

Exploring Effectiveness of Explanations for Appropriate Trust: Lessons from Cognitive Psychology

Ruben S. Verhagen**

Siddharth Mehrotra*†

Mark A. Neerincx‡

Catholijn M. Jonker§

Myrthe L. Tielman¶

Delft University of Technology, The Netherlands

ABSTRACT

The rapid development of Artificial Intelligence (AI) requires developers and designers of AI systems to focus on the collaboration between humans and machines. AI explanations of system behavior and reasoning are vital for effective collaboration by fostering appropriate trust, ensuring understanding, and addressing issues of fairness and bias. However, various contextual and subjective factors can influence an AI system explanation’s effectiveness. This work draws inspiration from findings in cognitive psychology to understand how effective explanations can be designed. We identify four components to which explanation designers can pay special attention: perception, semantics, intent, and user & context. We illustrate the use of these four explanation components with an example of estimating food calories by combining text with visuals, probabilities with exemplars, and intent communication with both user and context in mind. We propose that the significant challenge for effective AI explanations is an additional step between explanation generation using algorithms not producing interpretable explanations and explanation communication. We believe this extra step will benefit from carefully considering the four explanation components outlined in our work, which can positively affect the explanation’s effectiveness.

Index Terms: Human-centered computing—Visualization—Visualization techniques; Human-centered computing—Visualization—Visualization design and evaluation methods

1 INTRODUCTION

Humans and Artificial Intelligence (AI) systems are increasingly collaborating on tasks ranging from the medical to the financial domain. For such human-machine collaboration to be effective, mutual understanding and trust are of paramount importance [15, 18, 31]. AI explanations are a crucial and powerful way to increase human understanding and trust in the system. Explanations can help in explaining the decisions and inner workings of the “black-box” data-driven machine learning algorithms and actions & reasoning of goal-driven agents [2, 10, 11, 19, 25, 27].

Unfortunately, AI systems often lack transparency and explainability, providing a broad call for explainable AI (XAI). For example, EU GDPR requires organizations that deploy AI systems to provide relevant information to affected people about the inner workings of the algorithms [37]. Furthermore, in addition to helping people

understand AI systems, AI explanations can also contribute to identifying and addressing issues of fairness and bias, which are difficult to do without. Explanations can also help form appropriate trust in the AI system. For example, Dodge et al. [8] pointed out that when people trust the explanation, they are more likely to trust the underlying ML system.

A common challenge in popular explanation methods such as LIME (Local Interpretable Model-agnostic Explanations) by Ribeiro et al. [29], SHAP (Shapley Additive Explanations) by Lundberg & Lee [24], and global-local explanation by Lundberg et al. [23] is the effectiveness of the explanation to help users calibrate their trust and increase understanding of the explanation. Various aspects can impact the effectiveness of AI explanations, such as the user’s domain, system expertise or cognitive and perceptual biases. However, this list is far from conclusive. Therefore it is crucial to investigate the following two research questions:

1. *What should be the content of the explanation to make it effective?*
2. *How can the visual delivery or design of the explanation make it effective?*

2 EXPLANATION EFFECTIVENESS

In this work, we take inspiration from cognitive psychology, design, and data visualization techniques to explore ways how effective explanations can be designed. A large amount of work in cognitive psychology focuses on explanations in human-human interactions, what makes them effective and how context can affect the same. For example, Lombrozo [22] provides two properties of the structure of explanations that helps in reasoning, (1) explanations accommodate novel information in the context of prior beliefs, and (2) do so in a way that fosters generalization. The author showcase that explanations provides a unique window onto the mechanisms of learning and inference in human reasoning.

David B. Leake in his book “Evaluating Explanations” from cognitive psychology theories has described how context involving both explainer beliefs and goals can help in deciding an explanation’s goodness [21]. Similarly, Khemlani et al. shows how mental models represent causal assertions, and how these models underlie deductive, inductive, and abductive reasoning yielding effective explanations [17]. Since explanations determine how humans understand the world in fundamental terms, Tworek and Cimpian helps in exploring human biases towards the inheritance of judgments in people’s explanations over socio-moral understanding [35].

The previously mentioned works and recent research in human-AI interaction [12, 26, 32, 36, 43] helps us in exploring our two research questions based on four components (Perception, Semantics, Intent, and User & Context). These components are uniquely visible at the intersection of studies in human-AI interaction and cognitive psychology. We will now describe these components in detail.

