Figures and captions – The Scroll-maggedon in Modern Publications
PSYC 6135

Xianze Meng (Marcus)

It has always been a joy to immerse oneself in a paper with its rigorous logic, detailed methods, captivating figures, and compelling argument flow. This intellectual journey can be likened to the pleasure of listening to classical music, where the ideas and statements harmonize with the beauty of exploring the frontiers of knowledge. However, this joy can be marred when the author refers to a figure not present when the reader needs it the most. The situation worsens when the figures, crucial for understanding the paper, are accompanied by detailed captions that are nowhere to be found on the neighbouring pages! The pleasant journey suddenly halted with the attempt to piece together the argument from the paper with the misplaced visual aid several pages away. The constant page changes just drive people crazy! The frustration and urge to swear at the editors peak at that moment, "Why couldn't they put them ON THE SAME PAGE???!!"

Figure and caption history

From the earliest attempts to convey information beyond text alone and onto graphic form, the elegance and clarity of graphics have been equally important to the reader's comprehension, ensuring effective communication. Pre-historic diagrams relied on graphic representations to record and recall vivid information for future use; some graphics even gradually transformed into a systematic language (Video 1)! With the innovation of technology and scientific theories, people began to rely on graphics to explain measurements and naturalistic trends, such as Galileo's recordings of sunspot movements (Fig. 1); the supplementary texts live closely with the pictures to describe the contents that were inaccessible with graphical representations solely (e.g. explaining abbreviations, copyright attributions, further clarification of asterisks. Publication Manual of the American Psychological Association (7th Ed.)., 2020). It has been almost implicit that people comprehend the supplemental information as a part of the visualizations. Galileo's figures would be hardly understandable without the supporting texts, which would waste his exquisite, delicate, hand drawing and brilliant mind. The well-known cliché, "a picture worth a thousand words," is not always true when the audience can not fully grasp the picture alone! The figure captions bridge the reader to the niche, technical figures with straightforward language.

The exact origin of figure captions is nowhere to be found, but the development of captions goes hand-in-hand with the growing technology. Before the age of the printing press, a visual attribute could be used to recreate a scenario for education other than expressing art, in which a caption is needed in verbal or written form. The ancient Egyptians decorated the tomb of Menna with paintings of farming, hunting, and banqueting scenes with hieroglyph texts, serving purposes similar to modern figure captions (Fig. 2[1]; Strudwick & Strudwick, 1999). The "captions" would

[1] A picture of the the wall of the tomb of Manna, the Blue writing near the top were hieroglyph texts.

largely be hand-written in this era in blanks or alongside the figures.

With the invention of the printing press and the growth of scientific communications after the 15th century, the figures and their captions also innovated. The figures became more concise and standardized to reflect the increased information, with the captions printed and placed next to the figures instead of occupying the margins (Fig. 3; Playfair, 1805). In this era, with the rise of analytic geometry and the bloom in physics, chemistry, and mathematics, visualizations deviated from representing actual measurements and facts to showing theories, trends, and statistics with figures instead of mere replications (Hollister & Pang, 2022). Abstract theories in the 1800s, such as Probability theory, uncertainty, and binomial distributions, were well-supported by figures to enhance scientific communication and education. Compared to solving the formula, complex mathematical formulas, such as the representation of the least squares to find the appropriate slope in the bivariate plot, could be easily explained with figures and associated captions (Fig. 4[2]). The first scientific journal, the *Journal des Scavans*, was also birthed in the 18th century. It encouraged authors and publishers to include more standardized figures in their writing to support their scientific communications (Singleton, 2014).

Computers, coloured displays, and advanced graphing software propelled figures and captions into a new age of data visualizations. Automatic processes enabled higher-quality, less costly, and more customizable figures, meaning people could create more visualizations and encompass more information than ever before.

Figures in modern scientific publications mainly serve three purposes: 1) describing the experimental setup/pipeline for the study; 2) demonstrating the outcome/contrast of the experimental subjects; and 3) graphical representation of the results, often in the form of statistical graphs. However, the added complexity of advanced statistical methods and variable plots encompassed much more information within each figure, making them almost impossible to read without heavy captioning (Shah & Hoeffner, 2002). As the figures and captioning gained importance, complexity, and size, they posed a heavy perceptual challenge to the reader's (already chaotic) working memory when digesting the main text (Aguilar & Baek, 2019).

What is wrong with modern figures and caption placement?

