Phenomenon of "garbage" journals as a direct effect of science commercialization

Oleg V. MIkhaIlov

D. Sc. (Chemistry), Professor,
Department of Analytical Chemistry, Certification and Quality Management,
Kazan National Research Technological University,
Kazan, Russia;
e-mail: olegmkhlv@gmail.com

Abstract: The extremely serious problem associated with the emergence in recent years of a specific phenomenon in scientific journalism, namely with the so-called "garbage" journals, has been considered. The opinion that the first reason for this, on the one hand, is the commercialization of official science, has been expressed. On the other hand, the fact that there are many people in the modern scientific community whose low level of scientific research does not allow them to hope for the publication of their articles in authoritative scientific journals, and in particular those that are indexed in the *Web of Science* and *Scopus*. A number of recommendations concerning of this how to avoid undesirable publication of articles in such journals, was presented.

Keywords: «garbage» journals, *Open Access*, publication, impact-factor, *Web of Science, Scopus*.

References

Akoev M. A., Markusova V. A., Moskalyova O. V., Pislyakov V. V. (2014) Rukovodstvo po naukometrii. Ekaterinburg, Izd. Ural'skogo universiteta, 250 p. (in Russian)

Minobr ne dast deneg vuzam 5–100, ulichennym v samotsitirovanii [The Ministry of Education will not give money to higher education institutions 5–100, caught in self-citation] // Indikator. 17 maya 2017. URL: https://indicator.ru/news/2017/05/17/vuzy-5–100-samocitirovanie. (in Russian)

Mikhailov O. V. (2017) *Tsitirovanie i tsitiruemost' v nauke*. Moscow: URSS-LENAND. 208 p. (in Russian)

Kholodov A. S. (2015) Ob indexakh tsitirovaniya nauchnykh rabot // Vestnik Rossiyskoy akademii nauk. Vol. 85. No. 4. P. 310–320. (in Russian)

Nauchnaya elektronnaya biblioteka [Scientific electronic library] — eLIBRARY.RU. URL: https://elibrary.ru/ titles.asp

Upravlenie bol'shimi sistemami. Sbornik trudov. Spetsial'nyi vypusk No. 44 Naukometriya i expertiza v upravlenii naukoi (2013) / red. D. A. Novikov, A. I. Orlov, P. Yu. Chebotaryov. Moscow: IPU RAN. 568 p. (in Russian)

Yurevich A. V. (2011) K probleme otsenki vklada rossiyskoy sotsiogumanitarnoy nauki v mirovuyu // Vestnik Rossiyskoy akademii nauk. Vol. 81. No. 7. P. 617–620 (in Russian).

Alexander V. Shkurko

independent researcher
e-mail: khanovey@rambler.ru

УДК 001.3:05
DOI: 10.24411/2079–0910–2018–00005

How many knowledge claims are there in a scientific text? A study of three neuroscientific articles' content as reflected in the citing publications

Abstract: Citations are treated as a mark of scientific contribution indicating its value for the scientific community. The in-depth studies of citing behavior show different motivation to cite and various types of citations challenging the view that all citations should be treated as equal. One aspect of citations, however, remains underestimated. Not only researchers can interpret the cited article in different ways and attribute different significance to it, but they can focus on different pieces of knowledge embedded in the text. In this study, citations are used to identify how many different knowledge objects can be extracted from one scientific publication by the scientific community. Three academic articles in neuroscience are chosen as target articles. Samples of the publications citing target articles are analyzed to reveal the cognitive content associated with them by other researchers. The study showed that the diversity of knowledge associated with one scientific text is high: a number of different knowledge objects identified by other researchers varies from 12 to 56. Knowledge objects are distributed unevenly in the sample. A small number of knowledge objects are most popular whereas others are appearing seldom. However, the diversity of citations' content shows that a scientific publication can be valued for different reasons and there are numerous ways how it can contribute to scientific growth.

***Keywords*:** scientometrics, citation analysis; publication; citation context, citation content, knowledge claim.

Introduction

In science, citations, i.e. reference to previous works, is associated with several functions: information search and navigation, establishment and maintenance of social networks, self-presentation and promotion. Yet one of the most widely recognized function of citations is measuring one's scientific contribution (Merton, 1957). Measures based on the number of citations are widely used to evaluate publications, researchers, institutions, or countries. It is also recognized that simply counting a number of citations is a crude measure. Different fields of research vary in the size of their scientific community, rate of advance, or the culture of citing; citation statistics can be artificially manipulated, existing databases can only partially cover the relevant publications and correctly identify references, and so on. Studies show that to better measure the scientific significance of a contribution, one has to take into account in which part of the citing texts it is mentioned, how frequently it is mentioned in one publication, or how detailed is its description [*Voos and Dagaev*, 1976; Hu, Chen and Liu, 2013; *Ding et al.,* 2014].

Limitations of citation analysis based on citations frequencies, led to emergence of a more in-depth investigation of citations in terms of their types, context, and scientists' motivations [*Moravcsik and Murugesan,* 1975; *Erikson and Erlandson,* 2014; *Ding et al.,* 2014]. These studies show the actual diversity of citations, diversity which problematizes implicit assumption of all citations as equal to each other. Within a new approach named citation content analysis (CCA), G. Zhang, Y. Ding and S. Milojević (2013) proposed a classification scheme combining syntactic and semantic aspects of both cited and citing texts, as a basis for in-depth analysis of citing behavior measuring both qualitative and quantitative aspects of large sets of publications.