2.1 Perception

Perception in XAI can refer to the set of mental processes we use to make sense of a given explanation by an AI system. Perception

*e-mail: r.s.verhagen@tudelft.nl; *Author sharing first authorship.

†e-mail: s.mehrotra@tudelft.nl; *Author sharing first authorship.

‡e-mail:m.a.neerincx@tudelft.nl; Secondary affiliation: TNO, Soesterberg, The Netherlands

§e-mail:c.m.jonker@tudelft.nl; Secondary affiliation: Leiden University, The Netherlands

¶e-mail:m.l.tielman@tudelft.nl

is also frequently related to the modality in which explanation is presented and how our brain interprets it. For instance, example-based explanations are useful in cases where it's hard to explain AI reasoning. Here, developers can create an explanation in the form of a real-life example or a story to depict the system's processing, in line with how we as humans are primed to relate to real-life examples or episodes. Empirical research has demonstrated that when explanations take the form of a story, they can help humans in decision making [16]. According to Rumelhart [30], stories often contain initiating events, goals, actions, consequences, and accompanying states in a particular causal configuration. Knowing the structure of stories can allow humans to form appropriate trust concerning the perception of the explanation.

For saliency maps, categorical model confidence visualizations, n-best visualizations, and related ways of visual explanations, we propose that it is important to consider how the human brain is organized to see structures, logic, and patterns. Our brain organization helps us as humans to make sense of the world. A seminal work by Wertheimer [39] introduced seven Gestalt principles of visual perception in the form of heuristics. We now briefly describe those principles in relation to visual explanations in XAI:

1. **figure-ground**: visualization of the explanation can display objects as being either in the foreground or the background. For example, an identified object with a higher confidence score can be displayed in the foreground, leaving every other object in the background.
2. **similarity**: clustering of similar objects can be shown to the user such that the user can easily group similar objects together. For example, we can use a variety of design elements, like color and organization, to establish similar groups.
3. **proximity**: for pixel-wise importance in saliency map of human face detection, calculated using integrated gradients; explanations can focus on (a) the areas around the eyes, (b) lines anywhere on the face, and (c) regions around the mouth and nose. This relative nearness of the objects can strongly influence the user's understanding of the explanation.
4. **common region**: the principle of common region states that when objects are located within the same closed region, humans perceive them as being grouped together. Therefore, for any two distinct objects, a visual explanation can put them in different regions and vice-versa.
5. **continuity and closure**: continuity refers to the arrangement of elements while closure refers to identification of recognizable patterns. We can apply these principles in explanations presented in a commonly used dashboard. For example, these principles explain why only two axes, rather than a full enclosure, are required on a graph to define the space in which the data appears. Also, it can become obvious which groups belong to the subgroup when the hierarchy is expanded.
6. **focal point**: the premise of the focal point principle is that anything visually outstanding first captures and retains the viewer's attention. In terms of explanations, assigning each point a score reflecting its contribution to the ML model-recognition loss can follow this principle. The saliency map can then explicitly explain which points are the key to model recognition, as explained by Zheng et al. [42].

Zena O'Connor's contemporary work on colour, and contrast in combination with gestalt theories provide the effectiveness of visual communication design in graphical design, which can be helpful for the XAI community [28]. For example, the author examined the role of colour and contrast within the context of Gestalt theories of

perception. The author also showcased how colour and contrast not only help to distinguish contours, detail, and depth, but they also help to attract and divert attention, thereby clearly delineating key areas of text. This result can help design explanations when multiple models are used to describe a decision by an AI system.

2.2 Semantics

Semantics in XAI refers to how AI system designers can convey the meaning in language to their users. This meaning of the language can help users get a grip on what the expressions of a natural language contribute to the overall process of interpretation. Cann et al. [6] provide three basic types of sense relations to clearly communicate the meaning of the language:

1. **Synonymy**: sameness of sense; for example, X's pullover is yellow v/s X's sweater is yellow.
2. **Hyponymy**: sense inclusion; for example, X wants a hamburger v/s X wants to eat a hamburger.
3. **Antonymy**: oppositeness in sense; for example, This water is cold v/s This water is not hot.

It becomes essential to be aware of these sense relations when designing explanations for AI systems because as soon as AI system designers start probing what concept of meaning they should articulate, they may slip away into a quag of open-endedness.

