Quantitative Approaches

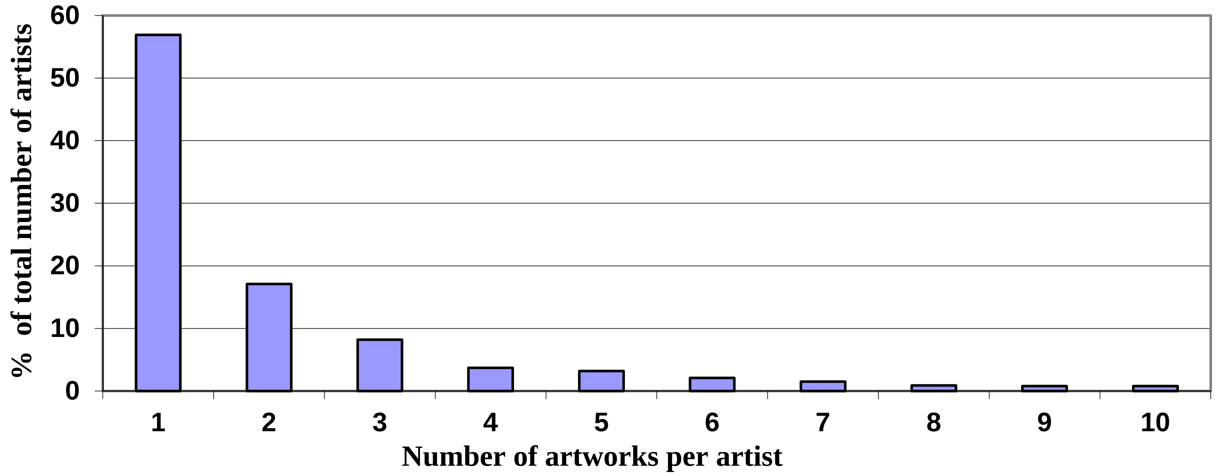


Figure 1: Plot of the number of artworks per artist in sample 1

Peer-Reviewed

Distant Viewing in Art History A Case Study of Artistic Productivity

K. Bender

Abstract: With reference to the concept of distant reading in literary history, distant viewing is a valuable analogy for a quantitative approach to art history. In this case study of artistic productivity eight samples are analyzed, extracted from a digital thematic research collection about the iconography of Aphrodite/Venus from the Middle Ages to Modern Times. The result is an empirical finding of regularity never before highlighted in art history. The artistic productivity fits perfectly the distribution known as Lotka's law of scientific productivity in bibliographic science. Issues of collecting and sampling are discussed and the meaning of this empirical finding is hinted. Suggestions for future research are made.

Keywords: artistic/scientific productivity, distant reading/viewing, Lotka, quantitative art history

Introduction[1]

The French philosopher Michel Tournier[2] discusses the difference between quantity and quality and cites a quotation:

Sans doute la qualité vaut mieux que la quantité, mais sur la qualité,
on peut discuter à l'infini, tandis que la quantité, elle, est indiscutable.

Edward Reinrot[3]

Franco Moretti[4], who initiated the concept of 'distant reading' in literary history, made the same statement in other words: "Quantitative research provides a type of data which is ideally independent of interpretations...". Moretti argues that literature isn't a 'sum of individual cases', but a 'collective system'. Scholars have focused on a select group of texts: the canon. In 'distant reading' the canon disappears into the larger literary system.

These arguments are equally valuable for art history, where traditionally 'quality' matters more than 'quantity' and monographs focus predominantly on works considered as the great masterpieces of art. However, quantitative data such as the number of replicas, of engravings and subsequent duplications or imitations by other artists are gaining greater attention, and studies about the economics and market related aspects of art production are increasingly popular. The analysis of the spreading and popularity of motifs and style also requires 'numbers'.

This quantitative aspect of art history needs specific types of data acquisition. Structured data collections, alongside standard bibliographies, are crucial for advanced quantitative studies[5].

