

What ‘Work’ Can Dataviz Do in Popular Science Communication?

Martin Engebretsen^a

Department of Nordic and Media Studies, University of Agder, Universitetsveien 25, Kristiansand, Norway


Keywords: Data Visualization, Popular Science Communication, Social Semiotics, Discourse Analysis, Design Practices

Abstract: Data visualizations have proliferated on public arenas for information and communication – in journalism, PR and governmental information as well as in popular science communication (PSC). In the existing literature on PSC, *simplicity*, *relevance* and *trust* are identified as critical factors for the communication to succeed. This position paper argue that data visualizations represent a semiotic resource with unique potentials regarding all of these criteria. The paper aims at presenting a theoretical and methodological framework for studying data visualizations applied in popular science discourses. The main goals are a) to introduce social semiotics as an advanced analytical tool for the scrutiny of data visualizations, b) to introduce PUS (Public understanding of science) as a field relevant for empirical studies of data visualization, and c) to present a method of analysis combining a small scale corpus analysis with multimodal close reading of selected visualizations.

1 INTRODUCTION

Data visualization (DV) represents a form of visual communication well integrated in scientific discourses. In disciplines where quantitative data play a central role – particularly in the natural and social sciences – graphs, charts and maps represent an irreplaceable semiotic resource for revealing patterns and relations in large amount of data (Tufte, 2001; Few, 2012). In the professional scientific discourses, the codes and conventions related to the production and interpretation of DVs are strong, and normally shared by the producer and the reader. In public dissemination of science, on the other hand, both verbal and visual forms of communication need to be adapted to the needs of non-scientists, who share neither verbal terminologies with the scientists, nor their forms of visual-numeric literacy. In this paper we ask: What role do DVs play in popular science communication (PSC)? What are their semiotic and social functions in the discourse? And how do they apply to the norms and expectations of the genre in which they are embedded?

During the latest decades, data visualizations have proliferated on public arenas for information and communication – in journalism, PR, governmental information etc. (cf. Cairo, 2013; Kennedy et al., 2016a and b; Engebretsen, 2017). This development is driven by, among other factors, a growing access to data from a variety of sources and a rapid development of cheap and easy-to-use visualization tools (Engebretsen et al., 2018). However, on these arenas, DVs meet a heterogenous group of readers, representing big variations concerning visual and numeric literacy (D’Ignazio & Klein, 2016; Pinney, Forthcoming). Advanced DVs, with complex visual codes and high density of variables and values, face a high risk of being misunderstood or neglected by members of the general public. On the other hand, a creative, surprising and visually attractive DV may receive a lot of attention and stimulate the motivation for further reading (cf. Cairo, 2016; Allen, 2018). What is regarded as best practices of data visualization in the public, is celebrated in awards like the *Malofiej Award*, the *Data Journalism Award*, and the *Kantar Information is Beautiful Award*¹.

^a  <https://orcid.org/0000-0001-8002-1423>

¹ See: <https://www.malofiejgraphics.com/>;
<https://datajournalismawards.org/> and
<https://www.informationisbeautifulawards.com/>

In the existing literature on public understanding of science (PUS), certain critical factors are identified regarding popular science communication. These values relate to *simplicity*, *relevance* and *trust* (cf. Bauer, 2009). Complex scientific problems and findings need to be simplified, in order to be understood by non-scientists. Further, these findings need to be contextualized and related to issues important to the general reader, in order to gain attention and emotional engagement. And, finally, the chosen forms of communication need to evoke trust, both in the communicator – be it a news organization or a professional mediator – and in the scientific source behind the popularized presentation. The argument underlying this paper, is that data visualizations represent a text type with particular affordances regarding both simplicity, relevance and trust. Simplification of complexity follows with the reductional nature of DVs, being the result of a number of choices made by the producer through the pipeline of production. Relevance and emotional engagement can be evoked by the message carried by the data itself, for instance when a chart shows the development of criminal incidents in a specific urban area. It can also be evoked by the contextualization provided by the verbal co-text, or by the interactive options offered in certain digital contexts. Finally; trust can be built through transparency regarding the production process; opening the “black box”. Such transparency can be achieved by informing broadly about the scientific sources and the data sources behind the DV, the methods used in the production of it, as well as the institution and the persons behind the popular version (Engebretsen, 2017).

This position paper aims at presenting a theoretical and methodological framework for studying DVs in popular science discourses. The main goals are a) to introduce social semiotics as an advanced analytical tool for the scrutiny of data visualization as a multimodal text type situated in particular social practices, b) to introduce PUS (public understanding of science) as a field relevant for empirical, discourse oriented studies of data visualization, and c) to present a method of analysis combining a small scale corpus analysis with multimodal close reading of selected DVs. The results of the study will be particularly relevant for science journalists and others working with science communication, as well as students and researchers in the field of information visualization and the field of Science and Technology in Society (STS).