Researchers have focused on resolving linguistic ambiguities in visual data communication. For example, Setlur et al. [33] provide heuristics to resolve partial utterances based on syntactic and semantic constraints. Similarly, Law et al. [20] have proposed ways such as encouraging skepticism and open-mindedness to reduce the illusion of causality when using question-answering systems. Similarly, Gaba et al. [9] have provided guidelines for designing natural language interfaces and recommendation tools to better support natural language comparisons in visual communication. These guidelines include use of basic charts as reasonable responses for a variety of comparisons, providing necessary information useful for the comparisons, and exposing the provenance of how implicit entities are such as "cheap" and "best-selling".

Humans often adapt cognitive approaches to explanations where any ambiguity and vagueness can affect human-AI communication. This fuzziness in communication can result in distrust in the AI system and can be disastrous in high-risk scenarios where lives can be at stake. Taken into consideration the fuzziness of explanations, Nauta et al. [26] provide a summary of Co-12 properties on explanation quality. The properties related to semantics (and perception) fall in the category of content focusing on correctness, completeness, contrastivity, continuity, consistency, and covariate complexity. These properties provide an aggregated view of what to evaluate in an explanation. We believe that an explanation's cohesiveness ties all these properties together. In cohesiveness, one can focus on creating semantically resonant explanations that enrich human understanding. When an explanation is cohesive it can also help in preserving users' mental models and memory. For example, providing infographic information with confidence scores based on reasoning with facts from verified and trusted sources can help make an explanation cohesive. Furthermore, exploring concepts around scaling or hiding information can create the right balance for explanation effectiveness.

Apart from focusing on effective explanations qualities, it is also essential to consider how data literacy can impact the perceived effectiveness of an explanation. Wolf's and Gottwald's popular work on *Tales of Literacy* sheds light on how digital literacy is differently represented in the human brain [40]. The authors present "vocabulary" of language and reading development to appreciate better the critical importance of specific literacy development aspects

with digital systems worldwide. On a similar note, Setlur and Cogley [32] provide an overview of data literacy and how it forces us to think about how everyone navigates the world. Authors call out to researchers working in this domain to focus on cultural differences in data literacy and how people around the globe can interpret an explanation in multiple ways.

Any recommendation in form of an explanation by an AI system is based on a probabilistic model. One of the challenges for effective explanations is identifying a baseline for communicating these probabilities. Based on the work by Wolf [41], we propose that ability of neural networks in the brain helps and hinders humans in their attempt to read and process information related to probabilities. Since our way of processing information differs at the individual level, so does our understanding of an explanation. Therefore, it becomes vital to consider diverse populations and the moral components of the design and use of visualizations for explanations in an easy to understand manner [7].

We believe that AI systems will not be explainable without first addressing the scarcity of semantics in explanations. Browne and Swift [4] highlights that “*producing satisfactory explanations for deep learning systems will require that we find ways to interpret the semantics of hidden layer representations in deep neural networks*”. There have been a few attempts looking at semantics in explanations, such as by Jin et al. [14] exploring hierarchical visualization of compositional semantics to help users create trust in deep neural networks. Similarly, Jacob et al. [13] investigating how users can guide the generation of counterfactual explanations by specifying a set of semantic regions of the query image. Although both reported works show promising results, extracting semantic explanations from hidden units is far from a solved problem.

2.3 Intent

The intent in XAI refers to the purpose of the explanation, ideally balancing semantics with goals. Therefore, intent can heavily influence the perception and semantics components when designing explanations. For example, explanations aimed at convincing users to engage in a healthier lifestyle should take different forms than explanations aimed at improving firefighters’ situation in high stake decisions. For recommending food options to a user to keep a healthy lifestyle, explanations can be designed to (a) convince users with confidence estimates of the nutrition of food, (b) predict the outcome of eating the healthy food, or (c) showcase low confidence when a system can’t recommend a healthy diet based on available food option to foster appropriate trust. In contrast, explanations aimed at improving the situation and moral awareness of firefighters can explain the potential consequences of the most important situational features to consider when making a decision.

From the literature of cognitive psychology, JS Bruner [5] sheds light on Macfarlane’s work, where authors studied how mothers are irresistibly imputing intent to the cries, gestures, expressions, and postures of newborns. As communication presupposes intent in communicating an objective, one can learn so much just by understanding a mother’s intent to understand her child. Bruner also highlights that in interpreting the infant’s communicative intent - correctly or incorrectly - the mother has a rich variety of cues to use, and so too for the child. Relating this to the XAI context, we propose that the AI system designers need to adapt their system designs to match the user’s intent based on available cues, similar to how a mother does for her child, as outlined by Bruner.

