

Howard D. White

Authors as Persons and Authors as Bundles of Words

1 Introduction

In 2002 Blaise Cronin and Debora Shaw published “Identity-creators and image-makers: Using citation analysis and thick description to put authors in their place”. This article was, to my knowledge, the first to adopt the “ego-centered” style of citation analysis I had put forward in White (2000, 2001a, 2001b)—“ego” implying not egotism but an individual whose links to others (“alters”) in a network are the focus of research interest (Wasserman & Faust, 1994). In my work, “ego” is an author’s name used to retrieve terms that co-occur with it in the records of a bibliographic database. It is thus a seed term to which the co-occurring terms (the “alters”) bear a fixed relation (e.g., “co-author-with”). When rank-ordered by their co-occurrence counts with the seed, these other terms, especially the high-ranked ones, characterize the seed economically and aptly. The characterizations are superficial, but as rapidly-formed synopses, they can aid in browsing and searching literatures and in gratifying curiosity about author-centered fields of study.

Of particular interest among co-occurring terms are other authors related to the seed in some way by citation. Cronin and Shaw’s “identity-creators” and “image-makers,” which I will explain, are relations of this sort. However, authors’ names are an unusual unit of analysis, in that the same name can simultaneously mean both a person and a body of writings. This fact bears on some major features of ego-centered bibliometric distributions described in sections below. I will first introduce the framework into which the Cronin and Shaw article fits—a framework for bibliometric distributions in general. That will lead to an explanation of ego-centered ones, which involves the relative ease with which terms in them can be associated.

2 Bibliograms

Authors, editors, and indexers routinely use terms of various kinds to identify and describe the writings that constitute literatures. Over time, such terms appear with differing frequencies in bibliographies, just as they appear with differing fre-

quencies in the full-text writings themselves. When ranked by their frequencies, the terms form bibliometric distributions. In White (2005a, b) I proposed “bibliogram” as a name for these distributions as verbal constructs. Bibliograms consist of (1) a seed term that sets a linguistic context, (2) terms of a fixed kind that co-occur with the seed across a database of bibliographic records, and (3) counts of term co-occurrences with the seed by which the terms can be ordered high to low. With appropriate changes, the same definition can be extended to non-bibliographic records, such as the texts of one or more items of discourse.

In return for merely knowing a seed term, bibliograms provide new information. The top terms in them can be visualized as word clouds, for example. Seeds of various kinds may be used to obtain various kinds of co-occurring terms. A classic Bradford distribution takes a subject term as seed and ranks the journals that co-occur with it by how many articles in the journals are indexed with that subject term. A classic Lotka distribution takes a subject term as seed and ranks the authors that co-occur with it by how many papers they have contributed to the literature it designates. Moving beyond bibliographic sources, a classic Zipf distribution takes an extended text as seed (e.g., Joyce’s *Ulysses*) and ranks the words that co-occur with its title by their frequencies in body text. If the last list is stripped of high-ranking function words (like “the” and “of”), the remaining content words often have thematic or stylistic interest.

For decades, information scientists have labeled such distributions by the mathematical or statistical properties of their counts, calling them skewed, empirical-hyperbolic, scale-free, power-law, size-frequency, reverse-J, core-and-scatter, and so on (see, in this volume, Furner, and Glänzel and Schubert). The point of adding “bibliogram” as a new designator is simply to emphasize semantic and other features of the ranked terms no less than their counts. The counts are content-neutral, but the ranked terms associated with a seed make a verbal object that is content-laden. As numbers people, many bibliometricians focus chastely on the counts, whereas the verbal object invites word people to interpret *why* the terms come together as they do—an instance of Geertz’s (1973) “thick description”. Until White (2005a, b), that object had no name.¹

¹ Bibliograms have analogues in fields allied to bibliometrics, such as webometrics and altmetrics. They also have an interesting parallel in word association lists from psycholinguistics. In White (2005a, b, and 2011) I call the latter “associograms” to bring out the parallel.

3 CAMEOs

The acronym CAMEO stands for Characterizations Automatically Made and Edited Online (White, 2001a). “Online” in White (2001a) refers to the now-defunct Dialog search system, whose Rank command could rapidly create CAMEOs in many databases. However, a similar command has been and could yet be implemented in other systems. The discussion here presumes Dialog’s near-instant response time in retrieving the terms that co-occur with a seed.

A CAMEO is a specific type of bibliogram—one in which an *author’s name* as seed retrieves a rank-ordered list of terms associated with the author’s publications. The terms characterize the author personally. When the list indicates subject matter, a CAMEO is a profile in the librarian’s sense of a customer’s subject interests expressed in indexing terms. (This of course plays on the *Webster’s New Collegiate Dictionary* sense of CAMEO as “a small medallion with a profiled head in relief.”) White (2001a), for instance, used two scientists’ names as seeds to rank the INSPEC and Ei Compendex descriptors applied to their papers. It also used the name of a famous Victorian author and artist, William Morris, as a seed to profile him by means of the Library of Congress subject headings applied to his books. (The idea can be expanded from one author’s books to multiple authors’ books combined, as in Ginda et al., this volume.)

Notably, however, White (2000) and (2001a) show how “ego” authors as seeds can be profiled in terms of their co-occurring “alter” authors in the citation databases of Thomson Reuters. This was done by exploiting four retrieval strategies that Dialog software made possible in these databases. (Dialog had tools not in the Web of Science, Scopus, Google Scholar, or CiteSeer.) Using the tags AU for *author* and CA for *cited author*, four different ego-alter relations could be revealed. Dialog’s Select command formed a set of documents defined by the seed, while its Rank command rank-ordered terms by their frequency of occurrence in the set. The four relations are given in Table 1.

These CAMEOs are drawn from bibliographic records only, because their tagged fields keep the relation between the seed and the co-occurring terms constant. That is, when a seed from one field is used to generate a CAMEO of co-occurring terms from the same or a different field, the tags will yield a single relation, as seen in the AU and CA combinations of Table 1. In associating names, a seed is *active* as a co-author or citer; *passive*, as a citee or co-citee. The individual

Tab. 1: Four types of CAMEOs.

