Title

Copy theory

Permalink

Journal

ISSN

Authors

Fremery, Wayne
Buckland, Michael K

Publication Date

2022-03-01

DOI

Copy Theory

Wayne de Fremery[1] | Michael K. Buckland[2]
[1]School of Media, Arts, and Science, Sogang University, Seoul, South Korea.
[2]School of Information, University of California, Berkeley, California, USA.
Correspondence: wdefremery@gmail.com

Abstract
In information science, writing, printing, telecommunication, and digital computing have been central concerns because of their ability to distribute information. Overlooked is the obvious fact that these technologies fashion copies, and the theorizing of copies has been neglected. We may think a copy is the same as what it copies, but no two objects can really be the same. "The same" means similar enough as an acceptable substitute for some purpose. The differences between usefully similar things is also often important, in forensic analysis, for example, or inferential processes. Status as a copy is only one form of relationship between objects, but copies are so integral to information science that they demand a theory. Indeed, theorizing copies provides a basis for a more complete and unified view of information science.

1 INTRODUCTION

Cognitive and cultural activities are characterized by various forms of copying: imitation (mimesis), representation, conformity, adaptation, and the emulation of behavior (Boon, 2010). Indeed, many areas of human activity involve a substantial degree of adoption, harmonization, and standardization that serve to fashion things, processes, and even people as usefully similar. Although the word *copy*, especially in legal contexts associated with copyright, is often used to mean "lacking originality" or "creativity," these activities are reproductive and generative and so are creative. Information science has tasked itself with the work of collecting, organizing, storing, retrieving, interpreting, and transmitting the results of these creative processes. Yet, the literature of information science and technology has paid relatively little attention to theorizing copying and copies.

There are large literatures on the nature and impact of writing, printing, telecommunication, and digital computing. In contrast, the attention paid in information science to the history and theory of copying and of copies has been very small. This is ironic because it is their ability to generate copies, either locally or at a distance, that makes writing, printing, telecommunication, and digital computing important. Even when the production of copies is recognized as important, copies and copying have not been considered foundational in information science. To be interested in a copy rather than an original might seem to be preferring a shadow to substance, but most of what we know is second-hand knowledge (Wilson, 1983) and the documents upon which we depend daily are ordinarily copies—the copy of our driver's license we send to an insurance company, the photocopy of useful passages from a book

we cannot take from the library. In brief, copying and copies have received relatively little direct attention in the history and theory of information science despite their centrality in everyday information practices and their enormous importance for administration, commerce, education, and scholarship as a convenience, for preservation, and in enabling further productivity.

After a brief description of some technologies for copying and of the etymology of *copy*, some basic concepts of copy theory will be presented before showing how copying and copies are integral to information science and other selected disciplines.

2 SOME TECHNOLOGIES FOR COPYING

Transcribing texts is presumably as old as writing itself. The earliest extant examples of xylographic (wood block) and movable metal type printing come from eighth- and fourteenth-century Korea, respectively. Tracing paper and tracing parchment were known already in the European middle ages. From the seventeenth century the pantograph (a pen mechanically linked to a pointer or to a writer's pen) was occasionally used to make copies, and, later, pressing thin moist paper against handwriting so that some ink seeps through to create a facsimile copy (letter press copying) was widely adopted in offices.

Copying was transformed once techniques for preserving photographic images were developed in the 1830s. In addition to copies of static images, pictures were arranged serially to produce "moving pictures." The many popular forms of document duplication developed in the late nineteenth and early twentieth centuries (ditto, mimeo, etc.) can be regarded as forms of printing because new masters are created. Although using photography to copy documents has a long history, widespread photocopying began only in the early twentieth century with the commercialization of the photostat process which used a large camera to make images directly onto sensitized paper. Compared with manual or typewritten transcription the overwhelming superiority of photostat technology in speed, simplicity, efficiency, and versatility was obvious, especially for images. In the 1930s, microfilming brought advantages in compactness, transportability, and reproducibility to libraries and many business applications (Buckland, 2021). In the 1960s, electrostatic copying (commonly known as xeroxing) became the technology of choice for copying machinery and for computer printers with 24 billion copies a year in the US by 1979 (Owen, 2004; Schwartz, 1996; Thompson, 2015).

Importantly, more convenient copying technologies inspire a desire to do more copying. In the 2000s, for example, Google worked to make copies of all the 130 million books it estimated had been produced worldwide. Google did not try to copy every copy of every book published. The premise was that a single copy of a book could stand as a useful substitute for all other copies of that book. There have been attempts to estimate the number of different books produced, but we are not aware of any estimates of the total number of copies of all books ever produced. The scale of copying has become very large. Digital technologies facilitate the copying of files on a vast scale. In 2018, YouTube announced that 500 hours of video content were copied to its platform every minute, or roughly 82 years of video a day (Frangoul, 2018).