2 THEORETICAL FRAMEWORK

The main framework for the planned study is social semiotic discourse analysis (SSDA). This approach to texts and other meaningful cultural artefacts can be traced back to the Australian linguist Michael Halliday (Halliday, 1978). His theories of meaning-making through verbal utterances have been adapted to multimodal and mediated contexts, e.g. by Gunther Kress and Theo van Leeuwen (Kress and van Leeuwen, 2006; van Leeuwen, 2005). The study is also inspired by Norman Fairclough's model of critical discourse analysis (Fairclough, 2012), aiming at revealing the social and political impact of specific textual choices. This theoretical framework invites an analytical approach first focusing on semiotic, meaning-making structures in the visualization itself, and then on the role that the DV has in the textual whole in which the DV is embedded. Finally, the findings on these semiotic-technological levels are discussed with regards to the norms and conventions of the specific genre and to the socio-cultural practice in which the DV is an integrated element (see more about methodology below).

A core concept in social semiotic theory, is that of the three metafunctions. The metafunctions refer to the ways in which an expression relates to different aspects of the context: to the experienced world, to the participants in the discourse and to the textual and technological resources applied in the production of the message. These metafunctions correspond to three types of meaning. *Experiential meaning* concerns what the text can say about (an aspect of) the world. *Inter-personal meaning* concerns the construction of social relations between the participants in the discourse – e.g. between the producer and the reader of a DV as well as between the reader and the persons/groups represented by the DV. Finally, *compositional meaning* concerns the ways in which semiotic and media-technological resources are used to create wholeness and coherence in the textual output (Halliday, 1978; Kress & van Leeuwen, 2006). All of these three dimensions of meaning are realized as semiotic potentials (although not necessarily identified by all readers), in any instance of meaningful expression – be it a 5-word e-message or a 50-minute documentary film – but they are obviously realized in very different ways. A relevant question is thus; how are the three metafunctions realized in DVs applied in public science communication, and how do they contribute to understanding, engagement and trust?

In a similar study conducted in 2017, the framework of SSDA was applied in an analysis of 17

journalistic data visualizations, collected from four major Norwegian news sites (Engebretsen, 2017). The study revealed that many of the DVs were designed in a way that made complex patterns in the data material easy to perceive, and thus supported the processing of experiential meaning. Interpersonal meaning potentials were, on the other hand, less focused. Only eight of the 17 DVs provided any information about the producers behind the visualizations, and only four offered substantial information regarding the methods used in the process of production. 10 of the DVs offered elements of interactivity, inviting the readers to explore the underlying data by themselves. This study provides a relevant model for analysis, although the DVs were collected from a different domain of public discourse than the PSC-discourse described in this position paper.

The other part of the theoretical framework for the study, is that of public understanding of science (PUS). PUS is a field of activity as well as an area of social research, closely related to the wider field of Science, Technology and Society (STS). PUS-related studies include the building of theory and models, qualitative case studies as well as surveys and other quantitative studies of science communication taking place in a range of public genres; science blogs, science journalism, popular science magazines, museum exhibitions etc. The PUS-discourse has historically, according to Bauer (2009), focused on three different problems. In the 60s and 70s, the focus was on a deficit of knowledge in the general audience. In the 80s, the dominating concern was a deficit of attention, interest and support of science in the general public. Since the 90s, much attention in the PUS-discourse is given to the lack of trust to science as well as to the media. The discourse of fake news (having grown in intensity in the Trump-period), is a symptom (or a driver) of the deficit of trust to the media. To illustrate the issue of low trust in science, one can look to Norway, a nation with a high educational attainment,² yet, with a very high density of climate skeptics. In the period 2013-2018, between 22 and 27 per cent of the adult Norwegian population were skeptical to the idea that climate changes are related to human activities, in spite of what is massively communicated by scientists.³

Modelling the interaction between the “esoteric” scientific discourses going on among scientists, and the “exoteric” discourses of science on the public and

private arenas, Bauer – referring to Flecks core-periphery-model from 1937 (see Fleck, 1979) – states that the further away from the esoteric core, the exoteric discourses are characterized by a growing gradient of simplification, concreteness and certainty of judgement. In other words; in a highly popularized presentation of a scientific result, one must expect to find a higher degree of simplification, of visual illustration and of certainty (i.e. a lack of reservations and modifications) than what is expected in e.g. a textbook in higher education. Bauer calls for more discourse-oriented studies – as a complement to the far more frequent quantitative studies – in future investigations of the PUS dynamics, where the relationships between the esoteric and the (different levels of) exoteric discourses of science ought to be closely investigated.