Neerinx et al. and Anjomshoae et al. [2, 27] provide three explanation phases to convey intent of an explanation: generation, communication, and reception. Explanation *generation* mostly concerns with the explainability methods used to construct relevant elements and information to share, such as SHAP or LIME explainers. Explanation *communication* refers to the form and content and *reception* relates to how the receiving user utilizes and understands

the explanation. In the work by Miller [25], the author did emphasize the crucial role of intention in explanations, arguing that determining the goal of an explanation is key to providing a good explanation. We resonate with this statement and propose that the current field of XAI can pay more attention to explanation intent. Specifically, intent can guide explanation communication (i.e., content/semantics and form/perception), which can show its benefits in explanation reception.

2.4 User and Context

We believe that defining explanation intent includes considering both user and context. Recent works emphasize the importance of personalized and context-aware explanations [2, 3, 25], as it is believed that they improve explanation effectiveness. For example, the same explanation used to convince user A to live healthier will likely not work if user B has a completely different value profile or personality. Likewise, when a firefighter is under extreme time pressure, visual explanations are probably much more effective than verbal ones. Taken together, we believe that simply defining explanation intent as to inform, convince, or support will not suffice. Instead, explanation designers can define the goal by combining explanation intent, receiver/user, and context (e.g., *convince conservative user X to start exercising more*).

3 USE CASE EXAMPLE

Wang and Yin provided three desiderata for designing effective AI explanations [38]. These desiderata include (a) designing explanations to improve people’s understanding of the AI model, (b) helping people recognize the uncertainty underlying an AI prediction, and (c) empowering people to trust the AI appropriately. In the following example of an AI explanation, we follow these desiderata together with the components of explanations that we discussed earlier.

Our use case example showcases a snapshot of the “estimating calories on a plate” user study, see Figure 1. We designed the task around nutrition as an approachable domain for everyone. In this user study, an AI agent helps the user to estimate the food calories with the help of explanations. The AI agent estimates the calories of a dish shown in the right corner of the figure. The agent also explains its confidence scores (based on an AI-driven food recognition model). In addition, the categorical visualization for each ingredient is displayed as a form of visual explanation. The user must decide among the four options to select the closest calorie count based on the AI agent’s suggestion and get a +10 score for a correct selection, whereas a wrong selection costs -10 points.

In the above use case, the confidence scores of ingredients are generated using a machine learning classifier in real-time. The visual explanations are based on categorical visualization (Figure 1, right) inspired by the example provided in the Google PAIR guidebook [1]. These visual explanations were designed keeping in mind the four components discussed in this paper as follows:

1. In our use case, **perception** refers to different categorizations such as best, good, or unsure match by providing clear meaning of the categorizations to the user.
2. For **intent**, we decided to use a single image for each categorization to convey the actual semantics we perceived as system designers. The text-based explanations in the figure are hand-crafted following the notion of situation vignettes [34]. Here every explanation starts with the intent the AI agent wants to convey, followed by the perception of the authority (in this case United Nations Food & Nutrition website) on which the suggestion is based.
3. The next part of the explanation is to clearly deliver the **semantics** of the reasoning behind the suggestion to the user in

Team Score: 0 points

Round 1

Question: What is the total calorie count on the food plate?

Step 2

Click on the below button to ask the AI agent & carefully read it's explanation!

Ask the Agent

I have selected 750 calories as the answer for this question. I think it is important to be transparent, so I'll tell you how I came to this decision.

I found a similar dish based on my training data from United Nations Food & Nutrition website that closely matches with the given plate for the calories count. The dish I found is a curry however I am not sure about which curry it is. The matching visualization is shown next to the identified ingredients. Based on my training data and similar dish search, the total amount of calories on should be 750 calories which is similar to the best match example.

300
 750
 900
 1100

Select this box if you think that the AI agent is better able to estimate the calories than yourself.