Co-authors	<i>Select AU = seed author</i> forms the set of papers by the seed. Then <i>Rank AU</i> ranks by frequencies the seed plus all other persons <i>with bylines in this set</i> . This is the <i>co-author</i> relation.
Identity	<i>Select AU = seed author</i> forms the set of papers by the seed. Then <i>Rank CA</i> ranks the authors in this set <i>by the frequencies with which the seed has cited them</i> . I call this relation the seed author's <i>citation identity</i> . It includes (removable) self-citations.
Image	<i>Select CA = seed author</i> forms the set of papers in which the seed author is cited. Then <i>Rank CA</i> ranks the authors in this set <i>by the frequencies with which they and the seed have been cited together</i> . This is the co-citation relation, which I call the seed's <i>citation image</i> . (Self-citation is again present, but removable.)
Image-makers	<i>Select CA = seed author</i> forms the set of papers in which the seed author is cited. Then <i>Rank AU</i> ranks the authors in this set by the frequencies of papers in which they cite the seed. I call these authors the seed's <i>image-makers</i> , since they create the seed's image through co-citation. (The seed can be one of them through self-citation.)

citer's identity is the basic citation relation, because it is from citers' identities that the counts for the image-makers and the image are aggregated.²

Co-authorships differ from the other relations, in that the data come solely from bylines rather than reference lists. (Co-authors frequently cite their own past collaborations, as well as each other; they are also frequently co-cited.) As one would expect, co-authorships are strongly influenced by geographic proximity (cf. Katz, 1994), but with contemporary communications technology, they no longer depend on it.

The CAMEOs in Cronin and Shaw (2002) present three seed authors: Cronin himself, Stephen Harter, and Rob Kling, all at the time at Indiana University. The relations displayed and analyzed are, first, their identities (authors they frequently cite) and, second, their image-makers (authors who frequently cite them). In Table 2, I give examples of all four relations, using as seeds Eugene Garfield (from White, 2000) and Belver Griffith (White, 2001a). For brevity, the distributions are severely truncated to show only the top four names from each ranking—that is, the inmost core of core-and-scatter distributions that are hun-

² Retrievals based on *bibliographic coupling* strength—the number of references any two papers share—are possible in Web of Science and Scopus, but were not possible in Dialog. In any case, they are not part of the ego-centered CAMEO framework presented here.

Tab. 2: Top names in four CAMEO types, with Eugene Garfield and Belver Griffith as seeds.

	Co-authors		Identity		Image		Image-makers
1264	<i>Garfield E</i>	815	<i>Garfield E</i>	3630	<i>Garfield E</i>	846	<i>Garfield E</i>
7	Welljams-Dorof A	78	Merton RK	520	Price DJD	52	Vlachy J
6	Small H	76	Price DJD	295	Small H	26	Braun T
3	Cawkell AE	67	Zuckerman H	285	Narin F	22	Small H
	Co-authors		Identity		Image		Image-makers
36	<i>Griffith BC</i>	11	<i>Griffith BC</i>	368	<i>Griffith BC</i>	17	McCain KW
7	Drott MC	9	Price DJD	207	Small H	15	Garfield E
6	White HD	9	Small H	147	Price DJD	15	<i>Griffith BC</i>
2	McCain KW	4	Garvey WD	130	Garfield E	13	Small H

dreds of names long. The data come mainly from Social Scisearch (in its surname-and-initials format). They are not current, merely illustrative.³

What stands out in the discussions of these CAMEOs, whether by Cronin and Shaw or myself, is our almost automatic tendency to treat the authors in them as persons—that is, as agents who write and cite on the basis not only of intellectual ties, but of social ties and geographic proximity (see Ekbia, this volume, for more on the somatics of science). Cronin and Shaw can do “thick description” of their seed authors because they have extensive knowledge of them as persons. In my own studies I used the Institute for Scientific Information’s Garfield and Drexel University’s Griffith for the same reason. For example, because of my non-bibliographic knowledge of the people involved, I can say that the seed authors in Table 2 are or were personally acquainted with every other author listed. Moreover, in its columns I see collegial ties, mentor-student ties, employer-employee ties, co-authorships, trusted assessorships, friendships, rivalries—the dense interpersonal linkages that research specialties typically exhibit. Many of these ties must be read into the data from one’s mental encyclopedia, but some, such as an adviser-author tie, may be formally acknowledged in publications (Cronin, 1995). *If* one has considerable domain knowledge of this sort, interpreting citation CAMEOs is easy, and there is much to say. Without it, there may seem little to write about.

³ I have used Small H (rather than Small HG) for Henry Small and chosen Cawkell AE and McCain KW from among co-authors who had tied counts with the seed. Garfield’s self-citation count of 815 in his identity differs from the 846 in his image-makers column; Griffith correspondingly has 11 as against 15. These discrepancies in the ISI data are at least partially explained in White (2000, 2001a).

Recall, however, that any author's name in a CAMEO can also mean an oeuvre, which connotes subject matter. This opens up wholly different analytic possibilities on two levels. On the primary level, an oeuvre consists of one or more full texts—a bundle of retrievable words and their verbal contexts. On the secondary (or meta-data) level, an oeuvre consists of the bibliographic records representing those works, and they, too, are a bundle of retrievable words. It is these word-bundles that allow us to document *intellectual* ties between authors. Intellectual ties pre-eminently include subjects shared within and across oeuvres. They are frequently signaled by exact or partial term-matches and by other easy-to-see terminological relations (e.g., synonyms, hypernyms, hyponyms) from work to work. Citation indexes contain millions of connections of this sort (see Stock, this volume, for the relationship between informetrics and knowledge organization systems). Aggregated in the Thomson Reuters databases on Dialog, they became a form of collaborative tagging.

4 Intellectual Ties and CAMEO Structure

A shallow but important connection is semantic closeness in title terms. For example, "Ritchie (2009)" is a journal article. Data from its bibliographic record in Social Scisearch include its author, AU = L. David Ritchie, and its title, TI = Relevance and Simulation in Metaphor. Among its 39 cited references (CR =) are a book, a handbook chapter, and a *Mind & Language* article. Abbreviated Thomson Reuters-style, they are:

SPERBER D, 1996, RELEVANCE COMMUNICAT

WILSON D, 2004, HDB PRAGMATICS, P607

WILSON D, 2006, MIND LANG, V21, P404

Table 3 spells out the works that Ritchie is actually citing. They obviously map onto his own title. The 2006 article matches Ritchie's in two key terms. The 1996 book and the 2004 book chapter match his on one, but by moving from the bibliographic level to passages in full text, it will be found that they, too, discuss metaphor. As Ritchie cites, he is actually indexing his own article *topically* with other works.

This indexing process feeds into the CAMEOs seen here. The authors' names in them were extracted from the CR strings designating works, as in the Sperber and Wilson examples above. Technically, an author's name becomes a metonym for the works in his or her oeuvre; that is, it can be substituted for them—or, more

Tab. 3: Three works cited in Ritchie (2009).

Authors	Titles
Dan Sperber and Deirdre Wilson	Relevance: Communication and Cognition
Deirdre Wilson and Dan Sperber	Relevance Theory
Deirdre Wilson and Robyn Carston	Metaphor, Relevance and the ‘Emergent Property’ Issue

hazily, for their subject matter—because of its frequent association with them in bibliographic and full-text sources (Falkum, 2011, ch. 6). Through the metonymic process, for example, the names Dan Sperber, Deirdre Wilson, and Robyn Carston have all come to stand for relevance theory, a major specialty in linguistic pragmatics. (Sperber and Wilson (S&W) created it, and Carston is probably now their foremost exponent.) At the bibliographic level, Ritchie’s own title and the works he cites suggest that his name, too, connotes S&W’s relevance theory to the knowledgeable reader, and the full text of his article confirms it.