Some commentators are less concerned about public deficits regarding knowledge, interest and trust in their approach to science communication, and more concerned about dialogue and active participation. They call for a less top-down and more interactive, mutual and dialogical view on the interaction between scientists, science communicators and the public audience (Riise, 2008; Santerre, 2008). Davies & Horst (2016) model science communication as a non-hierarchic ecosystem, and point to its large complexity of actors, epistemologies and discursive elements. In a dialogical, non-hierarchic approach to science communication, the style of expression and the interpersonal dimension of meaning making play a substantial role in the construction of the participants’ identities and their discursive roles and power.

In the intersection between these two frameworks, where the analytical tools of social semiotic theory are focused by core issues in the field of PUS, we can extract a more nuanced set of research questions in our study of data visualization in science communication, building on the broad and general questions formulated in the initial paragraph. We now ask:

- What characterizes the visual codes applied in DVs in successful PSC to inform about aspects of the world?⁴
 - What DV types are used? What visual codes, metaphors, forms and colors are applied? What is the level of information density?

² According to National Statistics Institute of Norway, 34 per cent of the Norwegian population have education on a bachelor’s level or higher. <https://www.ssb.no/en/utdanning/statistikker/utniv>

³ <https://www.bt.no/btmeneringer/debatt/i/LALy5V/slik-erde-norske-klimaskeptikerne>

⁴ In this paper, «successful PSC» refers to award winning instances of Popular Science Communication.

- What role do these DVs have with regards to the overall message of the texts in which they are embedded?
 - Do they carry core information? Do they document, illustrate, tell stories? Are elements of uncertainty represented verbally or visually?
- What characterizes the interplay between DV and co-text regarding the construction of identities, trust and discursive roles?
 - Are the DVs related to information concerning sources and methods? Are there any invitations for reader activity and/or dialogue?
- How do the findings related to the questions above apply to the current norms and expectations of popular science communication? Do they contribute to simplicity, engagement and trust?

3 METHODOLOGY

The method suggested to answer the questions above, is that of social semiotic discourse analysis (SSDA). This form of analysis belongs to the scientific paradigm of hermeneutics, linking descriptive, interpretive, and critical perspectives to generate a framework aimed at understanding how a certain semiotic artefact 'works' in a certain socio-cultural context (cf. van Leeuwen, 2005; Ledin & Machin, 2018). In SSDA, the analyst will always relate a close reading of selected elements on the micro level (the semiotic structures) to relevant elements on the macro level (the socio-cultural context) in order to understand the social functions of the semiotic choices made by the text producer. Sometimes the socio-cultural context is investigated directly, through observation, interviews etc.; sometimes it is investigated indirectly, through textual implication, other literary sources or general knowledge. The approach to the collected data follows an abductive method, combining the deductive use of certain theoretical perspectives as a starting point for the data analysis, with an inductive openness to the unique features of the investigated objects and their social contexts (Richardson & Kramer, 2006).

The first stage in a SSDA is most often to gather a selection of samples from the field under scrutiny (Aiello, forthcoming). The samples can be described according to their formal characteristics on a relevant level of detail. In the next stage, a small or large number of samples are analyzed with focus on the realization of the three metafunctions. The number of samples is dependent on the research questions as

well as the resources allocated to the study. When dealing with a larger corpus, a pilot-study with close reading of 2-3 samples can reveal what features that are most relevant to include in a more formalized and systematic study of the whole corpus. (cf. Engebretsen 2017)

The final stage of the analysis includes a discussion of the findings in relation to the socio-cultural context that they belong to. What semiotic and social work do the artefacts do? Do they answer to the norms and quality criteria of the genre? Do they represent change and innovation? Do they affect issues related to power, democracy and equality? This is the stage where the most important questions are dealt with, although the answers – as always in qualitative research – necessarily will involve interpretations and perspectivals on the side of the researcher.

3.1 Material for Analysis

In the suggested study, the analyst will follow the three stages described above, with a focus on the issues formulated in the four research questions. The corpus material belongs to the category of "best practices". It consists of 20 DVs, found in 20 price-winning instances of popular science communication. The price-winners are collected from three different awards, all relevant to the PUS-discourse.