Citation content analysis, a combination of quantitative citation analysis, and qualitative content analysis, is an important step toward a more adequate understanding of citing behavior. However, one aspect of citations remains to be underestimated, namely, their actual cognitive content. Ideally, citations refer to cognitive content perceived as valuable contribution by the citing author. The question is how many different contributions, i.e. distinguishable and valuable units of content can be extracted from one and the same publication by the scientific community. By default, any particular publication is considered as containing one specific *knowledge claim*. Sociology of scientific knowledge points that the actual content of these knowledge claims can be — and typically are — negotiated and socially constructed in scientific community [*Gilbert*, 1976; *Myers*, 1985]. This means that scientists (including editors and reviewers) can have different views on what exactly has been claimed, and how important is this contribution. Thus, citations can be treated as interpretations of the cited work [*Small*, 1982; *Cozzens*, 1985]. The question is how many *different* knowledge claims can be associated with one and the same publication.

The idea of knowledge diversity embedded in one texts is intriguing. S. Cozzens used citation context analysis to compare sciences with different patterns of citing behavior. She stressed that, basing on other scientists' citations, a number of different knowledge claims can be associated with one and the same text [*Cozzens*, 1985]. Among them, however, there is one 'main knowledge claim' which is the focus of attention by both authors of the original research and other scientists. In an experimental field such as neuropharmacology, the way how the research is cited changes: a more detailed view of the claim turns to a more general type of citing thus signaling that the claim is accepted by the community. This pattern is not observed in another field, sociology of science. Yet in this approach any citation context can be treated as one of four types: the 'main' knowledge claim, peripheral knowledge claim, 'interpretative' citation, or a 'method' citation. Interestingly, the fact that 43 % and 48 % of all the citations of the two articles analyzed by Cozzens, were treated as 'secondary' (peripheral), received almost no attention. In fact, they were considered as noise primary generated by outsiders. This diversity was considered as a background of scientific growth process resulting in establishing the eventual value of a contribution over time.

Similarly, comparative analysis of publications citing three management articles showed a number of different ways how the previous work is treated by scientists and how the initial knowledge claim is reformulated and changed through time in the citing texts [*Golden-Biddle, Locke and Reay*, 2006]. Author of this study also treat a publication as containing specific knowledge claim which can be separated from other knowledge content which is used by the citing authors and which *is not* related to the asserted knowledge claim. That is to say, some authors use and value the cited texts for content which is not considered as the main idea of the focal article. Although K. Golden-Biddle and colleagues also see them as 'peripheral' — as compared to other types of citations which somehow or other refer to the

focal article's explicit knowledge claims, — this citations heterogeneity supports the view that one and the same publication may contain different knowledge objects which are used by scientists situated in different contexts and driven by different cognitive interests.

The diversity of knowledge claims is not surprising for large publications such as textbooks and monographs. A study by M. H. Anderson (2006) is a good example of such diversity embedded in a book. Analyzing citation contexts of publications citing an influential book on social psychology of organizing, he identified 101 distinct concept or ideas embedded in the book (according to readers' interpretations) with 12 most popular concepts (knowledge claims) accounting for 67,6 % of all citations. At the same time, this study identified broad concepts rather than exact knowledge content.

The present study advances this line of investigation and attempts to answer the question directly addressing the issue of diversity embedded in a scientific text: how many *different* pieces of cognitive content can be identified in one and the same scientific article via analysis of the citing articles, and how these different knowledge objects (which should not probable be considered as true knowledge claims) are distributed across the whole set of the citing articles.

Method

The method used here to study the intellectual diversity of scientific publications is straightforward and based on a case study approach. Three arbitrary chosen articles published in academic journals have been taken as targets. All articles are in a broadly defined field of neuroscience. The articles were chosen arbitrary, on the basis of three main criteria: the type of the article is research (not review), the article is well-cited and has at least 200 citations in Google Scholar, and all the articles are from the more or less familiar (for author of the present study) field. The focus on highly cited articles is explained by the choice of method. The goal of the study (investigation of the plurality and diversity of knowledge objects embedded in scientific texts) can be achieved using other methods (e.g. multiple reviewing or even experiment) to become applicable to low-cited articles.

Then, a sample of the citing texts was constructed using Google Scholar. From the pool of the citing texts a number of articles were chosen using systematic sampling and their full texts were downloaded. When the selected text was unavailable for downloading, the next (closest) one has been downloaded instead. Although systematic sampling was used, the study does not pretend to be statistically rigorous in terms of sample representativeness due to its small size.

Identification and interpretation of the citation content was performed using manual analysis, taking into account that the study is about actual cognitive content and not about its linguistic form or syntax. Every citing text has been screened to identify every citation of the target article. According to the context and the content of a citation, it has been counted as a particular knowledge object — a cognitive object which is extracted from the cited text by authors of the citing text as a point of interest. I use the term 'knowledge object' instead of 'knowledge claim' as it refers to any possible way of the text usage not necessarily related to ideas explicitly stated as main contribution. In most cases, identification of a specific knowledge object was unproblematic. All such knowledge objects were consequently listed according to the order of their identification. Sometimes, one citation contained more than

one different knowledge objects. For this reason, a total number of knowledge objects is not equal to the number of citations embedded in the text. Some knowledge objects are appearing several times in one text. They are treated as 'entries'.