Although many of these individual technologies and the copies they produce have been the focus of research in information science, the nature of copies has received relatively little attention.

3 ETYMOLOGY

The word *copy*, like *copiousness*, comes from the Latin *copia*, meaning abundance or multitude. The present meaning of the noun *copy* seems to have come through the sense that transcribing a text increases the number available and so affords a more plentiful supply. The verb *to copy* means to transcribe or reproduce. Importantly, the term *copy* is ordinarily used in relation to documentary representations of symbols, signs, texts, images, and other objects. So, one would refer to a *copy* of a document but would use other terms for objects not regarded as documents. Tools or raw materials might be usefully similar to other tools or raw materials, for example, but we are unlikely to ask for a *copy* of a hammer or a nail, which do not ordinarily serve documentary roles. Any exception is likely to arise when the object has a documentary significance of a historical or cultural nature. Accordingly, here we are concerned with copies as documents in a broad sense of that word (Buckland, 1991; 2017; Malkan, 2005).

4 BASIC CONCEPTS

We may think of two copies of a document as being the same. But even when two copies are indistinguishable, they never can be entirely the same simply because they are distinct "particulars," meaning distinct material objects that, however much they may resemble each other, cannot physically occupy the same point in space and time. In practice, what we mean by "the same" is that two different things are usefully similar for some purpose in some situation.

4.1 Objects as Particulars

Every physical thing is a unique *particular* however much it may resemble other objects (Rettler & Bailey, 2017). Physical objects may be very similar and might, for some practical purposes, be substitutable ("fungible"), but they can never be entirely the same. "There are no fungibles in nature," write Harré & Llored (2018, p. 185). Every copy, being a material object, is itself necessarily a particular. If it were not, a copy would be what it copies. Strictly speaking, any assumption that two objects are the same or that they are entirely substitutable is either a fallacy or a convenient approximation that may be acceptable in some context (Hayes, 2011). An assumed equivalence may often be useful for some purpose.

4.2 Objects as examples

As a practical matter, two particulars may be equally useful for some purpose. Either of two hammers may serve to drive in a nail. Either hammer can substitute for the other and to that extent they are fungible. In other words, we can think of a class of hammers within which diverse individual hammers form instances. The key point is that in this situation these particulars (the hammers) are *regarded as* fungible. Each is a unique particular; but each is also, in a more abstract view, an example (instance, member) of a set of hammers.

The term "copy," for our purposes here, retains two different principal meanings. In a narrow, material sense, it denotes the outcome of a process of copying, a derived version, a literal copy, and is also a particular. These processes may often be mechanical but they are also generative because they create new objects. In another more abstract sense, it denotes more generally any member of a class of examples regarded as usefully similar to other objects or

acceptably substitutable (fungible) for some purpose. A lazy school child assigned to read a Shakespeare play will prefer a useful summary to the whole play. But a Shakespearean scholar or a book collector would insist on having the *right* copy for the purpose of making an academic argument or more money at an auction. For their purposes, any other copy is unlikely to be a satisfactory substitute (Tanselle, 1989).

4.3 Type, species, and set

Communities name and classify copies in different ways. *Type* and *token* are widely used terms for distinguishing items that are deemed the same and those that are not. It has been argued that communities, especially those that form academic disciplines, should identify "the types their discipline requires, and the properties such types satisfy" (Wetzel, 2009, p. 150). Types are abstract objects that have particular tokens which bear "a family resemblance" (Wetzel, 2009, p. xi and p. 124). *Class* and *instance* have similar meanings, although are not the same. In biology, the convention is to designate one individual particular, traditionally the first to be formally described, and to use the characteristics of that specimen for defining a species. This defining particular is called a *holotype*. All other particulars judged to be of the same species are considered *specimens*. A *set* and its *members* have a different similar meaning. A set is a quantity rather than a definition.

       The type-token relationship can be in a hierarchical structure: species are instances of a larger category (type) genus, and the species are then tokens of that type. A comparable type-token hierarchical structure can be seen in library collections: multiple copies of a particular edition; multiple editions of the same work; multiple works on the same topic; and so on.