Reference works and reports provide evidence about the increasing importance of quantitative data, generating new forms of knowledge in the digital age of art history[6]. They can be analyzed computationally, as demonstrated for example in the pioneering work of Schich and Ebert-Schifferer[7], a trend following innovative research in literary history and therefore termed 'distant viewing in art history'.

This paper presents a case study about artistic productivity with a distribution known in bibliographic science as Lotka's law. All data, extracted from a digital thematic research collection, have been published and are freely available. Hence, the results presented in this paper are verifiable and the data could be used to explore alternative models of productivity in art history.

The case study

The productivity in terms of number of artworks created by an artist has been examined with the help of eight samples. The samples are taken from a digital thematic research collection compiled for a project of topical catalogues of the iconography of the Greek-Roman goddess Aphrodite/Venus, depicted in sculptures, paintings, drawings, prints and illustrations from the Middle Ages to Modern Times[8]. The topical categorization in these catalogues is mutually exclusive: no work is listed more than once. This methodology allows for quantitative analyses of the popularity of topics, of the time distributions of works and artists and of the number of works per artist[9].

In the first sample of 1840 works by 649 identified Italian artists, the average number of works per artist is 2,8. However, the counting of works per artist shows a very unequal 'productivity': a large majority (57 %) of all artists created only one 'Venus'-work in a lifetime, only 17 % made two works, 8% made three works, 3,5% made four works, 3% made five works, 2% made six works, etc. ... 0,8% made 10 works as shown in Fig.1 (number of works per artist on the horizontal axis and percentage number of artists on the vertical axis)[10].

All other samples in this project yield identical distributions as explained below. This empirical finding has never before been highlighted in art history.

Analogy with Lotka's law of scientific productivity

The American statistician Alfred J. Lotka published in the Journal of Washington Academy of Science, 1926, an article 'The frequency distribution of scientific productivity' based on an analysis of publications by authors in two fields of the exact sciences. Potter[11] reveals "...that Lotka's article was not cited until 1941, that his distribution was not termed "Lotka's law" until 1949, and that no attempts were made to test the applicability of Lotka's law to other disciplines until 1973'.

Lotka found that the number of authors producing ***x*** publications is about

$$1/x^a$$

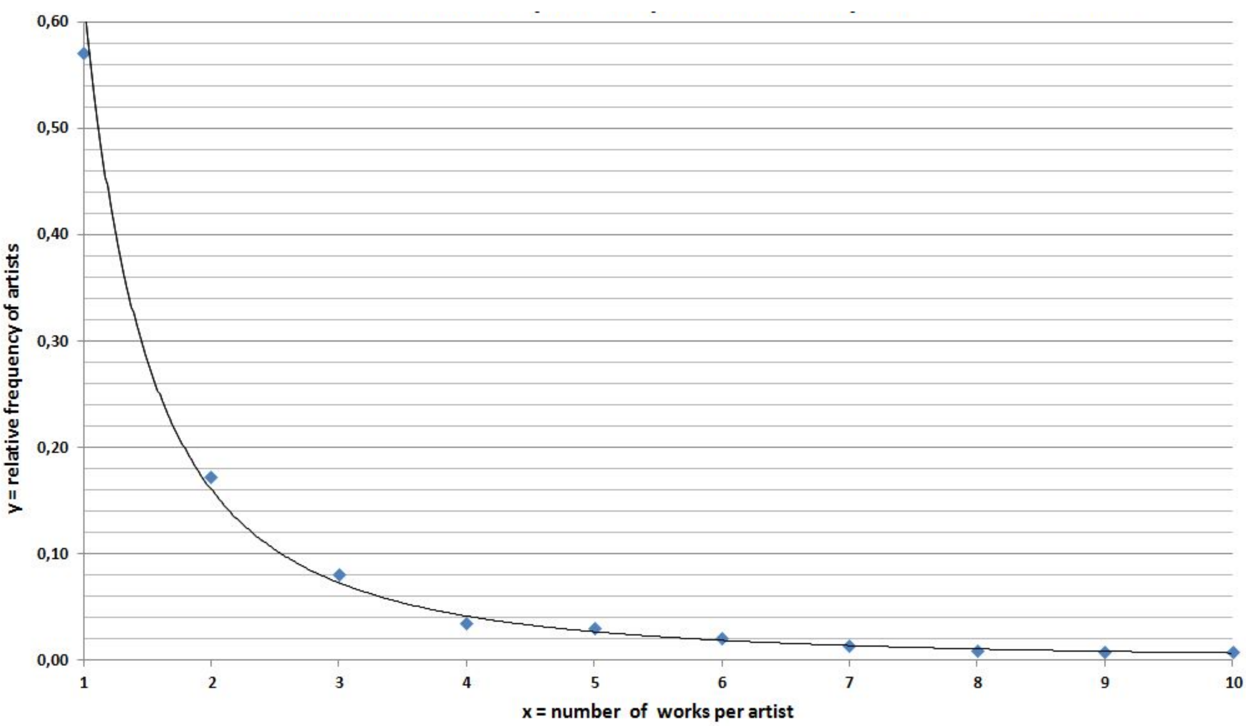
of those making one publication, or:

$$y=C/x^a$$

where ***y*** is the relative frequency (or proportional number) of authors with ***x*** publications and the constant ***C*** and the exponent ***a*** are parameters. Thus for ***x = 1, C = y***.

This is an inverse power function, now commonly referred to as 'Lotka's law'[12]. Lotka suggested that the exponent ***a*** nearly always equals ***2*** and the function can then be called an inverse square function. This means that the number of

Figure 2: Observed frequencies of number of works per artist and fitted inverse power equation for Sample 1
$y=0{,}6222\ x^{-1{,}948}$ $R^2 = 0{,}9929$



authors making 2 publications is $1 / 2{*}2 = 1 / 4 = 0.25$ of those making 1 publication; those making 3 publications: $1 / 3{*}3 = 1 / 9 = 0.11$ of those making 1 publication, etc. This surprisingly resembles the distribution as shown in Fig.1. Hence, it was a logic step to try out Lotka's law with the data of our first sample. By logarithmic transformation of the data and using the classical linear regression technique or 'least squares method', applied for instance automatically in the trend-line functionality in 'charts' of Microsoft Office Excel 2007, we can estimate the values of the parameters:

$$C = 0{,}6222 \text{ and } a = 1{,}948$$

and calculate a goodness-of-fit measure between the equation and the observed data, commonly called the correlation coefficient R^2 (with $0 < R^2 < 1$; the closer R^2 is to 1, the better fit):

$$R^2 = 0{,}9929$$

Thus the result of this test, plotted in Fig.2, shows a close resemblance to Lotka's law with an exponent a very near to the suggested value 2[13].

Further evidence for all samples

The next step was to analyze the data of all samples in the project compiled with the same methodology as sample 1. The data are extracted from the publications by Bender[14]. The basic data (N = total number of artists; X = total number of works; $\underline{x} = X/N$ average number of works per artist) for the eight samples are presented in Table 1 and the observed data of number n of artworks and relative frequency y of artists for each sample are given in Table 2.

Table 1: Basic data of the samples
(N = total number of artists; X = total number of works; average number of works $\underline{x}$ = X/N)

sample	***N***	***X***	***$\underline{x}$***	***country of artist's origin***	***references***
1	649	1840	2,8	Italy	9a
2	977	2997	3,1	France	9b
3	728	2636	3,6	Low Countries	9c
4	1506	3198	2,1	Germany, Switzerland, Central-Europe	9d
5	912	2113	2,3	Great Britain, Ireland	9e
6	184	291	1,5	Eastern Region	9f
7	220	503	2,3	Southern Region	
8	215	577	2,7	Northern Region	
total	***5401***	***14155***	***2,6***		