As a result of the textual analysis, the following information is gathered for every target article and the sample of the citing texts:

— total number of distinguishable knowledge objects, i.e. ideas and claims which authors of the citing texts extract from the target article;
— frequency of these knowledge objects, i.e. the number of citing texts containing particular knowledge object (mentioned at least one time);
— total number of entries across the whole sample, i.e. every knowledge object reappearing in a given citing texts was counted as a separate unit.

Initially, only one article has been analyzed. Then, to confirm the main finding, two other articles were found and analyzed using the same method but with smaller samples.

Results

Target article 1.

The first target article is 'Electrophysiological Studies of Face Perception in Humans', published in 1996 in Journal of Cognitive Neuroscience by Shlomo Bentin and four other researchers. The article presents empirical results of six neurophysiological experiments of face perception and has been cited for more than 2.4 thousand times since the year of publication.

A sample consisting of citing texts has been selected and downloaded via Google Scholar. Two texts have been excluded from further analysis, because I was unable to identify the knowledge object from the citation context. Thus, a total number of 103 texts have been analyzed.

The average number of entries (roughly similar to the number of citations — see explanation above) per article is 3.2. A total number of different knowledge objects identified in the sample is 27. Table 1 provides a list of all the knowledge objects identified across 103 citing texts, as well as basic descriptive statistics.

Table 1. Knowledge objects embedded in target article 1

ID	Knowledge object	No of texts containing the object	Total number of entries
1	General idea of N 170 potential in occipitotemporal regions as specific for face processing (Reaction to faces as compared to other objects including animals, hands etc.)	68	104
2	Role of the right hemisphere in face processing (N 170 is larger in right hemisphere)	20	23
3	N 170 is evoked by schematic line drawings of faces	2	2
4	Brain areas related to face processing (occipito-temporal regions, esp. Fusiform Face area)	27	32
5	Connection of N 170 and VPP (vertex positive potential)	4	7

Table 1 continuation

ID	Knowledge object	No of texts containing the object	Total number of entries
6	Effects of face inversion on N 170 (both on amplitude and delay). Inverted faces invoke delay in N 170. N 170 is larger in response to inverted than to upright faces.	27	39
7	The study is mentioned as a basis for the positioning of electrodes	3	3
8	The study is cited as an example of research dedicated to particular time interval	1	1
9	Activation of an internal eye-detector as a function of N 170	4	4
10	The 'structural encoding' of face (is independent of recognition of personal identity or emotional display or facial expression)	20	27
11	Temporal features of N 170	2	3
12	Time scaling of face processing	4	4
13	Scalp evoked potentials in eyes processing are not necessarily linked to a given eye direction	2	2
14	ERPs recorded over the lateral temporal scalp are larger to isolated eyes (or eyes with other inner components without face contours) than to full faces. N 170 is sensitive to face parts as well as faces.	20	30
15	Hypothesis on spatial differentiation of cell types existing in the human temporal lobe (two systems in face processing, holistic and piecewise)	3	4
16	General link, referring to a broad field of research, without specifying the cognitive content	8	9
17	Faces evoke a positivity (VPP) that is largest at the vertex and has a latency of 150–200 ms.	6	7
18	Negative ERPs about 50 msec later than N 170 elicited by noses	1	1
19	The peak latency of the N 170 face-sensitive potential is delayed for isolated face features, compared with the response to whole face stimuli	6	7
20	Inverted objects does not produce the same effect (N 170) as inverted faces	1	1
21	Small number of object categories used in the experiment (focus on experimental procedure)	1	1
22	Effects of face distortions on N 170	8	9
23	Topographic features of N 170 (when stressed)	6	6
24	The positive peak recorded from the lateral posterior scalp preceded N 170	1	1
25	N 170 (EEG potentials of intermediate latencies) is associated with explicit categorization of faces and facial expression	1	1
26	The study is treated as a source of normative values in neurophisiological tests of working memory and incidental recall	1	1
27	N 170 is elicited by visual stimuli (probably, incorrect generalization found in one text and apparently resulting from errors in citation).	1	1

Although one can discuss whether some of the identified objects should be treated as separate knowledge objects, and what is the actual cognitive distance between them, the multiplicity of the possible foci of attention for various readers still remains. The list includes objects varying in their generality, epistemological status (i. e. facts, interpretations, explanations, methodology and procedure, theme), and possibility to make connections to various fields of research. In some cases, I was even unsure whether a knowledge claim could be extracted from the target article in the form presented by an author of the citing article (as in case with object 27).

Different knowledge claims and other objects are appearing in the sample with varying frequencies. There are 6 most frequently knowledge objects. Each of them is cited by 19.4–66.0 % of the citing texts. In fact, 66 % of all the sample refer to the general idea of N 170 potential as specific for neural processing of faces (main knowledge claim, object 1), followed by knowledge objects 4 and 6, each of which is cited by 26.2 % of the articles. Figure 1 shows that the frequencies of the knowledge objects' appearances in the texts are distributed unequally.