4.4 "Work" and copy

In a library context, the word *work* has had four principal uses. Like the French term, *oeuvre*, it can denote a scholar's entire creative output. It is also widely used to refer to a particular text, regardless of edition, as in "a good work on this topic," or more familiarly to a particular document, as in the phrase "the work in hand." Finally, largely from the influence of Seymour Lubetzky, *work* has acquired a technical meaning within the Functional Requirements for Bibliographic Records (FRBR), a conceptual framework developed by the International Federation of Library Associations and Institutions (IFLA) (Lubetzky and Hayes, 1969; IFLA, 1997/2009; Svenonius, 2000). In FRBR, the products of intellectual or artistic endeavor are theorized in terms of four levels. Intellectual and artistic creations are considered abstractions apart from the ways they are manifest as objects. An abstraction called a *work* is realized through an *expression* in orthographies associated with specific languages, musical notation, or other methods of representation embodied in a *manifestation* (such as a specific printed edition) with *exemplars* called *items*. Items, manifestations, expressions, and works can be understood as a way that librarians have described kinds of similarity for their purposes in order to organize copies. This is to make the obvious point that cataloging and models such as the FRBR concern copies, even if copying and copies themselves are not theorized. The model facilitates a form of information representation, organization, and discoverability that aims to help users of the model identify and obtain copies of conceptualized items. FRBR is a specific method for describing useful similarity in order to "fulfill the objectives of catalogs" (Tillett, 2003, p. 3). After Wetzel

(2009) we can say that FRBR is how information scientists have identified the types and tokens that their discipline requires for the purposes of cataloging. The FRBR does not, however, as has been claimed, use a vocabulary that "free[s] us from the baggage of past terms that were ambiguous" (Tillett, 2003, p. 3). Rather, it functionalizes arbitrary terms to formulate the "universe as we know it today" (Tillett, 2003, p. 3), where "we" suggests information scientists. For new initiates to cataloging and FRBR, FRBR's terms will seem arbitrary in the root sense of the word. Arbitrary connotes judgements made according to opinions and preferences. Understanding that FRBR's terms are, in fact, arbitrary does not diminish their importance. Instead, it clarifies that the widely-used FRBR model helps to coordinate cataloging practices that suggest useful similarities among documents based upon consensus opinions and preferences of information scientists for the purpose of cataloging.

4.5 Copy and original

When a copy Y is made from source document X, the term *original* is commonly used to denote the source document. This usage does not imply that document X is original in any other sense than being the source for document Y.

Any copy, indeed any document, can be expected to be not only new but also to some extent novel. But originality is a matter of degree that, as described below, depends on purpose and context. Just as two documents cannot be identical, a single document cannot be completely original, i.e. a document created *ex nihilo* that has the quality of being independent of and different from anything that has gone before. This is true even if the idea of an original document is often useful. McGann (2014) makes a similar point when he writes: "Like the concept of origin, original documents are fictions we practice in order to manage their losses and our limits" (p. 5). McGann describes documents lost to history, but also what is "lost" when documents are copied, i.e. all that distinguishes a copy from what has been copied and how these distinctions multiply with every act of documentary reproduction. McGann is also describing the ways that documents limit our ability to know because of their elisions, even as they provide evidence of facts. Original documents can be considered useful fictions that facilitate certain ends, such as managing the *copia* of a library's many copies so that each might be useful. "Fictions," what librarians would call abstractions, provide boundaries that help to organize the overwhelming plenty of the bibliographical universe. They establish forms of bibliographical control that facilitate the exploitation of its resources for various ends. Tillett's (2003) description of the FRBR model suggests how original documents are abstractions information scientists practice to facilitate the bibliographical control FRBR can provide. Tillett uses *Gone with the Wind* to illustrate FRBR's power. *Gone with the Wind* "was expressed in many ways including an original text, translations, a critical edition and editions with illustrations, and many more" (Tillett, 2003, p.11). Each manifestation is related to the expression it is "based on" (Tillett 2003, p. 11). But what is the base? What counts as the "original text"? The one that was burned, perhaps, as legend has it, in the boiler room of Margaret Michell's Atlanta apartment the day after she died? The parts of the manuscript she typed on her secondhand 1923 Remington that recently reappeared, having, apparently, survived the flames? The portions of extant manuscript pages that Mitchell redacted? The documents her husband put in a safe deposit box in the Citizens and Southern National Bank on Marietta Street? Tillett does not fret these details and the ways that they challenge the idea of "*an* original text" because in the FRBR model

expressions, like *works*, are abstract entities (Tillett 2003, p. 9) used to delimit the copiousness of all that can count as "an original" copy of *Gone with the Wind*. "Original texts" are abstractions practiced in order to manage their loss and losses, as well as, in the case of the FRBR, to instantiate the limits of catalogs. By theorizing copies, we encounter the diversity of what can count as "original" and, with Yeo (2010), can consider more "nuanced" understandings of originality (p. 106). These more nuanced understandings, can in turn, help to articulate useful similarities among documents, as well as lead to a better understanding of how individuals and communities consider documents usefully similar for different purposes.