Scientific publications across different disciplines are becoming more complex as they pack more and more experiments into one single manuscript due to the increased understanding of previous research, growing needs for scientific communication, and elevated competition for exposure and impact. The trend is that researchers show a cohesive model in one publication, which encompasses results from different experiments. In some disciplines, there is also the rigorous need to validate the prior experiment before initiating the novel one, and every current result

---

[2] The left panel was from Wikipaedia's screenshot, functions provided by Rencher & Christensen, 2012 Williams, 2016. The right panel was from CUEMATH website (CUEMATH, n.d.).

shall be confirmed with more experiments. With the explosion of visualization quantity, a common way to organize the figures is assembling the “sub-figures” into a comprehensive figure panel for a single experiment and pairing it with heavy captioning. The readability and page organization dilemma becomes problematic when most academic publishers format their journals. The increased information forced the text font, such as axis labels and numbers, to be far smaller and more difficult to read, so the whole figure had to be enlarged for appropriate text size (Stewart et al., 2009). The bloated figures do not play well with the page formatting and captioning. Frequently, the figures would be found several pages away from where it was referred to in the main text (Fig. 5a[3]). Furthermore, the expanded and more complicated figure content usually requires more captioning to convey the message, which takes up even more page space to squeeze in the main text (Fig. 5b). The figures were beautiful in this publication, but still, trying to scroll back and forth made them frustrating to view.

When decomposing an academic paper, the disconnected text-figure-caption placement imposes an unnecessary perceptual challenge on the already-high cognitive load. The reader has to scroll back and forth between multiple pages to establish the link between text and graphic information, with high working memory utilization to hold onto the current thought. Even though resistance to distraction is an integral part of the autonomic executive functioning upon goal-directed behaviour, such as reading a paper, it is based on an efficient visual search with a fixated target in mind (Lavie & Cox, 1997; Lorenc et al., 2021). Repeated scrolling introduces perceptive discontinuity when encoding a stream of visual stimuli in different modalities – that is, it is a perceptual hurdle to comprehend figures when scrolling between text and figures is required (D’Addario & Donmez, 2019).

Moreover, the scrolling behaviour itself could act as a distractor, hindering comprehension and interpretability (Little et al., 2017). After initial encoding, the stimuli would be stored in a “buffer” area called working memory. The perceptual break introduced by repeated scrolling forces the reader to hold onto the viewed information into working memory, contaminating visual perception when acquiring new information by inducing interference and decreasing reading efficiency (Kang et al., 2011). The cognitive reserve to perceive novel visual information is identical to enabling visual working memory representation, which burdens cognition and makes it more prone to distraction (Kang et al., 2011; Wais et al., 2012).

From the perspective of working memory capacity, constant clearing and updating of working memory content are not beneficial for information chunking, a crucial preprocessing step before long-term storage. Chunking is the recoding of smaller working memory content into a meaningful whole, often regarded as bypassing the limited capacity of working memory (Oberauer et al., 2016; Thalmann et al., 2019). When the encoding behaviour is segmented by scrolling, chunking becomes significantly more difficult (Zacks, 2010). The encoding processes between each scroll could be regarded as discrete events, and a “boundary” could be

[3] The blue arrows denote Fig. 5a; the orange box denotes Fig. 5b. (Zhang et al., 2024)

introduced between them to hinder chunking (McGatlin et al., 2018; Radvansky & Zacks, 2017).

Physical books and manuscripts in Portable Document Format (PDF) suffer most from this formatting issue, as they cannot easily place the figure side-by-side with the text and caption. Less recent papers were also less plagued by this issue as 1) they have fewer and more straightforward figures, and 2) the publishers may have reformatted them when converting them into electronic media (Fig. 6[4]).

Currently, the study on the influence of distractors on graphs, especially academic figures, are lacking. Visual distraction studies more commonly use visual field distractors or auditory distractors to discuss visual performance (Banbury et al., 2001; Gao et al., 2019). Flipping the pages is more of an autonomic motor behaviour while reading a paper. How excess motor behaviour becomes a distractor and influences visual comprehension remains unknown. This study aims to systematically quantify the amount of distraction that happens to unnecessary motor behaviour and possibly propose a solution.