Table 2: Observed data of the samples
(x = number of works created by an artist; n = number of artists who created x works; relative frequency of artists who created x works: y = x/N; N = total number of artists)

sample		*x*									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	***n***	371	112	53	23	20	14	9	6	5	5
	y	0,5716	0,1726	0,0817	0,0354	0,0308	0,0216	0,0139	0,0092	0,0077	0,0077
2	***n***	586	158	75	31	28	19	11	8	9	9
	y	0,5998	0,1617	0,0768	0,0317	0,0286	0,0194	0,0112	0,0082	0,0092	0,0092
3	***n***	398	122	52	30	29	19	7	8	12	4
	y	0,5467	0,1676	0,0714	0,0412	0,0398	0,0261	0,0096	0,0110	0,0165	0,0055
4	***n***	1027	215	95	53	34	27	7	10	6	6
	y	0,6819	0,1428	0,0631	0,0352	0,0226	0,0179	0,0046	0,0066	0,0040	0,0040
5	***n***	628	119	65	25	20	10	7	4	4	5
	y	0,6886	0,1305	0,0713	0,0274	0,0219	0,0110	0,0077	0,0044	0,0044	0,0055
6	***n***	148	20	12	7	1	0	1	0	0	0
	y	0,7629	0,1031	0,0619	0,0361	0,0052	0,0000	0,0052	0,0000	0,0000	0,0000
7	***n***	162	30	12	7	2	1	2	1	1	0
	y	0,7364	0,1364	0,0545	0,0318	0,0091	0,0045	0,0091	0,0045	0,0045	0,0000
8	***n***	139	32	12	8	6	2	4	3	1	2
	y	0,6465	0,1488	0,0558	0,0372	0,0279	0,0093	0,0186	0,0140	0,0047	0,0093
all	***n***	**3459**	**808**	**376**	**184**	**140**	**92**	**48**	**40**	**38**	**31**
	y	**0,6404**	**0,1496**	**0,0696**	**0,0341**	**0,0259**	**0,0170**	**0,0089**	**0,0074**	**0,0070**	**0,0057**

The estimated values of the parameters ***C*** and ***a*** and the calculated goodness-of-fit measure R^2 for the individual samples vary between *0,5675* and *0,7506* for ***C***, *1,865* and *2,264* for ***a***, and *0,9423* and *0,9929* for R^2. The computation for all samples merged (last row in Table 2) yields:

$$y = 0{,}6505 / x^{2{,}089} \text{ with } R^2 = 0{,}9945$$

The estimated value of the constant ***C*** *= 0,65* is near the observed value ***y*** *= 0,64* and the value of the exponent ***a*** *= 2,089* is again very close to the value suggested by Lotka and thus the proposed inverse power function is practically an inverse square law (Fig. 3).

In this exercise the number of ***x*** has been deliberately limited to 10. However, it is known that observations of large values of ***x*** do not fit well the Lotka distribution: the so-called 'long-tail' problem. Therefore, an alternative model with three parameters was applied for samples 1, 2 and 3 when all values of ***x*** were included in the computations[15]. Though the model yielded slightly better goodness-of-fit measures R^2 , the values of ***C*** and ***a*** were not longer comparable among samples due to interaction with the third parameter and the model was discarded.