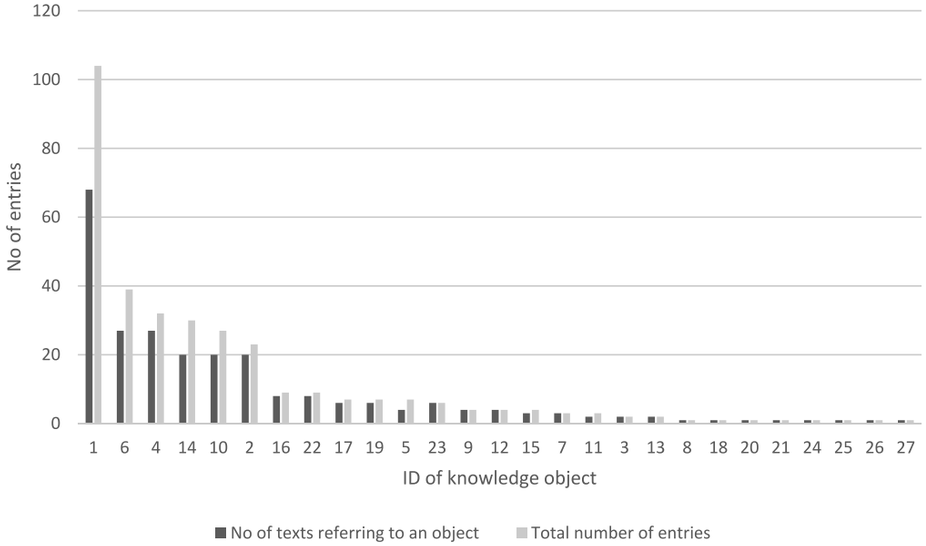

Fig. 1 Distribution of knowledge objects' appearances in the texts citing target article 1

The overall set of knowledge objects can be classified in three main groups according to the frequencies of their appearance. One knowledge claim — a general idea of N 170 as specific for face processing, is absolutely leading. In fact, this knowledge claim can be extracted from the title of the article and does not need the in-depth reading. Knowledge objects 6, 4, 14, 10, 2 are the second frequency group. They are more specific and refer to findings which are identified as important, typically by authors working in the same problem area as the authors of the source texts. The third group consists of significantly less frequently mentioned objects.

Within the sample of the citing texts, 8 knowledge claims are unique, i. e. they appear only once in the sample. Yet they contribute to 29.6 % of the knowledge diversity of the source text — more than the set of the 6 most frequently appearing knowledge objects.

To test whether this pattern of citation heterogeneity is typical, two other scientific texts were analyzed in the same manner.

Target article 2

The second article entitled 'Effects of α-Synuclein Immunization in a Mouse Model of Parkinson's Disease' was published in 'Neuron' in 2005 by E. Masliah and others. The article presents findings from experiments on mice in which the (human) synaptic protein α-synuclein (α-syn) was used as a therapeutic method of immunization in Lewy Body Disease (LBD) which includes the Parkinson Disease (PD).

Using the same procedure as in Study 1, a sample of 40 scientific publications citing the article has been selected using Google Scholar. The smaller size of the sample is partially explained by the smaller number of citations for this article which was slightly more than 400 (i. e., the sample size is about 1 % of the citing sources pool). An average number of citations (entries) across the sample is smaller than in Study 1 and is equal to 1.5. A detailed analysis of the citing texts led to identification of 12 different knowledge objects which attracted researchers' attention. Table 2 lists all the identified knowledge claims (objects) and frequencies of their appearances.

Table 2. Knowledge objects embedded in target article 2

ID	Knowledge object	No of texts containing the object	Total number of entries
1.	Immunization with α-syn as a strategy preventing LBD/PD	10	10
2.	Extracellular antibodies (vaccination) against α-synuclein can reduce accumulation of intracellular (α-syn) aggregates	24	30
3.	Vaccination as a strategy preventing neurodegenerative disease	2	2
4.	Activation of autophagy via immunotherapy might contribute to α-syn clearance in neurons	2	2
5.	Substantiation of the choice of transgenic mice as a model for the study of neurodegenerative diseases	3	3
6.	The role of lysosomal pathways as a possible mechanism of α-syn immunization	4	4
7.	Antibodies may recognize abnormal a-syn accumulating in the neuronal plasma membrane	2	2
8.	The most effective antibodies were against CT-epitopes of a-syn	1	2
9.	The procedure of antibody level determination in the brain and plasma of immunized mice	1	1
10.	The functions of a-syn in modulating presinaptic pool size / synaptic degeneration and motor deficits	2	2
11.	Injection of tagged, purified monoclonal, antihuman cL-synuclein Abs resulted in binding to cell bodies and synapses in cx-synuclein-expressing but not in wild type mice	1	1
12.	Higher molecular weight species for a-synuclein (trimers, tetramers, etc.) could not be observed	1	1

Again, there can be discussions about whether some claims should be differentiated from each other. For example, claims 1 and 3 can be considered as similar although the latter is more general than the latter one and refers to a strategic approach to neurodegenerative disease

prevention not linked with any particular method or any particular disease. On the other hand, knowledge claims such as 'immunization with α-syn prevents LBD' (general) and 'immunization with α-syn prevents LBD in transgenic mice' are treated here as one knowledge claim although the difference in some aspects can be important (when the use of transgenic mice model is explicitly expressed it is treated as a separate object, number 5). Subtle nuances in treating knowledge objects are common and indicate diversity rather than similarity.

Although the target article 2 contains a smaller number of knowledge claims to which the citing authors refer (given that the sample is smaller), and each citing article contains on average a smaller number of entries, the overall pattern of knowledge objects distribution across the sample is very similar to that of target article 1 (fig. 2).