Intellectual ties—broadly, shared specializations, shared journals, shared topical relationships—are even more fundamental than social ties in binding learned literatures together (White et al., 2004). As stated earlier, authors often have (or had) social ties with the people, living and dead, they cite. But they also know countless people they do not cite, just as they cite people they do not know. The latter include living persons with whom they have never communicated, but also, in many cases, the dead known only through reading. The tie underlying all these possibilities is an intellectual one—the relevance of one work to another, which frequently implies one or more topics in common.

However—and it is a big however—the relevance of one work to another *varies* in how obvious it is, prior to reading. At one extreme, it may immediately manifest itself on Cronin’s (1994) “opus” level, which is also the bibliographic level of titles, subject indicators, and abstracts. At the other, it must be traced to Cronin’s “quantum” level in body text—the passages actually cited—and it may not be obvious even then. I reproduce Cronin’s full list of levels, with his headings and examples, in Table 4.

Present readers can probably think of many ways in which the authors in Table 2 are connectable through their articles and books: it is no great stretch to get from, e.g., Garfield to Small, or Merton to Price, or Griffith to McCain through topically similar publications, even when the authors’ technical vocabularies differ. But what of a case like White (2000), in which I co-cited two papers by Garfield with poems by Walt Whitman and W. H. Auden? For someone not knowing my context, that is a stretch; the ties between Whitman, Auden, Garfield, and my chapter are neither obvious nor presumably strong.

Tab. 4: Tiered citation typology from Cronin (1994).

Focus	Scope	Level
Oeuvre (works of Freud)	Compound	V
Motif (punishment in Foucault; writers on sex)	Macromolecular	IV
Opus (Belkin's article)	Molecular	III
Chunk (the discussion section; chapter 3; opening paragraph)	Atomic	II
Quantum (formula; phrase; chemical compound; method; result)	Subatomic	I

As it happens, Garfield's image—his CAMEO as a co-cited author—conveys as much. Price, Small, and Narin are his top co-citees. The fact that citers refer to works by him and these authors in many different contexts suggests that their interrelatedness is easy to see. In contrast, authors like Auden and Whitman, whose works have no apparent connection with Garfield's, rank down in the tail of his image's frequency distribution. Moreover, like the vast majority of low-ranked names in citation images, they will *stay* in the tail, because they and Garfield do not jointly have rich implications for citers. There are no burgeoning Auden-Garfield or Whitman-Garfield studies. Put simply, citers see Price, Small, and Narin as vastly more relevant to Garfield than Auden or Whitman.

The same observation holds for citation identities, which are interpreted much like citation images. For example, because I cited Auden and Whitman once each in White (2000), they are in my identity, but they are among the many names in the tail of the distribution. My top-ranked names include several of the authors in Table 2—McCain, Small, Griffith, Garfield—and have for many years. My identity will continue to evolve, but none of the names in it are likely to change much in their rankings—certainly not Auden and Whitman.

The overarching lesson at this point is that the ease of associating the seed's name with other authors' names can be crudely measured by their co-occurrence frequencies over time. Table 5 recasts the CAMEOs of Table 1 in terms of this idea. If ease of association is linked to *relevance*, one must ask: what makes authors' works relatively easy or hard to associate and therefore of varying degrees of relevance to each other? I have already sketched an answer here and there, but it will be laid out more fully later.

Tab. 5: Hypothetical ease of association in four types of CAMEOs.

Seed as co-author	A seed author knows some other potential authors. <i>Co-authors</i> shows the relative ease with which the seed has been able to publish new works with them.
Seed as citer	A seed author knows some written works. <i>Citation identity</i> shows the relative ease with which the seed, in his or her own new publications, has cited their authors.
Seed as citee	Many citers know a seed author’s works. <i>Citation image-makers</i> shows the relative ease with which these citers, in their own new publications, have cited anything by the seed.
Seed as co-citee	Many citers know not only works by a seed author but works by others. <i>Citation image</i> shows the relative ease with which these citers, in their own new publications, have co-cited anything by the seed with anything by others.

5 Inference

The names in Table 2 connote oeuvres linked by subject matter, and the reader may also be able to interpret individual author’s names as implicit subject headings on the basis of their writings (e.g., Derek Price = Citation Networks, William Garvey = Communication in Science, Belver Griffith = Co-citation Mapping). Consider, then, a *global* subject heading for *all* the names in Table 2: what would it be? (Or if multiple global labels are allowed for subsets of names, what would they be?) This, of course, is the problem of labeling clusters of authors in a co-citation map, or factors of authors in a factor analysis. The same problem occurs if one is clustering or factor-analyzing research publications or journals.

I do not propose a label here myself; many readers will be able to supply at least one. Rather, I would point out that deciding on any label involves inference. The reader starts with existing assumptions about information science as a field and then is given as new input the names (i.e., the oeuvres) in Table 2. By combining the new input with the existing assumptions, one or more global labels are inferred as a new conclusion. Once the proposed labels are all in view, further inferences might be drawn as to their quality. “Information Science” and “Communication in Science” are too broad, “Co-citation Analysis” is too narrow, and so on.

The larger point is that *every* act of connecting scholarly and scientific texts involves inferences. As another example, take Table 6. These are the top names in Richard Smiraglia’s citation image—the authors with whom he had been co-cited

11 or more times in Social Scisearch as of 2009.⁴ Given this input, where does he fit in information science? What thematic threads can be inferred from the data? Let me imagine one response that you, the reader, as an information scientist, might make: “I don’t see bibliometrics or information retrieval or even information behavior in these names; in fact, I can’t make much of them at all.” Since Table 6 failed to interact with any of your existing assumptions about information science, it is largely or wholly *irrelevant* to you, like an unknown foreign language.

Tab. 6: Citation image for Richard Smiraglia.

