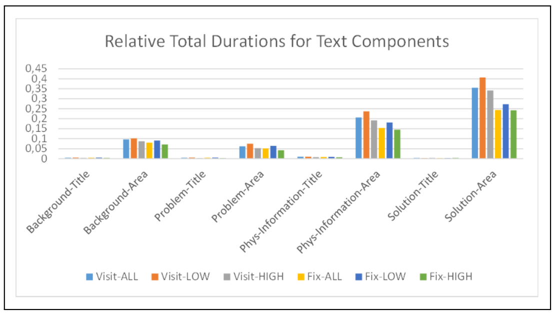


Fig. 14. Relative Total Visit- and Fixation-Duration for the Structure-Driven AOIs

Non-surprisingly, in general the title areas were very shortly looked at and the solution area the longest. The participants of the HIGH group spent less time fixating and less time visiting all areas. The length of fixations and visits could therefore indicate aptitude to the task, that is, it could give us a measure for *personal complexity of information*.

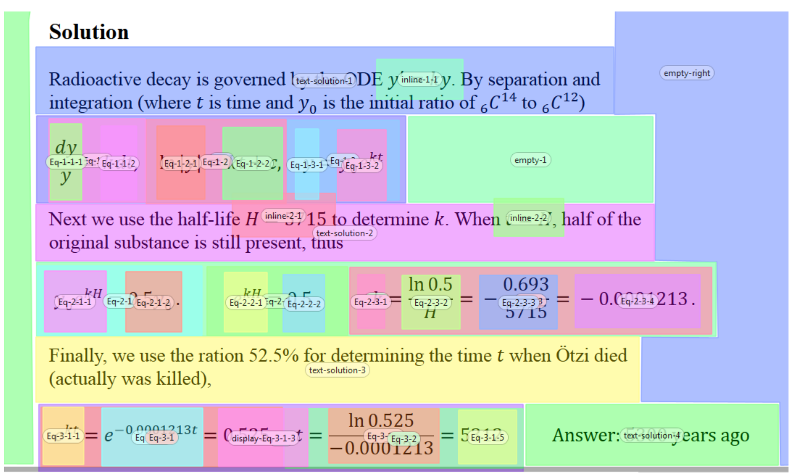


Fig. 15. Math-Driven AOIs

4.2 Math-Driven Areas of Interest

To get a better grasp about the discourse level when reading mathematics, we had a closer look to the solution itself as it contains paragraphs and formula areas. Except for the last text area all are of a comparable size, so that any differences can be attributed to the content itself. We distinguish four nested levels of formula areas (see Figure 15):

- the display equation level containing the display math in a line represents the **highest level**, e.g. "Eq-1".
- Each equation area contains several subareas with individual equations like "Eq-1-1" (separated by a comma), which we consider the **intermediate level**,
- which in turn consists of the sub-expressions on the left and the right of the equation sign of each su-bexpression on the **lowest level**, for instance "Eq-1-1-1".
- In the text areas in the solution there are several occurrences of **inline math** like "inline-1-1", that is math that is embedded in text.

We built the AOIs according to this structure, where Eq-1 – Eq-3 belong to the highest level, Eq-1-1 – Eq-3-2 belong to the intermediate level etc.

One would expect the number of the counts to equal the number of the sum of its parts. But this is not the case. Instead we can see in Figure 16 that the sum of the subarea visits is consistently larger than the higher level area, the sum of the subarea fixations in contrast consistently lower.