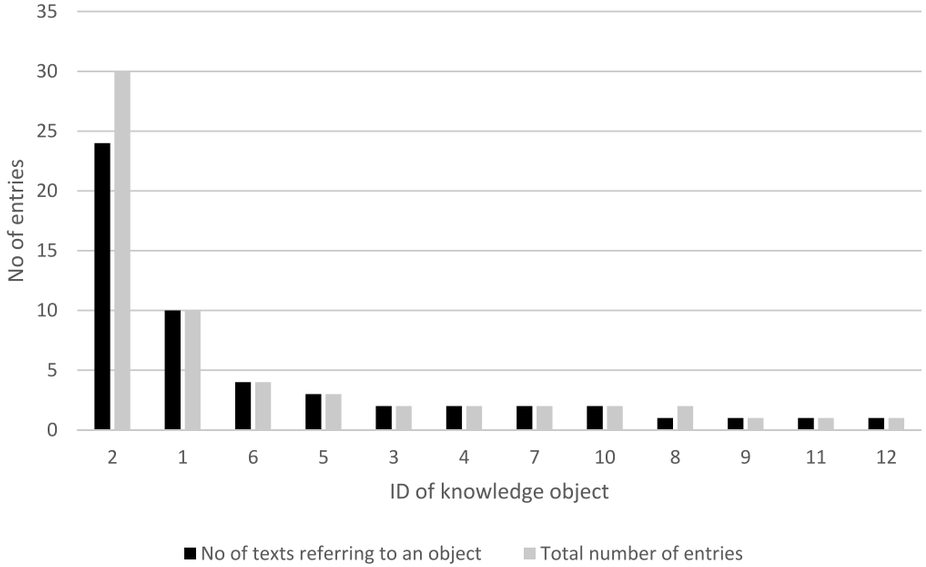

Fig. 2. Distribution of knowledge objects' appearances in the texts citing target article 2

Again, one object is identified by researchers much more frequently than others. Unlike the previous case however, it cannot be extracted from the title and refers to a specific neurochemical effect rather than its potential use in neurodegenerative diseases. Direct reference to α-syn immunization as a way of dealing with LBD/PD is a different knowledge object, second by frequency of use. There are four knowledge objects appearing only once in the sample. Given that the sample size is 2.5 times less than in the first case, one can expect that the actual cognitive diversity of the source text is larger in fact. Yet even the analyzed sample shows that the knowledge objects mentioned by authors diverse in their level of generality and epistemological status: they include identification of facts, hypothetical mechanisms, procedures, or even prospect of a more general medical recommendation.

Target article 3

The third target article is 'Molecular, anatomical, and functional organization of the Drosophila olfactory system', published by A. Couto, M. Alenius and B. J. Dickson in 2005. The article presents the results of authors' attempt to construct a map of odorant

receptor (OR) expression and olfactory receptor neurons (ORN) connectivity in the Drosophila olfactory system. This map shows the neural organization of Drosophila olfactory system, i. e. links between particular neurons and receptors.

To identify knowledge objects embedded in the article, 40 citing publications have been selected and analyzed. As well as target article 2, this focal study is considered as pursuing the goal of preliminary testing the outcomes of the first article analysis. Among the three cases, this one exemplifies the highest level of knowledge diversity. This case was probably the most difficult in terms of identifying particular claims (knowledge objects) and separating them from each other. Some citations contained information on different receptors, genes, or glomeruli and were treated as referring to different objects (e. g. ID 32–35). The sample contains 113 total entries (citations) and 56 different knowledge objects. The average number of entries per text is 2.8. The identified knowledge objects are listed in Table 3.

Table 3. Knowledge objects embedded in target article 3

ID	Knowledge object	No of texts containing the object	Total number of entries
1	37 of 62 ORs are expressed in specific subpopulations of antennal OSNs that display characteristic odor response profiles	2	2
2	Mapping (and classification) of the odorant receptor gene (expressed by most Drosophila ORN types) onto spatially stereotyped glomeruli in the antennal lobe (AL)	14	23
3	Within the antennal lobe, the six palp glomeruli are intermingled with the 43 antennal glomeruli	1	1
4	Information on Or43b associated with VM2 glomerulus	1	1
5	ORNs housed in the three major morphological classes of sensilla (basiconic, coeloconic, and trichoid) tend to project to nearby glomeruli and therefore define several (6) zones in the antennal lobe	4	5
6	Only a single type (of 62) of Or is expressed in most ORNs	5	5
7	Topographic arrangement of glomeruli (similar odorants exciting nearby glomeruli) / Glomeruli are highly ordered	3	3
8	Each gene is expressed in a spatially conserved subpopulation of ORNs either in the antenna, maxillary palp, or larval dorsal organ	6	7
9	There is no obvious molecular or evolutionary logic dictating which OR is expressed in a particular olfactory organ, or even a sensillum type	1	1
10	Instances of odorant receptor coexpression in Drosophila	1	1
11	The functional significance of Or83b receptor coexpression is unknown	1	1
12	Genetic tracing of 44 different OR-expressing populations of ORNs yielded a total of 46 different ORs definitively mapped to glomeruli	1	1
13	A possible chemotropic arrangement of glomeruli in the AL (glomeruli activated by similar odors cluster in the AL)	1	1
14	Adult fly expresses at least 47 ORs genes in 1300 ORNs	3	3
15	Information on Or7 (two adjoining Or7-positive ORNs are present within a single capitate-peg sensillum)	1	1