Count	Name
105	<i>Smiraglia RP</i>
23	Svenonius E
21	Wilson P
20	Tillett BB
16	Carlyle A
16	Yee MM
15	Gorman M
15	Hjorland B
14	Leazer GH
14	Lubetzky S
13	Buckland MK
13	Vellucci SL
12	Cutter CA
12	O’Neill ET
12	Smiraglia R
12	Taylor AG
11	Chan LM

A very different response to Table 6 would be: “It obviously represents the kinds of things that people in the International Society on Knowledge Organization write about—bibliographic control, bibliographic organization, works as cataloguing entities, conceptual foundations of cataloguing and indexing ... Smiraglia edits the ISKO journal ... His counts with Pat Wilson and Elaine Svenonius show he’s a theory guy ... In fact, I see theorists from long ago—Charles Ammi Cutter from the 19th century and Seymour Lubetzky—which is a sign of humanities-type re-

⁴ Social Scisearch on Dialog counted the *papers* in which co-citations occurred. At the time, exclusive of co-citation, Smiraglia had been cited with his middle initial in 105 papers and without it in 12. All the other counts reflect co-citation; for example, 23 papers cited at least one of his works with at least one by Elaine Svenonius.

search ... A practical cataloguer like Lois Mai Chan is lower down ... All these people are pretty library-oriented ... They're more qualitative than quantitative in style ... Greg Leazer was Smiraglia's doctoral student ..." A babbling-brook response like this would indicate that Table 6 is at least somewhat relevant to you, in that you can readily interpret it. In other words, as a new input, it interacts with your existing assumptions—items of domain knowledge—to produce many new inferences, which you then communicate. (More discussion on domain knowledge and inference in scientometrics can be found in Hjørland, this volume.)

Your existing assumptions further suggest you are at least somewhat interested in the various topics that the names in Table 6 imply. From what you know of their works, it is possible that you yourself might find contexts for citing them in a work of your own. In that event, knowledge of your own work-in-progress is part of your existing assumptions, and recollection or discovery of a related work is a new input. Combining the two, you infer new conclusions, such as: "Pat Wilson's paper reinforces what I'm saying here. But if I adopt his terminology, I've got to cite him." In so doing, you document the connection between your work and a previous one. Your motives for citing may be various, but one is constant across citers—a wish to observe the norm that use of other works will be documented. By citing, you intend that your readers, too, will experience the "proper documentation" effect, along with others having to do with your substantive claims. You may believe, for instance, that the earlier work provides evidence for your own argument, or negates someone else's, or supports a witty insight; there are numerous possibilities. In any case, you trust that your citations will produce in your audience the cognitive effects you intend. As a general principle, citers who intend readers to draw certain conclusions from their citations must first have reached those conclusions themselves.

Here, I have been imagining inferences by present readers, but surely the seed authors in Table 2 also drew inferences about the relevance of certain works to their own as they created their citation identities. In turn, their image-makers drew inferences about these same authors' works as they created their citation images. From cognitive and communicative processes along these lines, the entire edifice of citation is built.

6 Relevance Theory

Some may recognize that my discussions of "inferences" and "ease of association" have been informed by Sperber and Wilson's (1995) relevance theory, a tendency I now want to make explicit. In the article that introduced relevance theory (RT)

to information science, Harter (1992, p. 614) wrote: “The act of citing is the statement of an historical relevance relationship, captured for all time in a published article.” He was speculating about whether data from bibliometrics and citation analysis could help information scientists understand relevance as a dynamic cognitive process in S&W’s sense, as opposed to how it is commonly understood in retrieval system evaluations (relevance = a query-document match in topic). In the spirit of Harter, I will use RT to explain core and scatter in citation CAMEOs. My argument is developed much more fully in White (2011), and so the treatment of RT here will be brief and highly selective. I will relate it to CAMEOs in later sections.

The First Principle of RT is that “Human cognition tends to be geared to the maximisation of relevance” (Sperber and Wilson, 1995, p. 260). This is a cognitive universal, reached through evolution. Wilson and Sperber (2004, p. 608) ask, when is an input relevant?

Intuitively, an input (a sight, a sound, an utterance, a memory) is relevant to an individual when it connects with background information he has available to yield conclusions that matter to him: say, by answering a question he had in mind, improving his knowledge on a certain topic, settling a doubt, confirming a suspicion, or correcting a mistaken impression. In relevance-theoretic terms, an input is relevant to an individual when its processing in a context of available assumptions yields a positive cognitive effect. A positive cognitive effect is a worthwhile difference to the individual’s representation of the world—a true conclusion, for example. False conclusions are not worth having. They are cognitive effects, but not positive ones.

(Sperber & Wilson, 1995, 3.1–3.2)⁵

Wilson and Sperber (2004) proceed to define *the relevance of an input to an individual*:

- a. Other things being equal, the greater the positive cognitive effects achieved by processing an input, the greater the relevance to the individual at that time.
- b. Other things being equal, the greater the processing effort expended, the lower the relevance of the input to the individual at that time.

(p. 610)

Cognitive effects occur when newly presented information interacts with someone’s context of existing assumptions. “Context” in RT is thus a *psychological* notion, and existing assumptions are the thoughts, beliefs, perceptions, and knowl-

⁵ Wilson and Sperber (2004) introduce RT at chapter length. The summary of S&W’s ideas in Yus (2011) may be especially useful for present readers because it applies RT to Web-based communications, including document retrieval. Clark’s (2013) lucid book on RT is one of the Cambridge Textbooks in Linguistics.

edge (including knowledge of textual contexts—“co-texts”) that an individual can access in making sense of new inputs. Such inputs produce effects by (1) *strengthening* an existing assumption, (2) *contradicting and eliminating* it, or by (3) *combining with it to yield a new conclusion*—one derived from neither the new information alone, nor the context alone, but from both together.

S&W hold that, while human beings constantly receive inputs from many sources, they automatically heed those with the greatest relevance for them at a given time—that is, they heed the inputs that produce the greatest effects for the least effort. This in RT is defined as “maximal relevance.” Since the two factors operate simultaneously, one way of expressing them is as a ratio:

$$\text{Relevance} = \text{cognitive effects} / \text{processing effort}$$

Relevance thus varies directly with effects and inversely with effort (White, 2007, 2010a, 2011). But S&W also deny that effects and effort can be measured on ratio scales; the numerical values are not available to us either through introspection or by instrument. Human beings can only measure relevance *comparatively*—that is, by sensing it as a matter of degree. This dovetails with information scientists’ ubiquitous use of *ordinal*-level scales to measure the relevance of documents in retrieval system evaluations: one document will be more relevant to a query than another, but the difference between them cannot be gauged in exact units. Also, a document could be called “more relevant” than another either because it has greater cognitive effects for the same processing effort or because it has about the same effects as the other but costs less effort to process.⁶

As part of linguistic pragmatics, RT is primarily oriented toward spoken communication—toward explaining how hearers infer speakers’ intended meanings from what they actually say. Hearers automatically expect a speaker addressing them to be relevant, and speakers tend to comply, since they want their utterances to be understood. For hearers of an utterance, writes Clark (2013, p. 68), the relevance-theoretic comprehension procedure is:

- Follow a path of least effort in deriving cognitive effects.
- Consider interpretations in order of accessibility.
- Stop when expectations of relevance are satisfied.