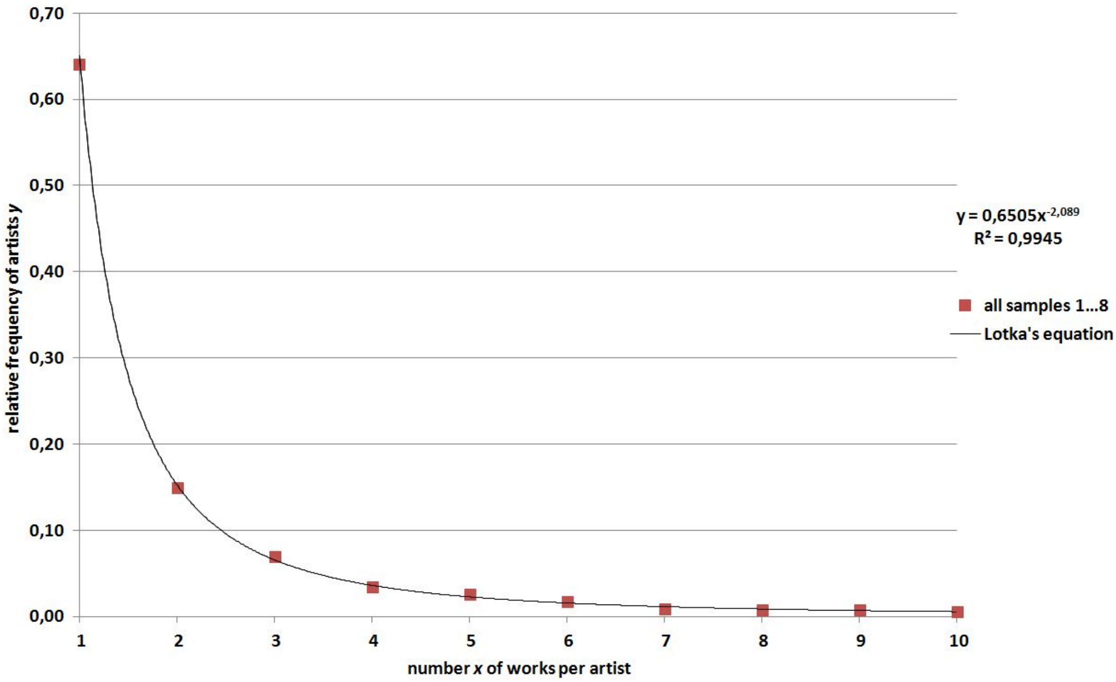


Figure 3: Observations and Lotka's law for all samples 1...8

Discussion of results

The perfect fit, with a very high value of the goodness-of-fit measure ***R²* = *0,9945***, to a large set of samples representing in total 14155 artworks, created over a period of more than 500 years by 5401 artists from all over Europe, leaves no doubt that the so-called "Lotka's law of scientific productivity" is applicable to this case study of art historical data. However, the sampling method, the thematic collection, the Lotka distribution and its 'long tail', and the meaning of the empirical finding, are issues deserving discussion and further study.

- The sampling method in this case study is not 'at random' where all artworks would have equal chance to be selected in the 'population' of the indefinite number of artworks by an unknown number of artists of a thematic research collection. In fact, the sampling is a 'convenient' one and always biased in a thematic collection because many artworks, never recorded, were lost and the information sources are limited to the collector. Hence, the representativeness and the size of the samples are always issues. The samples in this case study are presumably very large[16]; nevertheless their size can be always enlarged[17]. An advanced study of the sampling bias will eventually be performed in the future through methods of meta-analysis[18].

- More important are the formal concept of a thematic collection and its methodology of topical categorization: indeed, the series created form the basis for the distant viewing concept. Therefore, the series should be as homogeneous as possible in order to make quantification possible[19]. The fact that there is a remarkable regularity in all samples of this case study is an indication that the homogeneity of the thematic collection is high.

- Why does scientific/artistic productivity not follow the 'normal' Gaussian distribution of events that go by chance? Gaussian distribution offers an 'equal' chance to each event. Lotka's law, on the contrary, shows a very 'unequal' situation: 65% of the sources (authors/ artists) produce only 1 item (publication/work) and relatively few sources produce many items. The few artists producing more than 10 works in this case study – i.e. the so-called ‘long tail’ in the distribution – are, however, not the least known: on the contrary, many well-known masters are among the most prolific ‘Venus’-artists[20]. No doubt, this is related to problems of authenticity and attribution of the artworks as well as to the issue of workshop management of the production.

- Egghe[21] discusses at length the principle of 'success breeds success' or 'cumulative advantage' and demonstrates mathematically how it is related to Lotka's equation. The phenomenon is comparable to the economic or financial rule: 'the richer you are, the easier to get even richer'. One can interpret this as follows for the case study: there is always a probability that an artist with no 'Venus'-work in the past will create a first one. If this first 'Venus'-work has success, the greater probability will be that the artist will produce another 'Venus'-work and so on; if, however, this first work is a failure or has no success, the artist will probably not create a second 'Venus'-work. This may explain the high value of ***y = 65% for x = 1*** as well as the ‘long tail’ phenomenon of superstar artists with a large network of patrons and customers. This case study provides quantitative data for socio-economic models of creativity as discussed by Menger[22].