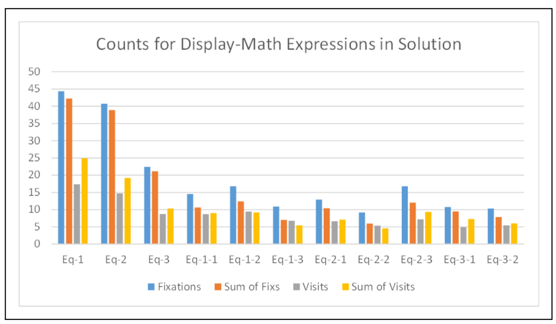


Fig. 16. Fixations and Visits in the Solution

As we left out some areas inside the higher level mathematical expressions to be covered by lower level ones, the higher number of fixations is explained by our subjects not always fixating on the sublevel areas but somewhere in-between. They didn't always fixate the formulae, only close-by - maybe on a recheck-jump in-between or to perceive the higher-level Gestalt information. The higher number of visits on the lower level areas means that the test subjects left and entered sublevel areas without leaving the higher level mathematical expression. Note that this inside-math-expression interactivity happened on all higher-levels and in a 50-50 ratio on the intermediate levels.

4.3 What are Cognitive Units (Words) in Math?

Another point of interest is the distribution of gaze intensity over the document. Typically, this is assessed visually by heatmaps. But the AOIs allow us to "do the numbers" to see how intensity differs by type of content, here between text, inline, and display math.

But how to define **gaze intensity**? Heat maps visualize fixations per space. With AOIs, we can calculate the number of fixations per AOI visit, but we still have to normalize for "space". For text the natural cognitive unit is a *word*, which is usually read holistically, or alternatively *word characters*, if we want to take word length into account. We can only compare gaze intensity on text and math, when we understand what the equivalent to a word or a word character in formulae is, but as we will see, that is not per se clear. To fortify our intuition, let us look at the following sentence from our document:

Radioactive decay is governed by the ODE $y' = ky$.

We can determine the number of words in the text part to be 7 and we count 34 letters ignoring the whitespaces. For the mathematical part it is indeed difficult: Does the entire equation represent a word, does the left part of the equation equal a word, or do we have to count the units explicitly spoken in the mathematical expression like "y prime equals k (times) y". Furthermore, we have to decide whether the equality sign is comparable to punctuation or what else is.

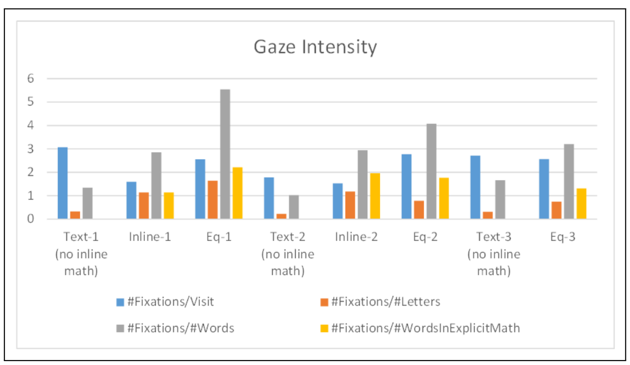


Fig. 17. Fixations per AOI Size and Word Number

Figure 17 shows the fixations per AOI and various space normalizations. Here, we looked only at the math-driven AOIs of type text, inline-math and display math on the highest level. Together these AOIs cover all non-empty areas in the solution section and each fixation falls into exactly one AOI. For #Words in formulae we counted the "equation sides" whereas in #WordsInExplicitMath we counted operators and constants which roughly corresponds to words in "spoken formulae", but normalizes for multi-word operators like "e to the power of k (times) t". If we compare the intensity columns for the various AOIs, we see that the values are significantly higher on formulae than text (by factors ranging from 2 to 5 for #Letters and #Words). Only for the intensity measure that normalizes with #WordsInExplicitMath we see comparable values.

The first (blue) column in Figure 17 shows the fixations per visit, which is not a "gaze intensity" as it is not normalized for (cognitive) space, but it is another intensity measure. The more fixations there are per visit, the more complex the information. We can interpret the results in terms of

i) *cognitive load* – and therefore probably information content and relevance to a STEM document: formulae are much more content-rich than text *ii*) a *complexity measure of formulae*: nodes in the operator tree representation of formulae are an adequate size measure for formulae.

Note that these interpretations do not compete as they address different aspects. The latter interpretation is consistent with the Gestalt Tree Hypothesis from [KKF17] and the finding of [Bau99] that mathematical formulae (especially inline ones) have grammatical function in the surrounding text and are best modeled by integration of the mathematical and sentence grammars.