⁶ In RT, effort is mainly discussed as it affects the spontaneous processing of utterances *at sentence level*. Larger units of processing, such as paragraphs or whole texts, and intuitions of future processing costs, as opposed to present ones, have received relatively little attention. A relevance theorist who does discuss the non-spontaneous interpretation of literary texts, as opposed to ordinary talk, is Furlong (1995, 2007).

For example, a wife says “It’s late” to her husband at a party, by which she means—and he understands—“It’s time to leave.” Later, on the train platform, he says to her, “It’s late,” by which he means, and she understands, “The train is past due.” The hearer in each case understands what is meant through inference—that is, a new input, the speaker’s utterance, combines with a context of assumptions readily available to the hearer to produce a new conclusion as a cognitive effect. Out of the various meanings an utterance *could* have, hearers automatically seek the interpretation that, in context, has the greatest cognitive effects for the least effort. Moreover, they stop at the *first* interpretation that meets these criteria as the *only* one, which allows talk to proceed without endless parsing of what was meant. (This eliminates consideration of “It’s late” as possibly meaning, e.g., “It’s hours after sundown” in the above example.) Nevertheless, hearers’ interpretations are merely hypotheses, and they can be wrong. Speakers, too, can be wrong (or deceitful), and so cognitive effects are by no means necessarily “positive.”

RT is not only a theory of interpersonal communication, however; it is also a theory of cognition and can accommodate processes that go on in a single head, such as a citing author’s. New inputs, that is, can come from an author’s perceptions or memories or trains of inference rather than talk. In White (2011), I wrote:

Citations in a sense are responses to stimuli that authors themselves have created. It is as if the speaker-hearer pairs typical of RT are no longer separate individuals ...but are lodged within each author. An internal speaker proposes possible utterances by recording them; an internal hearer acts like an editor or critic who accepts some utterances as maximally relevant while rejecting others. Such deliberations presumably account for the redraftings of documents, including the citations in them that depend on authorial reflection rather than reader input or feedback.

(p. 3348)

As noted, Harter (1992) called each citation “a statement of an historical relevance relationship”—in other words, it explicitly shows that an author judged a particular work to be relevant enough to cite, implying that, for the author, it produced acceptably high effects and cost acceptably low effort in a particular context. Any such judgment involves a bundle of inferences on the citer’s part—inferences in which the citing work sets a context of assumptions and the cited work is a new input.

In information science we sometimes read citers’ minds as to what these inferences may have been (cf. Clark, 2013, p. 349–351). By examining the prose in which citations appear, we infer *why* an author cited at that point (as in studies of citer motivations) or *what* a citation contributes to the argument at that point (as in studies of citation functions). Authors themselves sometimes explain why they cited something. In any case, before readers ever see a citation, it will have

had cognitive effects on the citer. It will also have been easy enough for the citer to process; otherwise, it would not occur. (Citations considered but dropped are thick on the ground.)

Table 7 presents some relevance judgments that are open to (pedestrian) mind-reading. Zhao and Strotmann (2008) cited both Schneider et al. (2007) and Zhao (2006). Zhao cites her own 2006 article not only because of its high cognitive accessibility, but also because its topic bears directly on her new topic and shows her developing a “research stream” within her oeuvre. Moreover, as she states, Schneider et al. (2007) themselves built on—and cited—Zhao (2006), and their title is a topical bridge between that article and Zhao and Strotmann (2008). Admittedly, I have chosen titles whose topical relation to each other is very plain, but then so did she.

Tab. 7: Titles in a citing-cited chain.

Year	Citing authors	Title
2008	Zhao & Strotmann	Comparing all-author and first-author co-citation analyses of information science
Year	Cited Authors	Titles
2007	Schneider, Larsen, & Ingwersen	Comparative study between first and all-author co-citation analysis based on citation indexes generated from XML data
2006	Zhao	Towards all-author co-citation analysis

Zhao is citing the earlier articles on Cronin’s (1994) “opus” level; they are relevant as entire works. When she sets up linkages like this, she implicitly justifies her present research in terms of research already published. She intends this justification to be inferred by readers—it is what RT calls an implicature—but it rests on her own inference that the prior articles strengthen her argument and should be discussed. These citer inferences occur whether the connections between citing and cited works will be obvious to readers (as in Table 7) or initially obscure (as in my linking Garfield to Auden and Whitman).

7 Explaining Core and Scatter

As Harter (1992) remarks, once a citation is made, the relevance judgment it represents is fixed over time. In contrast, citation CAMEOs are based on *aggregates*

of citations that gradually change over time. Furthermore, unlike individual citations, they facilitate comparisons of data. Take the citation identity. As an author's publications accumulate, it will develop the core-and-scatter, power-law shape of a bibliogram. Relatively few authors will be recited frequently, forming the core. Increasingly many authors will be recited less and less frequently, forming the scatter, which ends in a long tail of authors cited only once.

Looking only at the ranked counts, network physicists explain the power-law shape of citation bibliograms probabilistically, under such names as "cumulative advantage" (Price, 1976) or "preferential attachment" (Barabási & Albert, 1999). As glossed by Newman (2005):

... the probability of a paper getting a new citation is proportional to the number it already has. In many cases this seems like a natural process. For example, a paper that already has many citations is more likely to be discovered during a literature search and hence more likely to be cited again.

(p. 341)

But the depersonalized account of the physicists ignores the verbal *content* of the distribution that makes it a bibliogram. A relevance-theoretic explanation of this content at the level of individuals is compatible with the content-neutral probabilistic explanation at the level of citers in general.

By hypothesis, the shape of citations identities is determined by processing effort, in the RT sense. Recall that an identity ranks citees by the number of papers in which a seed author has cited them. The ranks of citees indicate their cognitive accessibility to the seed—that is, the effort it costs to associate his or her works with theirs—over years or decades of authorship. Authors as citers reduce this effort by stopping at certain names that spring to mind in the context of new writings. (Citation counts in identities are counts of these writings after publication.) The more frequently certain names have been cited in the past, the less effort is required to use them again. Names easy to process form the *core* of citation identities. Other, more numerous names cited in the past do *not* spring to mind in the context of new writings. The less frequently they have been cited in the past, the more effort required to process them again. They form the *scatter* in citation identities.

Wilson (2007, n.p.) notes that *frequency of use* is among the psychometric factors that affect an individual's effort in understanding utterances in conversations: "The more often a word, a concept, a sound, a syntactic construction or a contextual assumption is used, the less effort is required to process it." S&W (1995, p. 77) extend this observation to use of items held in long-term memory: "... as a result of some kind of habituation, the more a representation is processed, the more accessible it becomes. Hence, the greater the amount of processing involved in the formation of an assumption, and the more often it is accessed thereafter,

the greater its accessibility.” The effortful habituation that forms certain assumptions I take to include instances of intensive reading, writing, social interaction, and thought. But one devotes such effort to relatively few works, which thereby become more accessible than others.