- In his search to find an interpretation of Lotka's law, Price discusses the basic difference between creative effort in the sciences and in the arts: “The artist's creation is intensively personal, whereas that of the scientist needs recognition by his peers” [23]. Authorship of scientific articles is therefore an indication of prestige. The finding in this case study seems to prove that this distinction is mistaken: the artistic creativity follows a similar pattern as the scientific effort and obviously has also everything to do with 'prestige'. Moreover, the way how modern research is funded through targeted programs has some similarity with preferences of art patrons and fashion on the art market.

Conclusion and suggestions for future research

The empirical finding of this case study is remarkable and its interpretation 'success breeds success' has never been highlighted before in art history. The 'distant viewing' approach of a fairly homogeneous thematic collection and the quantification of data in eight large independent samples proves successful and could be an example for future quantitative research in art history. Are there other thematic collections in art history available which comply with the conditions of homogeneity and size? If yes, then one could further test the applicability of Lotka's law for other themes or explore more sophisticated models. A better understanding of the underlying regularity could give rise to unorthodox questions and offer new ways to decipher the complexity of artistic productivity.

Notes

[1] The author gratefully acknowledges the helpful comments of the anonymous peer-reviewers of the draft paper. He also thanks Paul Taylor of The Warburg Institute, London, for discussion of Fig.1 and for drawing his attention to the analogy with Lotka's law, and Béatrice Joyeux-Prunel of the Ecole Normale Supérieure, Paris, for her support regarding socio-economic reference material.

[2] Michel Tournier, "Quantité et qualité" in Le miroir des idées – Traité (Paris : Mercure de France, 1995), 205-208.

[3] Author's translation: 'Without doubt quality is better than quantity, but quality can be discussed ad infinitum, while quantity is indisputable'. Edward Reinrot is a pseudonym of Tournier himself.

[4] Franco Moretti, Graphs, Maps, Trees – Abstract models for literary history (London, New York: Verso, 2007) 9.

[5] Among many online data collections, such as 'Bildindex Foto Marburg' http://www.bildindex.de and 'The Warburg Institute Iconographic Database' http://warburg.sas.ac.uk/photographic-collection/iconographic-database/ of a general nature, one can also cite some specific ones: * the 'Census of Antique Works of Art and Architecture known in the Renaissance', started in 1947 at the Warburg Institute, University of London, and online http://www.census.de * the 'Montias Database of 17th Century Dutch Art Inventories', developed in the '80s by the economist John Michael Montias, online at the Frick Art Reference Library http://research.frick.org/montias/home.php

[6] Béatrice Joyeux-Prunel, éd., L' Art et la Mesure - Histoire de l'art et méthodes quantitatives: sources, outils, bonnes pratiques. (Paris: Editions Rue d'Ulm, 2010) https://ens.academia.edu/B%C3%A9atriceJoyeuxPrunel/Books. Hubertus Kohle, Digitale Bildwissenschaft. (Glückstadt, Verlag Werner Hülsbusch, 2013). http://archiv.ub.uni-heidelberg.de/artdok/2185/1/Kohle_Digitale_Bildwissenschaften_2013.pdf. Matthew Long and Roger C. Schonfeld, Supporting the Changing Research Practices of Art Historians. (Ithaka S+R, 2014) http://www.sr.ithaka.org/sites/default/files/reports/SR_Support-Changing-Research-ArtHist_20140429.pdf

[7] Maximilian Schich and Sybille Ebert-Schifferer, Bildkonstruktionen bei Annibale Carracci and Caravaggio: Analyse von kunstwissenschaftlichen Databanken mit Hilfe skalierbarer Bildmatrizen (ART-Dok report, 2008) http://archiv.ub.uni-heidelberg.de/artdok/volltexte/2009/712

[8] About the relevance of the motif of Aphrodite/Venus in Western art history, the author refers to reference Caroline Arscott and Katia Scott, eds., Manifestations of Venus – Art and sexuality. (Manchester and New York: Manchester University Press, 2000).