Go to Step 2

Ingredient	Confidence Score
meat	0.9904504
vegetable	0.9806911
beef	0.96175015
pork	0.94082177
sauce	0.92987496
pepper	0.9046277
tomato	0.8762977
soup	0.84438676
chicken	0.83761346
curry	0.8109094

Best Match
Good Match
Match unsure

Figure 1: In this snapshot of a user study, an AI agent provides an estimation of calories of the dish (shown in bottom right) out of four available options (shown in bottom left). Here the estimation of calories is accompanied by confidence scores of each identified ingredient(s) in a form of a table. Also, the AI agent provides both textual and visual explanations for its reasoning of selecting an option for calories estimation.

non-ambiguous language. For semantics, we adopted Rumelhart's [30] guidelines and followed gestalt heuristics for designing visual explanation.

- The **user and context** are also taken into account by relating the explanation with the closest dish in the database and providing logic-based reasoning that users can adhere to. Furthermore, the combination of visual and textual explanation is designed to elicit an appropriate amount of trust that the AI agent can foster, distinct from promoting the trust in the agent.

Our method of showcasing explanations is a way to incorporate the components that we have discussed in this work. This approach has known limitations, such as scaling the explanations with complex data or crafting these explanations in real-time. We believe that these limitations can be overcome with further advancements in the field of XAI with methods such as ensemble ML, feature-based, rule-based, and training-data based explanations.

4 CONCLUSION

Previous work has shown that the efficiency of human-AI interaction can be improved when AI systems explain their behavior and reasoning. However, how to create explanations that are both intuitive and easily understandable by users is a challenging task. In this work, we have discussed four components of designing effective explanations based on insights from cognitive psychology. Furthermore, we provided an example to incorporate those components in creating a textual and visual explanation for a use-case example. With this work, we propose the need to provide emphasis on these components to make explanations effective. We also showed how communicating results in an interpretable manner needs to be as intuitive as dialogue. However, we must note that explanations can increase the

chance that humans will accept the AI's recommendation, regardless of its correctness. Therefore, explanations should strive to create appropriate trust rather than helping promote trust in the AI system. Also, we highlight that it is important to avoid explanations biasing users towards certain actions or decisions, especially in high-stakes decision-making tasks.

ACKNOWLEDGMENTS

The authors wish to thank the three anonymous reviewers for their constructive comments and review of the manuscript.

REFERENCES

- [1] Google pair. people + ai guidebook, May 2019.
- [2] S. Anjomshoae, A. Najjar, D. Calvaresi, and K. Främling. Explainable agents and robots: Results from a systematic literature review. In *18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019*, pp. 1078–1088. International Foundation for Autonomous Agents and Multiagent Systems, 2019.
- [3] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115, 2020.
- [4] K. Browne and B. Swift. Semantics and explanation: why counterfactual explanations produce adversarial examples in deep neural networks. *arXiv preprint arXiv:2012.10076*, 2020.
- [5] J. S. Bruner. From communication to language—a psychological perspective. *Cognition*, 3(3):255–287, 1974.
- [6] R. Cann, R. Kempson, and E. Gregoromichelaki. *Semantics: An Introduction to Meaning in Language*. Cambridge University Press, 2009.