The verbal content of citation identities indicates *which* names (and associated works) are more accessible in memory. Some key evidence comes from *authors as persons*:

- Disproportionate self-citation;
- Disproportionate citation of selected acquaintances (including co-authors);
- Disproportionate citation of selected orienting figures (non-acquaintances) from the citer’s reading;
- Disproportionate citation of works *not* highly cited by others (*pace* Newman);
- Disproportionate citation of familiar works—personal anthologies—as opposed to those that must be newly read.

Evidence that certain *works* are easier for citers to process comes from *authors as bundles of words*:

- Repeated agreement in vocabulary in the full texts of citing and cited works;
- Repeated instances of matching or semantically close terms in citing-cited titles and abstracts;
- Repeated citation of the same works to symbolize the same concepts;
- Repeated citation of works in certain journals and from certain named specialties and disciplines.

This account grounds core-and-scatter citation distributions in the psychology of individual citers—the “natural process” that Newman and other network physicists explain probabilistically at a higher level of abstraction. White (2011) gives examples of several of the bulleted points above, including evidence from Newman’s own citation record (e.g., his personal anthology). Ideally, the bulleted points should be considered hypotheses that would be tested with massive amounts of data. That is not presently feasible, but readers can test them against what they know of citation practices. Readers can also test such hypotheses against the examples in the next sections.

The notion of “personal anthologies” was introduced in White (2011). Table 8 displays a typical one—the works I myself cited five or more times through 2009. The bulleted list above was based on other data, but mine show the predicted features: frequent recitation of relatively few works, heavy self-citation, recitation of acquaintances (McCain, Small), recitation of orienting non-acquaintances (Baldi, Harter, Schvaneveldt), and re-use of works to symbolize concepts (Schvaneveldt on Pathfinder networks; Small on “concept symbols” as an idea).

Tab. 8: Top works in Howard D. White's personal anthology.

Count	Works
9	WHITE HD, 1989, V24, P119, ANNU REV INFORM SCI
8	WHITE HD, 1981, V32, P163, J AM SOC INFORM SCI
8	WHITE HD, 1997, V32, P99, ANNU REV INFORM SCI
8	WHITE HD, 1998, V49, P327, J AM SOC INFORM SCI
6	BALDI S, 1998, V63, P829, AM SOCIOL REV
6	MCCAIN KW, 1990, V41, P433, J AM SOC INFORM SC
6	WHITE HD, 1990, P84, SCHOLARLY COMMUNICAT
5	HARTER SP 1992, V43, P602, J AM SOC INFORM SC
5	SCHVANEVELDT RW, 1990, PATHFINDER ASS NETWO
5	SMALL HG, 1978, V8, P327, SOC STUD SCI
5	WHITE HD, 1982, V38, P255, J DOC
5	WHITE HD, 1986, V5, P93, INFORM TECHNOL LIBR
5	WHITE HD, 2000, P475, WEB KNOWLEDGE FESTSC

Count	Titles
9	Bibliometrics
8	Author cocitation: A literature measure of intellectual structure
8	Visualization of literatures
8	Visualizing a discipline: An author co-citation analysis of information science, 1972–1995
6	Normative versus social constructivist processes in the allocation of citations: A network-analytic model
6	Mapping authors in intellectual space: A technical overview
6	Author cocitation analysis: Overview and defense
5	Psychological relevance and information science
5	Pathfinder associative networks: Studies in knowledge organization
5	Cited documents as concept symbols
5	Authors as markers of intellectual space: Cocitation in studies of science, technology and society
5	Cocited author retrieval
5	Toward ego-centered citation analysis

Very few authors know they can be profiled in this way. Citation CAMEOs and personal anthologies reveal patterns that citers themselves do not monitor or see reported. As my own anthology grew over time, I was largely unaware of it. Having seen it, I am not surprised; it is a window into a long-term cognitive dynamic I recognize. But I did not intend to create this anthology as such; like every other citer, I was simply behaving. However, I *do* intend this chapter's tables as overt communications, and so, for readers, they fall under the cognitive effects / processing effort formula of RT. They do not convey single, sharp implicatures on the order of "It's late" = "It's time to leave." Rather, they convey an indeterminate number of

what RT calls “weak implicatures”—various meanings they *could* be taken to have (as in the case of Smiraglia’s image above). In this, as relevance theorists might point out, they resemble poems.

The strings identifying works in Table 8 reappear in the lower part as bundles of words—titles again. Certain terms explicitly link works, and there are many implicit linkages. Clearly, global *topical* relevance is a binding force here; Harter (1992) was wrong to dismiss it summarily (cf. Bean & Green, 2001, pp. 120–121). But he was right that there is far more to relevance than inferences about whether topics match. For instance, no one would infer that the topics “Bibliometric Distributions” and “Marketing” are even remotely similar, yet the former is highly relevant to the latter once the case is made, as in Chris Anderson’s (2006) book *The Long Tail*.

8 Identities, Images, and Effort

The examples below focus on cores in citation identities and images, which reflect citers’ past cognitive processes. But CAMEOs are also verbal constructs for present readers to interpret. Readers can judge degrees of accessibility and processing effort in specialties they know, because they have the same cognitive mechanism as citers: Relevance = cognitive effects / processing effort. By hypothesis, readers with domain expertise will find names in core easier to relate to a seed than names in scatter. If core and scatter are divided into zones, the descending zones of scatter will tend to be progressively harder to interpret. In other words, readers will be able to give reasonably quick and accurate accounts of what a set of top-ranked names implies. By contrast, bottom-ranked names, taken together, will mean little or nothing to them (cf. White, 2011). Interpretations of top-ranked names are themselves hypotheses that can be tested for accuracy by examining further bibliographic or full-text data. For example, a domain expert might be able to look at the names of an ego-alter pair in a citation image and correctly predict the exact *works* of theirs that are being co-cited.

Table 9 presents 30 names from the information scientist Peter Ingwersen’s citation identity in 2010. His name is italicized as the seed. At left are authors he has cited in eight or more papers. At right are alphabetically selected authors he has cited in one paper each. I discussed the core of this identity at some length in White (2010b). Here, let me simply recapitulate a few points.

Although I recognize several names in the scatter column, they seem virtually meaningless as a group. Core names, in contrast, are vividly suggestive. Ingwersen’s interests as a citer put him at the center of the last few decades of in-

Tab. 9: Peter Ingwersen's identity: Names from top core and bottom scatter.