Study design

The proposed study will aim to recruit 150 participants from third or fourth-year undergraduate students with molecular biology and mice study backgrounds. Only participants between 18-28 years of age and similar test scores from upper-year molecular biology classes will be included. Participants with reading comprehension difficulties, visual abnormalities, and neurological disorders will also be excluded. This study aims to recruit participants with a similar background knowledge level to ensure that the topic in the Zhang et al. (2024) paper is not entirely unfamiliar. The age limit also ensured the participants would have similar distraction resiliency as cognitive aging, which has been shown to impact the ability to ignore distraction severely, even with healthy aging (McNab et al., 2015). The participants could be recruited online for greater exposure and convenience as long as a protected, distractor-free experimental environment can be ensured.

The participants would be given the Zhang et al. (2024) paper and instructed to read and comprehend the paper as much as possible, but with no aids allowed to write notes. Three individually-completed paper knowledge tests will be given to the participants, covering the paper's experimental results at three different time points – 1 hour, 2 hours, and 4 hours. Each test for an individual participant will consist of 50 questions regarding the whole paper, but different versions will be given across different tests to avoid question familiarity. The participants will be told to digest the paper as best as possible and will be tested three times. Correctness of the question will not be disclosed to the participants until the end of the study session to avoid feedback-based learning. The aim is to examine the mean group comprehension level, as well as the score increase across groups.

The Zhang et al. (2024) paper mentioned before is an ideal stimuli candidate for its poor figure and caption placement in the PDF format. It requires repeated

[4] (Kimble, 1968)

page flipping and is paired with a well-designed webpage where all the figures and captions are placed in the "correct" places with no page break. The control group will be given the webpage version of the paper, and the first experimental group will be given the PDF version. The third version would receive a version similar to the webpage, but no figure would be displayed by default. The figure will only appear after the reader clicks on the mentions in the main text, and only the corresponding sub-figure will be shown with captions. For example, if Figure 3a is referred to in Zhang et al. (2024), only Figure 3a will be shown as a floating panel if the participant clicks the hyperlink with appropriate captions (Fig. 6). The participant can choose to show or hide the figure as they desired. This proposal believes that reducing the figure size also helps alleviate the heavy cognitive load on the participants and reduces distraction by the other sub-figures, which increases visual comprehension; hence, all the figures in this proposal are attempted to be represented in this method, but unfortunately limited by the capability of Google Docs. All participants will be using an iPad to view the paper.

The participants' test scores will be collected to compare mean comprehension levels between groups and learning curves. The participants were not expected to fully comprehend this paper within 4 hours without viable aids, avoiding the ceiling effect. This study predicts that the control group will yield the lowest paper comprehension level, measured by the group mean of test scores at all time points; their learning will also be the slowest, measured by subtracting scores between different tests and then taking the group mean. Two experimental groups will yield significantly better comprehension levels and learning curves, with the second experimental group performing better.

Possibilities of improvement?

The main aim of improvement will be to reduce the perceptual break introduced by constant page flipping. The first and most straightforward solution would be to avoid page-flipping while reading. If the reader does not read the mentioned Zhang et al. (2024) paper with the more common PDF but on the original webpage, the figures and captions are very nicely formatted alongside each corresponding section. There is also a convenient button to retract the figure caption if it is unnecessary. The supplementary video for schematic and real-time experimental comparison video is also an excellent resource for understanding this study without introducing perceptive breaks.

The solution of split screening, which refers to having a separate window open with the figures alone, only marginally reduces page flips. Again, with bloated figures occupying the majority of the screen, it is hard to situate the main text in an appropriate space. It would need constant adjustments to lock in the correct sub-figures with the main text, which is also a distraction. The split screening does not match the effective visual searchlight achieved by the large perceptive field after laying different pages simultaneously on the table (Atkinson & Braddick, 1989; Schwarzkopf et al., 2019). This proposal aims to improve visual perception and journal comprehensibility on electronic media.

The authors could also introduce interactive figures to reduce figure size and cognitive load. With current software enhancements such as Shiny in R studio and coverage and bandwidth increases on the Internet, an interactive figure could be easily incorporated into manuscripts. Ancker et al. (2011) showed the potential of interactive graphs to demonstrate elevated interpretability in complex, abstract data. The reduced graph density also helps reduce the cognitive load upon complex plots, as the interactive plots could appear one at a time or simultaneously to increase contrast. When there are fewer items in the visual searchlight, the attention span and efficiency can also increase the interpretability and memorability of a graph (Etzel et al., 2013; Soni et al., 2018).

The recent breakthrough in generative AI modelling could also be incorporated into complicated graph comprehension to generate autonomic figure captions. Eye tracking could also monitor the reader’s progress on the publication, enabling customized captions tailored to the reader's knowledge level. For example, the caption would focus more on explaining the figure on the initial encounter and be increasingly informative about its implications and linkage with other studies on repeated reads.