4.6 When differences do matter

Although copies may normally be regarded as acceptably substitutes, differences between copies often also matter. Sameness, in practice, is relative to some purpose and purposes may differ. Handwritten documents in Shakespeare's time, for example, used a style of handwriting, "secretary hand," that was very different from modern handwriting and is now difficult for us to read (Secretary hand, 2020). A transcription into a modern orthography would be very useful for many purposes but it would not be suitable for a specialist in the history of handwriting, who, if the sixteenth or seventeenth-century original were not available, would want a copy that looks the same (a facsimile) as Shakespeare's manuscript. The transcript in a modern orthography and the facsimile are both copies but different purposes make each acceptable under different circumstances.

Reproduced documents are considered copies when perceived similarities between two objects are useful for some purpose. But the differences between similar objects can also be useful. A nice example is provided by Lodewyk Bendikson (1932) who presents the case of a printed document in which a censor had obscured (redacted) several lines of text with ink marks. Figure 2 shows two photographic copies of the same page. On the left is the redacted version. On the right, the redacted lines are clearly legible. Both images are photocopies of the same original. The explanation is that the ink used by the censor was chemically different from the ink that had been used by the printer. Crucially the censor's ink absorbs light in the visible spectrum but not infrared light, whereas the printer's ink absorbs both. Wherever light is absorbed appears black and where it is not absorbed is transparent. Consequently, in the left-hand image photographed using light in the visible spectrum the censor's marks appear black and block the underlying text. But in the right-hand image, photographed using infrared light, the light passes through the censor's marks rendering them transparent and is absorbed by the printer's ink making the printed text appear black. The two documents are both copies but the difference between them is of importance.
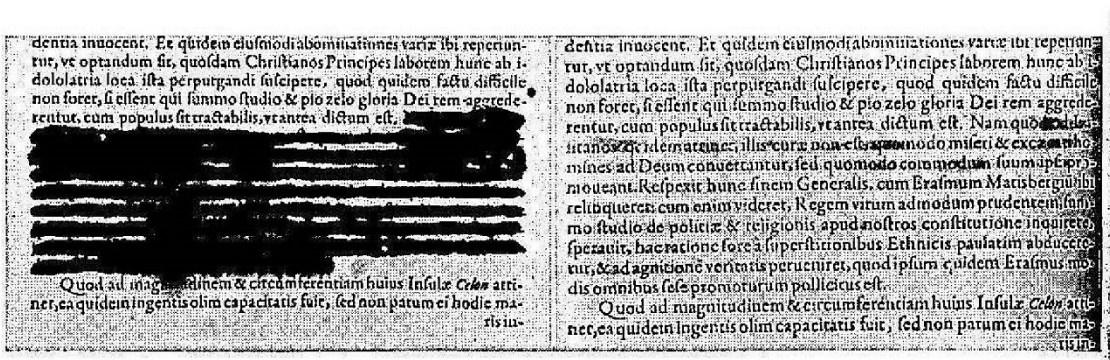
FIGURE 1 Redacted document photographed using visible spectrum (left) and infrared light (right). (Bendikson, 1932, p. 792).

4.7 Inheritance

Another important feature of copies is that when they differ, their differences are ordinarily inherited by subsequent copies. In Figure 1 where dirt blocked light during the photostat copying process small white areas resulted. Subsequent copies of that photostat (e.g. Figure 1 itself) will inherit this difference unless it is corrected. Manual transcribing of any long text results in errors, omissions, additions, and corrections, some deliberate, some unconscious. Whatever the cause, these changes are not random and create patterns that may be discerned by careful scholars investigating the different transcriptions (witnesses) to understand their textual genealogy (stemma).
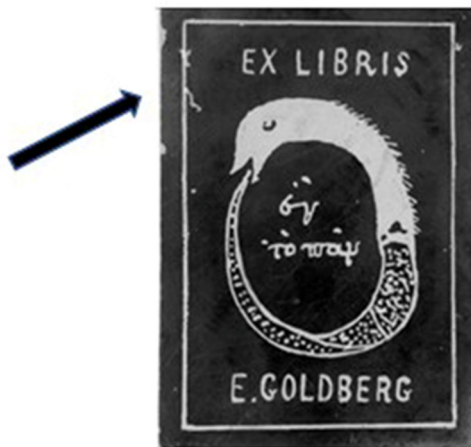

FIGURE 2 Photostat copy of bookplate of search engine pioneer Emanuel Goldberg, c. 1910. The arrow upper left indicates where dirt has caused white areas. (Buckland, 2006, 33–34).