Core		Scatter	
<i>Ingwersen P</i>	41	Abraham RH	1
Wormell I	17	Babbie ER	1
Belkin NJ	16	Calza L	1
Garfield E	11	Dahlberg I	1
Saracevic T	11	East H	1
Bates MJ	10	Fagan JL	1
Christensen F	10	Gaillard J	1
Cronin B	10	Hahn U	1
Jones KS	9	Jacob EK	1
Rousseau R	9	Kaplan D	1
Brookes BC	8	Larsen HL	1
Croft WB	8	Machlup F	1
De Mey M	8	Nance RE	1
Ellis D	8	Oddy RN	1
Pejtersen AM	8	Paice CD	1

formation science—the intersection of informetrics (Wormell, Garfield, Cronin, Rousseau,); cognitive information science (Belkin, Saracevic, Brookes, De Mey, Ellis); experimental information retrieval (Croft, Jones); and online interactive behavior (Bates, Christensen, Pejtersen). The Festschrift editors independently drew much the same conclusion about Ingwersen as an integrator of information science fields (Larsen et al., 2010). His identity also bears out claims from the bulleted list above, such as disproportionate self-citation and disproportionate citation of works not highly cited by others. His social ties are evident as well: Wormell (his wife), Christensen and Pejtersen (workplace colleagues), and just about everyone else (friends and acquaintances).

Another kind of test, which appears in White (2010b, 2011), involves comparing a seed author's citation identity with his or her citation image. I have claimed that names in the core of an identity are relatively easy for the seed to cite. It would strengthen this assumption, however, if the seed's image-makers frequently *co-cited* the seed with at least some of the same names. For example, Blaise Cronin might cite certain authors heavily for both intellectual and social reasons. If those authors are also repeatedly *co-cited* with him, it confirms that names he finds easy to relate to his work are similarly easy for his image-makers. But the image-makers need not consider Cronin's social ties, if any, with his co-citees—with authors as persons. They can base their co-citations on intellectual ties between oeuvres—on authors as bundles of words.

Tab. 10: Top cores in identity and image for Blaise Cronin.

Identity		Image	
Cronin B	106	Cronin B	1237
Davenport E	21	Garfield E	308
Garfield E	20	Small H	192
Merton RK	18	Merton RK	132
Small H	18	White HD	132
White HD	16	Rousseau R	128
Chubin DE	15	Price DJD	121
McCain KW	14	Borgman CL	116
Kling R	11	Ingwersen P	112
MacRoberts MH	11	van Raan AFJ	112
Porter ME	11	Oppenheim C	108
Ingwersen P	10	Harter SP	106
Machlup F	10	Egghe L	99
Meadows AJ	10	MacRoberts MH	96
Mullins NC	10	Glanzel W	95
Porat MU	10	Moed HF	94
Swanson DR	10	Leydesdorff L	92
Toffler A	10	McCain KW	90
Edge D	9	Thelwall M	89
Hyland K	9	Bar-Ilan J	88
Latour B	9	Chubin DE	88
Shapin S	9	Lancaster FW	83

Table 10 has the identity-image comparison for Cronin, with counts as of 2009. The identity displays authors he has cited at least eight times, an arbitrary core of 21 names plus his own. Accordingly, his top 21 co-citees are also shown, and identity-image matches are bolded. Unboldfaced names from his identity could be matched in lower ranks of his image; even so, the top names show considerable overlap.⁷

Both matched and unmatched names in Table 10 are open to interpretations that are applicable, I believe, to many other seeds (Cronin, 1981). Particulars first. The bolded names in Cronin's identity imply that he specializes in citation theory and analysis. By co-citing him heavily with these same names, his image-makers reveal that the same associations are obvious to them as well, which supports the

⁷ Cronin's identity in Cronin and Shaw differs somewhat from the one here. If Dialog were available, I could probably reduce the differences. For example, Derek Price should be above the threshold in my data. He is not, because his count is fragmented among his various name-forms in ranks I excluded in my 2009 retrieval.

ease-of-processing hypothesis. Cronin and Shaw (2002, p. 36) call Cronin's specialty "citation analysis," but note that he has other strings to his bow. The first—"social studies of science"—is attested in Table 10 by Meadows, Mullins, Edge, Latour, and Shapin; the second—"business strategy"—by Porter. The remaining unboldfaced names in Cronin's identity suggest his wide-ranging interests in information science as a discipline.

These additional interests, however, are not the most salient ones for Cronin's image-makers. They pigeonhole him among citationists and bibliometricians, including mathematical ones (Egghe, Rousseau, Glänzel) that he, as a non-mathematician, is less inclined to cite. If asked to convert his name to a subject heading, as I did with some authors earlier, his image-makers would probably just say "Citation Analysis." In effect, they tend to *stereotype* him (called *typesetting* in White, 2010b). Identities generally imply the grounds for the stereotype—"the author as subject-heading"—while also picking up a seed author's idiosyncratic choices. They reveal the seed's own reading, as opposed to what is widely read by the image-makers.

Stereotyped associations, like clichés, spring to mind with relatively little effort. They abound in research fields (as elsewhere) because they speed interpretation. Above, I described RT's stopping rule for conversations:

... hearers automatically seek the interpretation that, in context, has the greatest cognitive effects for the least effort. Moreover, they stop at the *first* interpretation that meets these criteria as the *only* one, which allows talk to proceed without endless parsing of what was meant.

For individual citers, the first and only interpretation of an author—the one they stop at—is defined by relatively few works. This allows them to cite without endlessly considering possibilities from an author's oeuvre. Moreover, they tend to converge on the *same* works, suggesting an economy of attention. Cronin as a very big bundle of words is thereby reduced to a much smaller bundle of words, such as *The Citation Process* (1984), which is cited repeatedly and hence co-cited repeatedly. Out of hundreds of works he has written, that book and 10 or so of his other top-ranked items account for about a third of his citations (as of mid-2014) in Google Scholar. The content of his most cited works helps to determine the top co-citees in his citation image. RT, in which maximal relevance goes up as processing effort goes down, allows us to make sense of these phenomena.

9 Conclusion

To summarize, the foregoing remarks illustrate two broad claims, based on my reading of RT. Suitably adapted, they apply to all bibliograms, but I will confine them to citation CAMEOs:

- CAMEOs *store cognitive effects*—the associations that authors have inferred among cited works.
- CAMEOs *reflect processing effort*—the ranked frequencies with which authors have made those associations.

Hypothetically, as authors write, their effort varies with the cognitive accessibility of terms—the ease with which the mind supplies them in particular contexts. Authors as citers reduce processing effort by citing people, works, journals, and vocabulary that are relatively accessible to them over time. In CAMEOs, more accessible terms will appear more frequently and become cores; less accessible terms will appear more rarely and become scatter. Determinants of accessibility are what citers know of their citees as persons and as useful bundles of words.

At the same time, CAMEOs are intentional communications for readers to interpret. Those excerpted above are meant to lead readers to certain inferences I myself have drawn, and perhaps to other inferences I have not foreseen. This is merely to say that they permit my claims from relevance theory to be intuitively tested.