Limitations and Implications

At the current stage, due to the lack of systemic measurement of the perceptual “distance” between the figure and the main text, the current project would most likely be a pilot study to measure visual perception with procedural distractors. The individual differences in reading ability and reading order introduced many unavoidable confounds of the experimental procedure, which prevented this project from receiving generalizable results other than trends. Although relatively large, the sample is unrepresentative due to the background knowledge required to comprehend a complex academic publication. This study would need simpler stimuli, more generalizable procedural distractors, and more straightforward measures for visual comprehension to yield persuasive results, especially when some publishers were still hesitant about colouring their figures. The increased readability in figures should increase the citation quantity of a paper and enhance the impact factor for the journal.

Page flipping is only an iconic example of this greater “procedural distractor” class, with promising potential to explore. At the very least, this project will serve as a guide for journal publishers to improve page design and increase interactability. It could also help the author to design easier-to-read figures. On the broader side, distraction studies have been heavily implemented to improve road safety by introducing stringent laws on cell phone use and vehicle and road designs (Chan & Singhal, 2013). The development of artificial intelligence could also benefit from distraction studies to improve machine vision and mimic human reactions to a distractor (Bidollahkhani et al., 2024).

Overall, improving the page arrangement of modern journals could benefit everyone. Sometimes, the editors need a bellringer to remind them that the journal they formatted could be hard to read. And it avoids students’ desperate hair-pulling or the rage to yell at the editor, “Why couldn’t they put them ON THE SAME PAGE???!!”

References

Aguilar, S. J., & Baek, C. (2019). Motivated Information Seeking and Graph Comprehension Among College Students. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 280–289. https://doi.org/10.1145/3303772.3303805

Ancker, J. S., Weber, E. U., & Kukafka, R. (2011). Effects of Game-Like Interactive Graphics on Risk Perceptions and Decisions. *Medical Decision Making*, *31*(1), 130–142. https://doi.org/10.1177/0272989X10364847

Atkinson, J., & Braddick, O. J. (1989). 'Where' and 'What' in Visual Search. *Perception*, *18*(2), 181–189. https://doi.org/10.1068/p180181

Banbury, S. P., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory Distraction and Short-Term Memory: Phenomena and Practical Implications. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(1), 12–29. https://doi.org/10.1518/001872001775992462

Bidollahkhani, M., Raahemi, M., & Haskul, P. (2024). Distracted AI: Integrating Neuroscience-Inspired Attention and Distraction Learning in ANN. *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*, 1–8. https://doi.org/10.1109/AISP61396.2024.10475279

Chan, M., & Singhal, A. (2013). The emotional side of cognitive distraction: Implications for road safety. *Accident Analysis & Prevention*, *50*, 147–154. https://doi.org/10.1016/j.aap.2012.04.004

Chou, M. (Director). (2022, February). *The Story of Chinese Character: 月*. https://www.youtube.com/watch?v=ZIWmEqnIaMM

CUEMATH. (n.d.). *Least Square Method*. CUEMATH. https://www.cuemath.com/data/least-squares/

D'Addario, P., & Donmez, B. (2019). The effect of cognitive distraction on perception-response time to unexpected abrupt and gradually onset roadway hazards. *Accident Analysis & Prevention*, *127*, 177–185. https://doi.org/10.1016/j.aap.2019.03.003

Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: Promise, pitfalls, and potential. *NeuroImage*, *78*, 261–269. https://doi.org/10.1016/j.neuroimage.2013.03.041

Friendly, M., & Denis, D. J. (2022). *1800-1849: Beginnings of modern data graphics*. Milestones in the History of Thematic Cartography, Statisticial Graphics, and Data Visualization. https://datavis.ca/milestones/index.php?group=1800%2B

Galilei, G., Reeves, E. A., Van Helden, A., & Scheiner, C. (2010). *On sunspots*. University of Chicago Press.