Both static and interactive DVs will be included in the sample. The difference between a static (explanatory) and an interactive (exploratory) DV may affect all three of the semiotic metafunctions mentioned in the theory section. E.g.; a static DV may have a stronger narrative power, while an interactive DV may lead to a more active and emotionally engaged reading process.

The selection of samples is obviously not representative of all DVs applied in PSC. Thus, the findings cannot be taken as indicative of the total population. On the contrary, they most probably stand out from their population, which is the reason for their status as price-winners. However, this status also gives them the function of being models for other DV designers. Close reading of a 'golden sample' is particularly relevant in a genre with such a rapid development concerning visual and technological design. What wins prices today, will most probably affect the mainstream of tomorrow.

ACKNOWLEDGEMENTS

The research reported in this article was supported by The Norwegian Research Council (NFR).

REFERENCES

- Aiello, G., Forthcoming. Inventorizing, situating, transforming: Social semiotics and data visualization. In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society*, Amsterdam University Press. Amsterdam.
- Allen, W., 2018. Visual brokerage: Communicating data and research through visualization. In *Public Understanding of Science*, 27(8).
- Bauer, M.W., 2009. The Evolution of Public Understanding of Science – Discourse and Comparative Evidence. In *Science, Technology & Society* 14:2, pp. 221–240. DOI: 10.1177/097172180901400202
- Cairo, A., 2013. *The functional art. An introduction to information graphics and visualization*, New Riders. Berkeley, CA.
- Cairo, A., 2016. *The truthful art. Data, charts, and maps for communication*, New Riders. Berkeley, CA.
- Davies, S. & Horst, M., 2016. *Science Communication. Culture, Identity and Citizenship*, Palgrave Macmillan. London.
- D’Ignazio, C. & Klein, L., 2016. Feminist Data Visualization. In Proceedings from the Workshop on Visualization for the Digital Humanities at IEEE VIS Conference, Baltimore, Maryland.
- Engebretsen, M., 2017. Levende diagrammer og zoombare kart. Datavisualisering som nyskapende fortellerform i journalistikken. [Living diagrams and zoomable maps. Data visualization as an innovative form of storytelling in journalism]. In *Norsk Medietidsskrift* 02/2017.
- Engebretsen, M.; Kennedy, H. & Weber, W., 2018. Data Visualization in Scandinavian Newsrooms. Emerging Trends in Journalistic Visualization Practices. In *Nordicom Review* 39(2). DOI:10.2478/nor-2018-0007.
- Few, S., 2012. *Show me the numbers. Designing tables and graphs to enlighten*, Analytics. Oakland, CA. 2nd ed.
- Fairclough, N., 2012. *Critical discourse analysis: the critical study of language*, Routledge. London.
- Fleck L., 1979 [1935]. *Entstehung einer wissenschaftlichen Tatsache*, Suhrkamp. Frankfurt
- Halliday, M. A. K., 1978. *Language as a social semiotic. The social interpretation of language and meaning*, Arnold. London.
- Kennedy, H., Hill, R., Allen, W. og Kirk, A, 2016a. Engaging with (big) data visualisations. Factors that effect engagement and resulting new definitions of effectiveness. In *First Monday*, vol. 21 (11)
- Kennedy, H., Hill, R., Aiello, G. & Allen, W., 2016b. The work that visualisation conventions do. In *Information, Communication and Society*, vol. 19 (6).
- Kress, G., & van Leeuwen, T., 2006. *Reading images: The grammar of visual design*, Routledge. London, 2nd ed.
- Ledin, P. & Machin, D., 2018. *Doing Visual Analysis. From Theory to Practice*, SAGE. LA, London.
- Pinney, L., Forthcoming. Is Literacy What We Need in an Unequal Data Society? In M. Engebretsen & H. Kennedy (Eds.), *Data visualization in society*, Amsterdam University Press. Amsterdam.
- Richardson, R. & Kramer, E.H., 2006. Abduction as the type of inference that characterizes the development of a grounded theory. In *Qualitative Research* 2006; 6/4, pp. 497–513.
- Riise, J., 2008. Bringing Science to the Public. In D. Cheng et al. (Eds.) *Communicating Science in Social Contexts*. Springer Science+Business Media B.V. Pp 301-309.
- Santerre, L., 2008. From Democratization of Knowledge to Bridge Building between Science, Technology and Society. In D. Cheng et al. (Eds.) *Communicating Science in Social Contexts*. Springer Science+Business Media B.V. Pp. 289-300.
- Tufte, E. R., 2001. *The Visual Display of Quantitative Information*, Graphics Press. Cheshire, CT. 2nd Ed.
- van Leeuwen, T., 2005. *Introducing Social Semiotics*, Routledge. Abingdon, New York.