While Dialog-based CAMEOs were possible only from 1992 (when Dialog got the Rank command) to 2013, CAMEOs remain instructive bibliometric displays. Practically speaking, their creation has been limited to a few people who could search Dialog through institutional accounts. The underlying programming, however, has turned up in various document retrieval systems since the days of NASA's RECON system in the 1960s (cf. White, 1990, pp. 453–454; White, 1996, pp. 244–245). The idea of featuring CAMEOs within citation-bearing databases, such as a future Web of Science or Scopus, is therefore not entirely far-fetched. Meanwhile, WoS can rank co-authors or image-makers, but cannot do a seed's identity, image, personal anthology, or co-cited works. It seems unlikely that these constructs, especially their cores, have lost their interest. Rather, they have not yet been made easy enough for enough people to generate.

Finally, I would add that segments of bibliograms, such as those seen here, are examples of the *citation objects* that Wouters, this volume, discusses in the last three sections of his chapter on semiotics. They are constructs that can exist only through the retrieval and reorganization of data from citation indexes, and their whole *raison d'être* is to be interpreted as numerically conjoined names or

words. Although they are closely related to the raw data that underlie the maps of science Wouters mentions, they are themselves interesting objects. If explicated, they can make the abstractions of semiotics more concrete. (One can even make testable predictions about their contents.) I have focused on grounding them in what I take to be citers' psychological processes, but that is not the only line of semiotic analysis they can support. Likening them to poems, above, was mainly tongue-in-cheek. But not altogether.

Cited References

- Anderson, C. (2006). *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion.
- Barabási, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.
- Bean, C. A., & Green, R. (2001). Relevance relationships. In *Relationships in the Organization of Knowledge*. Dordrecht, The Netherlands: Kluwer. 115–132.
- Clark, B. (2013). *Relevance Theory*. Cambridge, UK: Cambridge University Press.
- Cronin, B. (1981). The need for a theory of citing. *Journal of Documentation*, *37*, 16–24.
- Cronin, B. (1984). *The Citation Process: The Role and Significance of Citations in Scientific Communication*. London: Taylor Graham.
- Cronin, B. (1994). Tiered citation and measures of document similarity. *Journal of the American Society for Information Science*, *45*, 537–538.
- Cronin, B. (1995). *The Scholar's Courtesy: The Role of Acknowledgement in the Primary Communication Process*. London: Taylor Graham.
- Cronin, B., & Shaw, D. (2002). Identity-creators and image-makers: Using citation analysis and thick description to put authors in their place. *Scientometrics*, *54*, 31–49.
- Falkum, I. L. (2011). *The Semantics and Pragmatics of Polysemy: A Relevance-Theoretic Account*. PhD dissertation, University College London.
- Furlong, A. (1995). *Relevance Theory and Literary Interpretation*. PhD dissertation, University College London.
- Furlong, A. (2007). A modest proposal: Linguistics and literary studies. *Canadian Journal of Applied Linguistics*, *10*, 323–345.
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In *The Interpretation of Cultures: Selected Essays*. New York: Basic Books.
- Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, *43*, 602–615.
- Katz, J. S. (1994). Geographic proximity and scientific collaboration. *Scientometrics*, *31*, 31–43.
- Larsen, B., Schneider, J. W., & Ångström, F. (2010). Foreword. In *The Janus Faced Scholar: A Festschrift in Honour of Peter Ingwersen. E-Newsletter of the International Society for Scientometrics and Informetrics*, v. 6-S: 7–9.
- Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*, 323–351.

- Price, D. de S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27, 291–306.
- Ritchie, L. D. (2009). Relevance and simulation in metaphor. *Metaphor and Symbol*, 24, 249–262.
- Schneider, J. W., Larsen, B., & Ingwersen, P. (2007). Comparative study between first and all-author co-citation analysis based on citation indexes generated from XML data. *Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics* (pp. 696–707), Madrid, Spain, June 25–27, 2007.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. 2nd ed. Oxford: Blackwell. (1996). *Ibid.* Cambridge, MA: Harvard University Press.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.
- White, H. D. (1990). Profiles of authors and journals in information science: Some trials of ORBIT's Get command. *Proceedings of the Eleventh National Online Meeting*. Medford, NJ: Learned Information. 453–459.
- White, H. D. (1996). Literature retrieval for interdisciplinary syntheses. *Library Trends*, 45, 239–264.
- White, H. D. (2000). Toward ego-centered citation analysis. In *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*, Blaise Cronin and Helen Barsky Atkins, eds. Medford, NJ: Information Today. 475–496.
- White, H. D. (2001a). Author-centered bibliometrics through CAMEOs: Characterizations automatically made and edited online. *Scientometrics*, 51, 607–637.
- White, H. D. (2001b). Authors as citers over time. *Journal of the American Society for Information Science*, 52, 87–108.
- White, H. D. (2005a). On extending informetrics: An opinion paper. *Proceedings of ISSI 2005*, the 10th International Conference of the International Society for Scientometrics and Informetrics. Stockholm, Sweden: Karolinska University Press. Vol. 2: 442–449.
- White, H. D. (2005b). Bibliogram. In *Wikipedia, the Free Encyclopedia*.
- White, H. D. (2007). Combining bibliometrics, information retrieval, and relevance theory: Part 1. First examples of a synthesis. *Journal of the American Society for Information Science and Technology*, 58, 536–559
- White, H. D. (2010a). Some new tests of relevance theory in information science. *Scientometrics*, 83, 653–667.
- White, H. D. (2010b). Ingwersen's image and identity compared. In *The Janus Faced Scholar: A Festschrift in Honour of Peter Ingwersen. E-Newsletter of the International Society for Scientometrics and Informetrics*, v. 6-S: 219–227.
- White, H. D. (2011). Relevance theory and citations. *Journal of Pragmatics*, 43, 3345–3361.
- White, H. D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure? Longitudinal evidence from the 'Globenet' interdisciplinary research group. *Journal of the American Society for Information Science and Technology*, 55, 111–126.
- Wilson, D. (2007). Relevance: The cognitive principle. Lecture 3 in Pragmatic Theory [PLIN 2002], an online course of the Department of Phonetics and Linguistics, University College London. [Unpaginated; no longer available online.]
- Wilson, D., & Carston, R. (2006). Metaphor, relevance and the 'emergent property' issue. *Mind & Language*, 21, 404–433.
- Wilson, D., & Sperber, D. (2004). Relevance theory. In *The Handbook of Pragmatics*. Oxford, UK: Blackwell. 607–632.

- Yus, F. (2011). *Cyberpragmatics: Internet-Mediated Communication in Context*. Amsterdam/Philadelphia: Benjamins.
- Zhao, D. (2006). Towards all-author co-citation analysis. *Information Processing & Management*, 42, 1578–1591.
- Zhao, D., & Strotmann, A. (2008). Comparing all-author and first-author co-citation analyses of information science. *Journal of Informetrics*, 2, 229–239.