- [7] M. Correll. Ethical dimensions of visualization research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.
- [8] J. Dodge, Q. V. Liao, Y. Zhang, R. K. Bellamy, and C. Dugan. Explaining models: an empirical study of how explanations impact fairness judgment. In *Proceedings of the 24th international conference on intelligent user interfaces*, pp. 275–285, 2019.
- [9] A. Gaba, V. Setlur, A. Srinivasan, J. Hoffswell, and C. Xiong. Comparison conundrum and the chamber of visualizations: An exploration of how language influences visual design. *arXiv preprint arXiv:2208.03785*, 2022.
- [10] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Comput. Surv.*, 51(5), Aug. 2018. doi: 10.1145/3236009
- [11] D. Gunning. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web*, 2(2), 2017.
- [12] M. Hearst, M. Tory, and V. Setlur. Toward interface defaults for vague modifiers in natural language interfaces for visual analysis. In *2019 IEEE Visualization Conference (VIS)*, pp. 21–25. IEEE, 2019.
- [13] P. Jacob, É. Zablocki, H. Ben-Younes, M. Chen, P. Pérez, and M. Cord. Steerx: Steering counterfactual explanations with semantics. *arXiv preprint arXiv:2111.09094*, 2021.
- [14] X. Jin, Z. Wei, J. Du, X. Xue, and X. Ren. Towards hierarchical importance attribution: Explaining compositional semantics for neural sequence models. In *International Conference on Learning Representations*, 2019.
- [15] M. Johnson and A. Vera. No ai is an island: The case for teaming intelligence. *AI Magazine*, 40(1):16–28, Mar. 2019. doi: 10.1609/aimag.v40i1.2842
- [16] D. H. Jonassen. Designing for decision making. *Educational technology research and development*, 60(2):341–359, 2012.
- [17] S. S. Khemlani, A. K. Barbey, and P. N. Johnson-Laird. Causal reasoning with mental models. *Frontiers in human neuroscience*, 8:849, 2014.
- [18] G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich. Ten challenges for making automation a “team player” in joint human-agent activity. *IEEE Intelligent Systems*, 19(6):91–95, 2004. doi: 10.1109/MIS.2004.74
- [19] P. Langley, B. Meadows, M. Sridharan, and D. Choi. Explainable agency for intelligent autonomous systems. In *AAAI*, vol. 17, pp. 4762–4763, 2017.
- [20] P.-M. Law, L. Y.-H. Lo, A. Endert, J. Stasko, and H. Qu. Causal perception in question-answering systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2021.
- [21] D. B. Leake. *Evaluating Explanations: A Content Theory*. Psychology Press, 1992.
- [22] T. Lombrozo. The structure and function of explanations. *Trends in cognitive sciences*, 10(10):464–470, 2006.
- [23] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67, 2020.
- [24] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’ 17*, p. 4768–4777. Curran Associates Inc., Red Hook, NY, USA, 2017.
- [25] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019. doi: 10.1016/j.artint.2018.07.007
- [26] M. Nauta, J. Trienes, S. Pathak, E. Nguyen, M. Peters, Y. Schmitt, J. Schlötterer, M. van Keulen, and C. Seifert. From anecdotal evidence to quantitative evaluation methods: A systematic review on evaluating explainable ai. *arXiv preprint arXiv: 2201.08164*, 2022.
- [27] M. A. Neerincx, J. van der Waa, F. Kaptein, and J. van Diggelen. Using perceptual and cognitive explanations for enhanced human-agent team performance. In D. Harris, ed., *Engineering Psychology and Cognitive Ergonomics*, pp. 204–214. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-91122-9_18
- [28] Z. O’Connor. Colour, contrast and gestalt theories of perception: The impact in contemporary visual communications design. *Color Research & Application*, 40(1):85–92, 2015.
- [29] M. T. Ribeiro, S. Singh, and C. Guestrin. “why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pp. 1135–1144, 2016.
- [30] D. Rumelhart. *Understanding and Summarizing Brief Stories*. Basic processes in reading: Perception and comprehension, 1979.
- [31] E. Salas, D. E. Sims, and C. S. Burke. Is there a “big five” in teamwork? *Small Group Research*, 36(5):555–599, 2005. doi: 10.1177/1046496405277134
- [32] V. Setlur and B. Cogley. *Functional Aesthetics for Data Visualization*. Wiley, 2022.
- [33] V. Setlur, M. Tory, and A. Djalali. Inferencing underspecified natural language utterances in visual analysis. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 40–51, 2019.
- [34] M. Strack and C. Gennerich. Personal and situational values predict ethical reasoning. *Europe’s Journal of Psychology*, 7(3):419–442, 2011.
- [35] C. M. Tworek and A. Cimpian. Why do people tend to infer “ought” from “is”? the role of biases in explanation. *Psychological Science*, 27(8):1109–1122, 2016. PMID: 27485133. doi: 10.1177/0956797616650875
- [36] G. Vilone and L. Longo. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76:89–106, 2021.
- [37] P. Voigt and A. Von dem Bussche. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 10(3152676):10–5555, 2017.
- [38] X. Wang and M. Yin. Effects of explanations in ai-assisted decision making: Principles and comparisons. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2022.
- [39] M. Wertheimer. Gestalt theory. 1938.
- [40] M. Wolf and S. Gottwald. *Tales of Literacy for the 21st Century*. Oxford University Press, 2016.
- [41] M. Wolf and C. J. Stoodley. *Proust and the squid: The story and science of the reading brain*. Harper Perennial New York, 2008.
- [42] T. Zheng, C. Chen, J. Yuan, B. Li, and K. Ren. Pointcloud saliency maps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1598–1606, 2019.
- [43] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger. Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics*, 10(5):593, 2021.