5 A Nascent Theory for Mathematical Documents

We summarize the most relevant observations to be used for explorative design in a first, admittedly very basic nascent theory.

NT1: The top level block structure of the experiment's mathematical document was noticed and used while reading – reconfirming previous results for general text reading for mathematical document reading.

NT2: Participants mostly read the text linewise in document order, starting at the beginning. The more was read, the more regressions happened, especially when starting the solution area which has a high intensity of formulae. Here, we recognized several distinct patterns.

- *Local Regressions* to *i*) identifier declarations, *ii*) equations the current one is derived from, and *iii*) justifications of the current equation. Contrary to [Cha+16] – which admittedly was not focused computer science papers, not on mathematics – we found that students with a high level of understanding used regressions – probably to deepen their knowledge.
- *Non-Local Regressions* to the problem and background descriptions after they have read the solution. Recaps are largely driven by the display formulae in the derivation, whereas re-orientations are driven by terminology coreference.

NT3: Participants spend significantly more time on formulae than on text. This also holds for inline formulae, but only if they are more than one letter long. As a rule of thumb, one operator or constant in a formula is worth one word in a text.

NT4: There seems to be a lot of looking to the right margin and the section title areas, which can be interpreted as a need for "spaces to think with".

5.1 Design Application: Interactive Documents

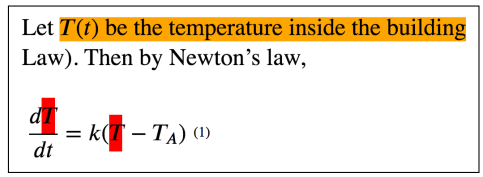
We give some very first examples for explorative design ideas based on **NT1** - **NT4** to showcase the value of a nascent theory.

For instance, if structural elements are easily findable, then we strengthen the effect of **NT1**: What about personalizing structural layout? That way everyone loves and lives with her/his own recognition clues.

When reading/understanding formulae, people look for declarations (**NT2-**i). With active documents – i.e., interactive documents that adapt themselves to the user; see [Koh+11] – we can support this process.

Consider for instance a simple instrumentation like the one shown in Figure 18, where we have bound the hover event on the sub-formula T to highlight the declaration text and other occurrences of the same identifier via two lines of `jQuery`. Of course, we also systematically need label/ref annotations on corresponding sub-formulae, which is the real bottleneck.

We have shown in [WG10] that the vast majority of identifiers is declared near where they are used. [Kri+12; Sch16] present a simple algorithm and framework for "declaration spotting", which could be used to generate such annotations automatically. Of course, this needs to be improved substantially to be practical for document instrumentation.


Let $T(t)$ be the temperature inside the building
Law). Then by Newton's law,

$$\frac{dT}{dt} = k(T - T_A) \quad (1)$$

Fig. 18. Instrumenting Documents

Similarly, we can instrument the document to support the re-check and re-orientation phases diagnosed in (**NT2-***Non-Local Regressions*). For re-checking, we would instrument the formula dependency relation: for a formlula F focused by the reader[8] we could highlight all formulae (or alternatively their relevant parts) that compute objects in F. For instance, if F is e^{kt} on the left of the last equation line, then we could highlight the dependency $k = \frac{\ln 0.5}{H} = \frac{0.693}{5715} = 0.0001213$ and the declaration "t is time". Again, we are presuming annotations for the dependency relation in the document. Note that the recap/re-orientation patterns discussed in Section 3 suggest that a highlighting instrumentation is more effective than e.g. generated summary (at least when all recaps are on the same page): the subjects seemed to have a very clear notion of where to find the information they were looking for in the document.

Based on **NT3** we could explore the effect of enlarging inline math or we use the complexity measure based on mathematical words to assess the necessity of assistance features.

Introducing explicit thinking space to a mathematical document layout seems rather unusual, but might be helpful for the reflection process in individuals according to **NT4**.