Table 3 continuation

ID	Knowledge object	No of texts containing the object	Total number of entries
16	VC 1, VA4 and VM7 glomeruli have ORN axons from the maxillary palp via the labial nerve	1	1
17	Information on Or67d associated with DA1 glomerulus	4	7
18	Specific K_a/K_s ratio in larval Or genes	1	1
19	There are 45 odorant receptors expressed in adult olfactory neurons	1	1
20	One receptor — one glomerulus principle (glomerular identity) / One neuron — one receptor principle	7	7
21	Antennal lobe glomeruli vary widely in size, and the size of each glomerulus is stereotyped across individuals	1	1
22	The adult D. melanogaster only expresses 43 OR genes	1	1
23	Fluorescence in DM2 glomerulus	1	1
24	Four distinct classes of trichoid sensilla, housing one to three OSNs	2	2
25	Information on Or65a associated with DL3 glomerulus	2	5
26	Ubiquitous receptor dOr83b (co-expressed with other receptors)	1	1
27	Information on Or47b associated with VA1 glomerulus	2	2
28	Or10a is expressed in ORNs that project to glomerulus DL1	1	1
29	There are 50 glomeruli in fruit fly	1	1
30	The number of glomeruli is closely related to the number of expressed olfactory receptors	2	2
31	Or83b is also co-expressed with Or35a in a broadly tuned class of coeloconic OSNs, but the remaining OSNs in coeloconic sensillae, specialized to select volatiles including small amines, have not been reported to express Or83b, Or or Gr genes	1	1
32	Or46a-expressing OSNs are housed in pb2 sensillum and are associated with VA7l glomerulus	1	1
33	Or67a-expressing OSNs are housed in ab10 sensillum and are associated with DM6 glomerulus	2	2
34	Or98a-expressing OSNs are housed in ab7a sensillum and are associated with VM5 glomerulus	1	1
35	Or46a, Or67a or Or98a — expressing OSNs are not co-expressed with other functional ORs	1	1
36	The differential profiles and the expression of different subsets of genes in different organs and life stages would support their functional divergence	1	1
37	Or22a and Or22b genes are expressed in the same neuron (ab3A) of basiconic type 3 sensilla	1	1
38	Information on dOr13a associated with DC 2 glomerulus	1	1
39	dOr13a might be expressed in intermediate sensilla (this claim is rejected by the citing authors)	1	1
40	Reference to Couto et al. (2005) as a source of data on the morphology of glomeruli	1	2
41	ORs in tandem arrays are co-expressed	1	1

Table 3 continuation

ID	Knowledge object	No of texts containing the object	Total number of entries
42	Duplicate and co-expressed D. melanogaster OR 49a and OR 85f genes	1	1
43	Several OR promoter reporters only incompletely recapitulate endogenous OR expression, presumably reflecting a lack of all necessary regulatory elements	1	1
44	OSNs that express recently duplicated receptors and now project to neighboring glomeruli (e. g. Or98a/Or98b) (Explanation of how one glomerulus splits into two)	1	1
45	Information on Or23a innervating at2	1	1
46	Information on Or59b innervating ab2A	1	1
47	Compared to the maxillary palp, the antenna has about 6 times more OSN types (38 vs 6)	1	1
48	ab9 and ab10 sensilla neurons are proposed to exist	1	2
49	ab9 contains one neuron that expresses Or69a and Or69b and a second that expresses Or67b	1	1
50	Target article as a study of Drosophila olfactory system	2	2
51	12 OR genes map to individual ORNs in trichoid sensilla	1	1
52	Information on Or88a corresponding to at4 sensillum	1	2
53	Distribution of at4 sensilla	1	1
54	In each pair (Or85b, Or85d) and (Or59b, Or59c), one gene is expressed in the antenna and one in the maxillary palp	1	2
55	The third segment of the antenna is patterned in zones along the proximal–distal axis	1	1
56	There are at least 48 ORN classes in sensilla of the trichoid and intermediate category	1	1

The target article contains information about different receptors. This explains why there are so many separate knowledge objects and why they are often appearing only once in the sample. Depending on research interests, scientists can be interested in information on one particular receptor or connectivity pattern.

As in previous articles, separation of knowledge objects from each other is sometimes difficult and debatable. For example, claim 22 looks very similar to claim 14, both referring to the number of OR genes expressed in Drosophila. However, they use different estimations and connotations basing on nuances of the target article's content. The article allows to point to different number of receptors, genes, glomeruli and ORN classes thus contributing to the diversity of its use even in regard to seemingly the same information.

The distribution pattern of the third article knowledge objects is similar to the previous cases (see figure 3). Again, one claim is much more frequently identified by authors than others. Claim 2 referring to the general map of OR genes onto glomeruli in Drosophila is appearing in 35 % of all the citing texts followed by another general idea of 'one neuron — one receptor' and 'one glomerulus — one receptor' principles appearing in several forms in 17.5 of the citing texts. 41 knowledge objects appeared only once in the sample although some of them are reappearing in the citing text.