[9] Details in several posts and especially in the series *'Statistics in Art History'* in the author's Blog *'Iconography in Art History'* http://kbender.blogspot.be/?view=magazine

[10] For practical reasons of visualization the graph is limited to 10 artworks per artist. See below about the 'long-tail' issue.
[11] William Gray Potter, "*Lotka's Law Revisited*"; Library Trends 31,2 (1981)): 21-39. https://www.ideals.illinois.edu/bitstream/handle/2142/7191/librarytrendsv30i1e_opt.pdf?sequence=1#page=1&zoom=auto,-87,590
[12] It would be better called 'Lotka's equation or distribution' since it is not a precise law, a term used in physics.
[13] The more exact 'maximum likelihood method' to estimate the parameters ***C*** and ***a*** yields similar results for all numbers ***x***: ***C = 0,6095*** and ***a = 2,0047***. Details on the author's webpage *'LOTKA's Law of Productivity'* https://sites.google.com/site/venusiconography/home/research-papers/lotka-s-law-of-productivity
[14] K. Bender: The Iconography of Venus from the Middle Ages to Modern Times. Volumes 1.1 to 6.1. (2007, 2009, 2010, 2012, 2013, 2014). Physical books published by www.lulu.com and www.shopmybook.com/en/, https://archive.org/search.php?query=K.%20bender%20Venus%20AND%20mediatype%3Atexts
[15] K. Bender, "Time Distribution, Popularity, Diversity and Productivity of the Iconography of Venus in the Low Countries, France and Italy", Research Paper 5 in the Series 'Quantitative Iconography of Venus' (2011) https://independent.academia.edu/KBender/Papers
[16] For comparison reasons: a search "Venus since the 6th century" in the above cited general collections 'Bildindex Foto Marburg' and 'The Warburg Institute Iconographic Database' yields, respectively, 5406 and 2699 images, all attributions confounded.
[17] This would especially be useful for samples 6, 7 and 8. The author is presently revising the Topical Catalogue 'The Italian Venus' (reference 9a), leading to a much larger sample which then can be used for a meta-analysis.
[18] John E. Hunter and Frank L. Schmidt, Methods of Meta-Analysis – Correcting Error and Bias in Research Findings. (Thousand Oaks: SAGE Publications, 2004).
[19] General collections like the ones quoted above do not easily allow to quantification because thematic search terms in the database do not necessarily retrieve mutually exclusive artworks, i.e. the same artwork can be retrieved more than once. The same problem occurs in standard thematic reference works like Pigler's 'Barock-Themen' or the Oxford Guide to 'Classical Mythology in the Arts, 1300-1990s', both unfortunately not yet digitized.
[20] See author's post of March 27, 2014 'The Venus of the Eastern, Southern and Northern European Regions' http://kbender.blogspot.be/2014/03/the-venus-of-eastern-southern-and.html
[21] Leo Egghe, Power laws in the information production process: Lotkaian informetrics. (Oxford: Elsevier, 2005) 45. http://yunus.hacettepe.edu.tr/~tonta/courses/spring2011/bby704/power%20laws%20in%20information%20production%20processes-e-book-qvt7lUnRLk.pdf
[22] Pierre-Michel Menger, The Economics of Creativity – Art and Achievement under Uncertainty. (Cambridge, Mass: Harvard University Press, 2014), 179f, 274 inter alia.
[23] Derek J. De Solla Price, Little Science Big Science. (New York: Columbia University Press, 1963), 69.