We have experimented with these and other instrumentations during the course, but the eye-tracking studies on this were inconclusive, as the number of tested students were too small and the participants untrained in the new features. The latter was a main problem in all of the instrumentations: they were not easily discoverable – we did not want to change the appearance of the document too much – and subjects needed time for understanding what they were seeing.

[8] There is of course the practical problem of how to determine whether F is focused. We could use in-place-and-time eye-tracking data instead of forcing the reader to e.g. hover the mouse over F to "focus" it, as this might be too distracting.

6 Conclusion

In this paper we report on an eye-tracking experiment, observing engineering students reading a mathematical document. In contrast to other studies which focus on text or formulae alone, we focus on the discourse-level interaction of text and formulae. We have identified various conspicuous patterns in the data from the experiment and shown how these could be used to improve the reading and understanding experience of STEM practitioners. The nascent flavor of our theory notwithstanding it is already useful; we can e.g. refute the assumption that the occurrence of backjumps indicate a low level of understanding as suggested in [Cha+16].

In the future we plan to systematically design interactive features for STEM documents and evaluate the effectiveness and efficiency of reading and understanding STEM documents. On the other hand we plan eye-tracking studies that further elucidate the cognitive processes behind perceiving mathematical documents, hopefully ending with a "mature" theory, which "*present[s] well-developed constructs and models that have been studied over time with increasing precision [...] resulting [...] in points of broad agreement*" [EM07, p. 1158].

Acknowledgements The authors gratefully acknowledge the Erasmus+ Staff Mobility for Teaching program that provided the institutional setting of the design research, and the students of the course for their participation in the eye-tracking study.

References

[Bau99] Judith Baur. "Syntax und Semantik mathematischer Texte – ein Prototyp". MA thesis. Saarbrücken, Germany: Fachrichtung Computerlinguistik, Universität des Saarlandes, 1999.

[BL07] George Buchanan and Fernando Loizides. "Investigating Document Triage on Paper and Electronic Media". In: *Research and Advanced Technology for Digital Libraries*. Ed. by László Kovács, Norbert Fuhr, and Carlo Meghini. Springer Verlag, 2007, pp. 416–427.

[Cha+16] S. S. Mozaffari Chanijani et al. "An eye movement study on scientific papers using wearable eye tracking technology". In: *2016 Ninth International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*. IEEE, 2016, pp. 1–6. DOI: 10.1109/ICMU.2016.7742085.

[EM07] Amy Edmondson and Stacy McManus. "Methodological Fit in Management Field Research". In: *Academy of Management Review* 32 (Oct. 2007), pp. 1155–1179.

[HC10] Alan Hevner and Samir Chatterjee. *Design Research in Information Systems: Theory and Practice*. 1st. Springer Publishing Company, Incorporated, 2010.

[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. DOI: 10.1037/0096-1523.25.1.210.

[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`.

[KKF17] Andrea Kohlhase, Michael Kohlhase, and Michael Fürsich. "Visual Structure in Math Expressions". In: *Intelligent Computer Mathematics (CICM) 2017*. Ed. by Herman Geuvers et al. LNAI 10383. Springer, 2017. DOI: `10.1007/978-3-319-62075-6`.

[Koh+11] Michael Kohlhase et al. "The Planetary System: Web 3.0 & Active Documents for STEM". In: *Procedia Computer Science* 4 (2011): *Special issue: Proceedings of the International Conference on Computational Science (ICCS)*. Ed. by Mitsuhisa Sato et al. Finalist at the Executable Paper Grand Challenge, pp. 598–607. DOI: `10.1016/j.procs.2011.04.063`.

[Kre06] Erwin Kreyszig. *Advanced Engineering Mathematics*. 9th ed. Wiley, 2006.

[Kri+12] Giovanni Yoko Kristianto et al. "Extracting Definitions of Mathematical Expressions in Scientific Papers". In: *The 26th Annual Conference of the Japanese Society for Artificial Intelligence*. 2012. URL: `https://kaigi.org/jsai/webprogram/2012/pdf/719.pdf`.