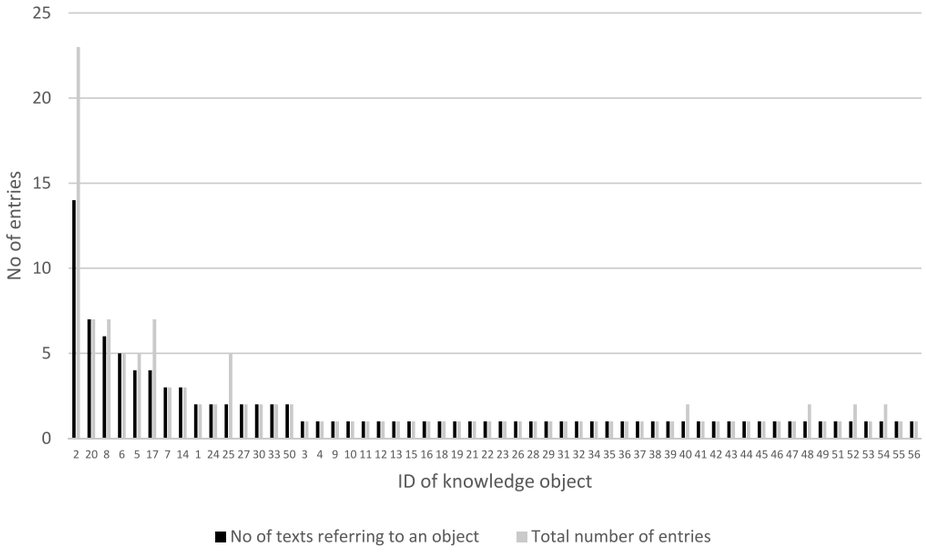

Fig. 3. Distribution of knowledge objects' appearances in the texts citing target article 3

Discussion

Citations embedded in a scientific text can be analyzed for different purposes. One of them is to better understand the essence and value of the cited publication. Researchers present their finding trying to persuade their readers that their knowledge claim is both important and true. However, researchers themselves don't have a monopoly over the cognitive content they produce. Other researchers may treat their findings in a different way and focus on various aspects of their publications. In fact, they even can interpret the knowledge claim in such a way which could be treated as irrelevant and false by authors of the original publication. For scientific growth, all of these interpretations are equally valid and compose the overall body of scientific knowledge.

The so called summarization studies attempt to use citations as a means for producing the shortest description of a 'contribution' embedded in one scientific publication, on the basis of others' interpretations rather than contributor's self-descriptions [*Abu-Jbara*, 2011]. The task of the present study was exactly the opposite — to find out the multitude of different knowledge claims a publication can consist of, the multitude of contributions which can be associated with one and the same publication.

Analysis of cognitive content of sentences containing citations of a target article showed the possibility to use this approach for studying internal knowledge diversity of scientific articles. The current study supports the idea of citations heterogeneity proposed by several authors [*Cozens*, 1985; *Golden-Biddle, Locke and Reay,* 2006; *Anderson*, 2006]. Each of the three target articles analyzed in the current study contains a number of clearly distinguishable knowledge objects, ranging from 12 to 56. This indicated that the richness of cognitive

interests within the scientific community conditions ways how an article is read and interpreted. The diversity can be explained as resulting from relevance detection mechanisms which focus readers' attention on elements of publications which better suit their cognitive and, probably, social interests.

In all three cases, a common pattern of citation distribution can be found, similar to that found by M. Anderson (2006) in regard to the perceived value of an organizational psychology book. A small number of knowledge objects is most frequently associated with a given publication. It is reasonable to suppose that 'summarization' efforts are focused exactly on them. Moreover, the logic of the summarization task claims that the text should be reduced to one particular knowledge claim and consider others as noise. However, the most popular knowledge object covers only a part of readers' interpretation of the target article's contribution. Among the three analyzed texts, the most popular knowledge object accounts for 35–66 % of the relevant sample of the cited texts. This means that at least one third of the citing articles do not mention what can be considered as the 'core idea' or main contribution of the article. Does this mean the interests and interpretation of this part of the scientific community should be ignored? There are a number of reasons why not.

First, the frequency of knowledge objects identified via citations can result from scientific growth in a field making some object less relevant than others. Citation patterns are evolving over time indeed [*Cozzens*, 1985; *Hu, Chen, and Liu*, 2013]. This does not however mean that objects withdrawn from circulation were useless at their stage or would never become important in the future. Although a preliminary study of the distribution of knowledge objects embedded in target article 1, over time (not reported here) showed no any clear time dependencies, this does not mean they could not be found in a more focused study.

Second, the cognitive diversity of scientific publications should take into account the actual diversity of cognitive interests within the scientific community and between different fields of research. A large portion of knowledge objects found in the citing articles consists of objects appearing one or two times in the sample. Are they less important than more popular interpretations of target articles? In terms of overall scientific influence, probably. However, such examples show that these more seldom knowledge claims *can* be extracted from the texts. Researchers with more specific interests, probably working in a different field or elaborating more specific aspects of the problem, found the value in the cited text other than the mainstream interpretation. For them, however, these untypical knowledge claims are important for their own endeavor. Researchers from other fields have less probability to read this article but probably can found in them something which is not interesting or relevant for other readers. The more researchers from other field read the article, the higher frequency of 'less popular' knowledge claims we might expect to find. This hypothesis can be another focus of further research.