Bibliography

1. Tournier, Michel. *Le miroir des idées – Traité.* Paris: Mercure de France, 1995. 269 pp.
2. Moretti, Franco. *Graphs, Maps, Trees – Abstract models for literary history.* London, New York: Verso, 2007. 119 pp.
3. Joyeux-Prunel, Béatrice, editeur. *L' Art et la Mesure - Histoire de l' art et méthodes quantitatives : sources, outils, bonnes pratiques.* Paris: Editions Rue d'Ulm, 2010. 600 pp. https://ens.academia.edu/B%C3%A9atriceJoyeuxPrunel/Books
4. Kohle, Hubertus. *Digitale Bildwissenschaft.* Glückstadt, Verlag Werner Hülsbusch, 2013. 189 pp. http://archiv.ub.uni-heidelberg.de/artdok/2185/1/Kohle_Digitale_Bildwissenschaften_2013.pdf
5. Long, Matthew and Roger C. Schonfeld. *Supporting the Changing Research Practices of Art Historians.* USA, Ithaka S+R, 2014. 54 pp. http://www.sr.ithaka.org/sites/default/files/reports/SR_Support-Changing-

Research-ArtHist_20140429.pdf
6. Schich, Maximilian and Sybille Ebert-Schifferer. *Bildkonstruktionen bei Annibale Carracci and Caravaggio: Analyse von kunstwissenschaftlichen. Databanken mit Hilfe skalierbarer Bildmatrizen.* Digital Repository Art History: ART-Dok report. http://archiv.ub.uni-heidelberg.de/artdok/volltexte/2009/712
7. Arscott, Caroline and Katia Scott, editors. *Manifestations of Venus – Art and sexuality.* Manchester and New York: Manchester University Press, 2000. 233 pp.
8. Potter, William Gray. “Lotka's Law Revisited”. *Library Trends* 31,2 (1981): 21-39. https://www.ideals.illinois.edu/bitstream/handle/2142/7191/librarytrendsv30i1e_opt.pdf?sequence=1#page=1&zoom=auto,-87,590
9. Bender, K. *The Iconography of Venus from the Middle Ages to Modern Times.*
   9a.Vol. 1.1 ‘The Italian Venus’ (2007) 156 pp.
   9b.Vol. 2.1 ‘The French Venus’ (2009) 194 pp.
   9c.Vol. 3.1 ‘The Venus of the Low Countries’ (2010) 202 pp.
   9d.Vol. 4.1 'The German, Swiss and Central-European Venus' (2012) 302 pp.
   9e.Vol. 5.1 'The British and Irish Venus' (2013) 197 pp.
   9f.Vol. 6.1 'The Venus of the Eastern, Southern and Northern European Regions' (2014) 154 pp.

Physical books published by www.lulu.com and www.shopmybook.com/en/ https://archive.org/search.php?query=K.%20bender%20Venus%20AND%20mediatype%3Atexts
10. Bender, K. Time Distribution, Popularity, Diversity and Productivity of the Iconography of Venus in the Low Countries, France and Italy. *Research Paper 5 in the Series ’Quantitative Iconography of Venus’.* 26 pp. https://independent.academia.edu/KBender/Papers
11. Hunter, John E. and Frank L. Schmidt. *Methods of Meta-Analysis – Correcting Error and Bias in Research Findings.* Thousand Oaks: SAGE Publications, 2004. 582 pp.
12. Egghe, Leo. *Power laws in the information production process: Lotkaian informetrics.* Oxford: Elsevier, 2005. 447 pp. http://yunus.hacettepe.edu.tr/~tonta/courses/spring2011/bby704/power%20laws%20in%20information%20 production%20processes-e-book-qvt7lUnRLk.pdf
13. Menger, Pierre-Michel. *The Economics of Creativity – Art and Achievement under Uncertainty.* Cambridge, Mass: Harvard University Press, 2014. 405 pp.
14. Price, Derek J. De Solla. *Little Science, Big Science.* New York: Columbia University Press, 1963. 119 pp.

K. Bender is an independent researcher in Belgium with an academic background in the 'hard' sciences and developing his digital thematic research collection since 2004.

Correspondence e-mail: bender@telenet.be
https://twitter.com/bender_k