Gao, Y., Bing, L., Li, P., King, I., & Lyu, M. R. (2019). Generating Distractors for Reading Comprehension Questions from Real Examinations. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 6423–6430. https://doi.org/10.1609/aaai.v33i01.33016423

Kang, M.-S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, *18*(5), 860–869. https://doi.org/10.3758/s13423-011-0126-5

Kimble, D. P. (1968). Hippocampus and internal inhibition. *Psychological Bulletin*, *70*(5), 285–295. https://doi.org/10.1037/h0026470

Lavie, N., & Cox, S. (1997). On the Efficiency of Visual Selective Attention: Efficient Visual Search Leads to Inefficient Distractor Rejection. *Psychological Science*, *8*(5), 395–396. https://doi.org/10.1111/j.1467-9280.1997.tb00432.x

Little, D. F., Zhang, Y.-X., & Wright, B. A. (2017). Disruption of Perceptual Learning by a Brief Practice Break. *Current Biology*, *27*(23), 3699-3705.e3. https://doi.org/10.1016/j.cub.2017.10.032

Lorenc, E. S., Mallett, R., & Lewis-Peacock, J. A. (2021). Distraction in Visual Working Memory: Resistance is Not Futile. *Trends in Cognitive Sciences*, *25*(3), 228–239. https://doi.org/10.1016/j.tics.2020.12.004

McGatlin, K. C., Newberry, K. M., & Bailey, H. R. (2018). Temporal Chunking Makes Life's Events More Memorable. *Open Psychology*, *1*(1), 94–105. https://doi.org/10.1515/psych-2018-0007

McNab, F., Zeidman, P., Rutledge, R. B., Smittenaar, P., Brown, H. R., Adams, R. A., & Dolan, R. J. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings of the National Academy of Sciences*, *112*(20), 6515–6518. https://doi.org/10.1073/pnas.1504162112

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Publication manual of the American Psychological Association (7th ed.). (2020). American Psychological Association. https://doi.org/10.1037/0000165-000

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, *17*, 133–140. https://doi.org/10.1016/j.cobeha.2017.08.006

Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis* (Third Edition). Wiley.

Schwarzkopf, D. S., Moutsiana, C., & Panesar, G. (2019). *Validation of population receptive field estimates in human visual cortex*. https://doi.org/10.31219/osf.io/479cr

Shah, P., & Hoeffner, J. (2002). Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review*, *14*(1), 47–69. https://doi.org/10.1023/A:1013180410169

Singleton, A. (2014). The first scientific journal. *Learned Publishing*, *27*(1), 2–4. https://doi.org/10.1087/20140101

Soni, U., Lu, Y., Hansen, B., Purchase, H. C., Kobourov, S., & Maciejewski, R. (2018). The Perception of Graph Properties in Graph Layouts. *Computer Graphics Forum*, *37*(3), 169–181. https://doi.org/10.1111/cgf.13410

Stewart, B. M., Cipolla, J. M., & Best, L. A. (2009). Extraneous information and graph comprehension: Implications for effective design choices. *Campus-Wide Information Systems*, *26*(3), 191–200. https://doi.org/10.1108/10650740910967375

Strudwick, N., & Strudwick, H. (1999). *Thebes in Egypt: A guide to the tombs and temples of ancient Luxor*. Cornelll University Press.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 37–55. https://doi.org/10.1037/xlm0000578

Wais, P. E., Martin, G. M., & Gazzaley, A. (2012). The impact of visual distraction on episodic retrieval in older adults. *Brain Research*, *1430*, 78–85. https://doi.org/10.1016/j.brainres.2011.10.048

Wieck, R. S. (1988). *Time sanctified: The Book of hours in medieval art and life* (1st ed). G. Braziller in association with the Walters Art Gallery, Baltimore.

Williams, J. H. (2016). *Quantifying measurement: The tyranny of numbers*. Morgan & Claypool Publishers.

Yu, H., Agarwal, S., Johnston, M., & Cohen, A. (2009). Are figure legends sufficient? Evaluating the contribution of associated text to biomedical figure comprehension. *Journal of Biomedical Discovery and Collaboration*, *4*(1), 1. https://doi.org/10.1186/1747-5333-4-1

Zacks, J. M. (2010). How we organize our experience into events. *Psychological Science Agenda*, *24*(4).

Zhang, X.-Y., Wu, W.-X., Shen, L.-P., Ji, M.-J., Zhao, P.-F., Yu, L., Yin, J., Xie, S.-T., Xie, Y.-Y., Zhang, Y.-X., Li, H.-Z., Zhang, Q.-P., Yan, C., Wang, F., De Zeeuw, C. I., Wang, J.-J., & Zhu, J.-N. (2024). A role for the cerebellum in motor-triggered alleviation of anxiety. *Neuron*, *112*(7), 1165-1181.e8. https://doi.org/10.1016/j.neuron.2024.01.007