[Mar09] Catherine C. Marshall. *Reading and Writing the Electronic Book*. Morgan and Claypool Publishers, 2009.

[MGA14] Pascual Martínez-Gómez and Akiko Aizawa. "Recognition of Understanding Level and Language Skill Using Measurements of Reading Behavior". In: *Proceedings of the 19th International Conference on Intelligent User Interfaces*. IUI '14. Haifa, Israel: ACM, 2014, pp. 95–104. DOI: `10.1145/2557500.2557546`.

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

[RRS10] Erik D. Reichle, Andrew E. Reineberg, and Jonathan W. Schooler. "Eye Movements During Mindless Reading". In: vol. 21. 9. 2010, pp. 1300–1310. DOI: `10.1177/0956797610378686`.

[Sch16] Jan Frederik Schaefer. "Declaration Spotting in Mathematical Documents". B. Sc. Thesis. Jacobs University Bremen, 2016. URL: `https://gl.kwarc.info/supervision/BSc-archive/blob/master/2016/schaefer-frederick.pdf`.

[WG10] Magdalena Wolska and Mihai Grigore. "Symbol Declarations in Mathematical Writing: A Corpus Study". In: *Towards Digital Mathematics Library, DML workshop*. Ed. by Petr Sojka. Masaryk University, Brno, 2010, pp. 119–127. URL: `http://dml.cz/bitstream/handle/10338.dmlcz/702580/DML_003-2010-1_14.pdf`.

[ZC12] Daniela Zambarbieri and Elena Carniglia. "Eye movement analysis of reading from computer displays, eReaders and printed books". In: *Ophthalmic and Physiological Optics* 32.5 (2012), 390–xb396. DOI: `10.1111/j.1475-1313.2012.00930.x`.

[ZSF10] John Zimmerman, Erik Stolterman, and Jodi Forlizzi. "An Analysis and Critique of Research Through Design: Towards a Formalization of a Research Approach". In: *Proceedings of the 8th ACM Conference on Designing Interactive Systems*. DIS '10. Aarhus, Denmark: ACM, 2010, pp. 310–319. DOI: `10.1145/1858171.1858228`.

A The Mathematical Document of our Eye-Tracking Study

1 Radioactive Decay

Background

In September 1991 the famous Iceman (Ötzi), a mummy from the Neolithic period of the Stone Age found in the ice of the Ötztal Alps (hence the name Ötzi) in Southern Tyrolia near the Austrian-Italian border, caused a scientific sensation.

Problem

When did Ötzi approximately live and die if the ratio of carbon ${}_6C^{14}$ to carbon ${}_6C^{12}$ in this mummy is 52.5%?

Physical Information

In the atmosphere and in living organisms, the ratio of radioactive ${}_6C^{14}$ (made radioactive byt cosmic rays) to ordinary ${}_6C^{12}$ is constant. When an organism dies, its absorption of ${}_6C^{14}$ by breathing and eating terminates. Hence one can estimate the age of a fossil by comparing the radioactive carbon ration in the fossil with that of the atmosphere. To do this one needs to know the half-life of ${}_6C^{14}$, which is 5715 years.

Solution

Radioactive decay is governed by the ODE $y' = ky$. By separation and integration (where t is time and y_0 is the initial ratio of ${}_6C^{14}$ to ${}_6C^{12}$)

$$\frac{dy}{y} = k\,dt, \quad \ln|y| = kt + c, \quad y = y_0 e^{kt}$$

Next we use the half-life $H = 5715$ to determine k. When $t = H$, half of the original substance is still present, thus

$$y_0 e^{kH} = 0.5 y_0 \,. \qquad e^{kH} = 0.5 \,. \qquad k = \frac{\ln 0.5}{H} = -\frac{0.693}{5715} = -0.0001213 \,.$$

Finally, we use the ration 52.5% for determining the time t when Ötzi died (actually was killed),

$$e^{kt} = e^{-0.0001213t} = 0.525, \quad t = \frac{\ln 0.525}{-0.0001213} = 5312 \qquad \text{Answer: 5300 years ago}$$