Finally, amount of efforts invested in the reading of articles, can lead to identification of more than one knowledge claim. The current study showed that texts containing more than one citation of target article typically include several different knowledge objects. For example, one article citing target article 2 [*Masliah et al.*, 2005] contains 7 citations and 5 different knowledge objects. Another publication citing target article 3 [*Couto et al.*, 2005] contains 18 citations and 11 different knowledge claims. Generally, researchers working on the same problem and composing 1 % of those scientists who is 'going through' in Latour's terms [*Latour*, 1987, p. 60], easily identify and discuss different knowledge claims or knowledge objects embedded in a given text. The actual diversity and, consequently, preserved value of the texts can thus be a function of time and efforts spent by researchers on the text.

In sum, the present study shows that there are many ways how a scientific publication can be used by other researchers. Any publication can be a source of ideas, facts, methods contributing to other scientists' work which constitute the different content of the publications' scientific 'value'.

Conclusion

Three arbitrary chosen scientific articles were analyzed to identify how many different knowledge claims and other objects they can contain. For this purpose, samples of the citing texts have been studied to identify and analyze citation contexts. In all three cases, readers identify numerous ideas, facts and other knowledge objects as valuable contribution to their own research. In all the cases, these knowledge objects are distributed unevenly within the whole set, with a few number of most popular knowledge claims. Whether this pattern of citations distribution is typical is a matter of future research. However, the current study shows that *potentially* any scientific article can be treated in numerous ways and contain many cognitive elements which can be differently valued by various subpopulations of scientists. In practical sense, this may indicate the overall significance, or contribution of scientific texts cannot in principle be judged by one or two experts (e.g. reviewers, editors, advisors, etc.). It seems reasonable to suppose that the more efforts the scientific community invest in reading an article, and the more diverse the community of readers is, the higher will be cognitive diversity of a text, and, consequently, the more value it will generate for science as a social endeavor. Although the scientists' limited resources of time and effort are a natural constraint for increasing investments in reading any given article, the plausible hypothesis of knowledge diversity opens possibilities for a more flexible approach to researchers' publishing strategies. In terms of publications' evaluation, this cognitive diversity can be used as a possible new dimension which might be taken into account when researchers and their publication activity is evaluated with citation-based measures.

References

Abu-Jbara, A. and Radev, D. (2011). "Coherent citation-based summarization of scientific papers" in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, USA, June 19–24. Portland, USA: Association for Computational Linguistics, pp. 200–509.

Anderson, M.H. (2006). "How can we know what we think until we see what we said?: A citation context analysis of Karl Weick's The Social Psychology of Organizing" in: *Organization Studies*, 27, pp. 1675–1692.

Bentin, S., Allison, T., Puce, A., Perez, E., and McCarty, G. (1996). "Electrophysiological Studies of Face Perception in Humans" in: *Journal of Cognitive Neuroscience,* 8 (6), pp. 551–565.

Couto, A., Alenius, M. and Dickson, B.J. (2005). "Molecular, Anatomical, and Functional Organization of the Drosophila Olfactory System" in: *Current Biology,* 15, pp. 1535–1547.

Cozzens, S.E. (1985). "Comparing the sciences: Citation context analysis of papers from neuropharmacology and the sociology of science" in: *Social Studies of Science,* 15, pp. 127–153.

Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., and Zhai, C. (2014). "Content-based citation analysis: The next generation of citation analysis" in: *Journal of the Association for Information Science and Technology,* 65, pp. 1820–1833.

Erikson, M.G. and Erlandson, P. (2014). "A taxonomy of motives to cite" in: *Social Studies of Science,* 44, pp. 625–637.

Gilbert, G.N. (1976). "Transformation of research findings into scientific knowledge" in: *Social Studies of Science,* 6, pp. 281–306.

Golden-Biddle, K., Locke, K., and Reay, T. (2006). "Using knowledge in management science: An investigation of how we cite prior work" in: *Journal of Management Inquiry,* 15, pp. 237–254.

Hu, Z., Chen, C., and Liu, Z. (2013). "Where are citations located in the body of scientific articles? A study of the distributions of citation locations" in: *Journal of Informetrics,* 7, pp. 887–896.

Latour, B. (1987). *Science in Action. How to follow scientists and engineers through society*. Cambridge, Mass.: Harvard University Press.

Masliah, E., Rockenstein, E., Adame, A., et al. (2005). "Effects of α-Synuclein Immunization in a Mouse Model of Parkinson's Disease" in: *Neuron,* 46, pp. 857–868.

Merton, R.K. (1957). "Priorities in Scientific Discovery: A Chapter in the Sociology of Science'" in: *American Sociological Review,* 22, (6), pp. 635–659.

Moravcsik, M.J. and Murugesan, P. (1975). "Some results on the function and quality of citations" in: *Social Studies of Science,* 5, pp. 86–92.

Myers, G. (1985). "Texts as knowledge claims: The social construction of two biology articles" in: *Social Studies of Science,* 15, pp. 593–630.

Small, H. (1982). "Citation context analysis" in: Dervin, B.J. and Voight, M.J. (eds.) *Progress in Communication Sciences*. Norwood: Ablex, pp. 287–310.

Voos, H. and Dagaev, K.S. (1976). "Are all citations equal? Or, did we op. cit. your idem? " in: *Journal of Academic Librarianship,* 1, pp. 19–21.

Zhang, G., Ding, Y., and Milojević, S. (2013). "Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content" in: *Journal of the American Society for Information Science and Technology,* 64, pp. 1490–1503.