5 COPIES IN DIVERSE COMMUNITIES

Interest in copies arise in many contexts according to local needs. Here we note a few concerns with copies in diverse communities.

5.1 Copies in information science

The website of the Association for Information Science and Technology (2020) offers three canonical explanations of information science. These three explanations are interesting from a copy-theoretic perspective. The first is a widely quoted explanation by Harold Borko:

> Information science is that discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with the body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, and utilization of information. (Borko, 1968)

The second is by Martha E. Williams:

> Information science brings together and uses the theories, principles, techniques and technologies of a variety of disciplines toward the solution of information problems. Among the disciplines brought together in this amalgam called information science are computer sciences, cognitive science, psychology, mathematics, logic, information theory, electronics, communications, linguistics, economics, classification science, systems science, library science and management science. They are brought to bear in solving the problems with information — its generation, organization, representation, processing, distribution, communication and use. (Williams 1987/1988)

The third, by Tefko Saracevic, states:

> Information science is the science and practice dealing with the effective collection, storage, retrieval, and use of information. It is concerned with recordable information and knowledge, and the technologies and related services that facilitate their management and use. More specifically, information science is a field of professional practice and scientific inquiry addressing the effective communication of information and information objects, particularly knowledge records, among humans in the context of social, organizational, and individual need for and use of information. The domain of information science is the transmission of the universe of human knowledge in recorded form, centering on manipulation (representation, organization, and retrieval) of information, rather than knowing information. (Saracevic, 2009)

These three explanations of information science selected and presented by the society are consistent. In each case, the author wants information science to be concerned with an entity (information) that remains undefined, but which, in the end, is operationalized in terms of physical documents, with "information as thing," such as records and comparable documents. In each case the final sentence makes little sense unless understood to mean that information science is based on and concerned with physical copies of documents.

5.2 Shannon's copy theory

Claude Shannon's landmark paper, "A Mathematical Theory of Communication" (Shannon 1948; Shannon and Weaver 1949), is often considered foundational in information science. Shannon, a mathematician and electrical engineer, was concerned with how to efficiently and reliably reproduce a message at some distant destination. He formulated the process in the following terms: A message is encoded on the sender side, sent through a channel, then decoded on the receiver side. However, distortion ("noise") may occur during the process much like transcription errors in manuscript copying. Shannon's concern was strictly with an engineering task, with the characters or other physical signals sent, not with meaning. His concepts and methods had an enormous impact in a range of fields and led to a specialty known as information theory. However, the problem he addressed was, quite simply, how could one make a copy at a distance efficiently and faithfully. So, Shannon's communication theory would more aptly be known as Shannon's copy theory. (For a convenient introduction to Shannon's and comparable theories see Badia (2019)).
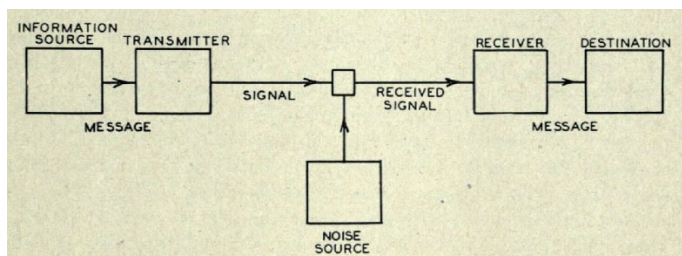

FIGURE 3 Schematic diagram of a general communication system (Shannon, 1948).

5.3 Information Retrieval

Information retrieval in its simplest form involves the retrieval of suitable copies. A collection exists and there is a binary division between the searchable collection and whatever else may exist but is not searched. Then, for any given search, there is the binary division between the retrieved set and the unretrieved set. Refining the retrieval result into more than two categories increases the number of subsets and divides the retrieved items across them. In the extreme case of strict ranking this becomes a ranked list of sets each containing a single member.

The commonest form of searching, submitting search words (more precisely, character strings) to retrieve documents containing those words, can be understood as a search for copies. The principle is that the greater the textual similarity between a query and a retrievable document the more usefully similar it is and the more relevant it is presumed to be.

5.4 Libraries

Libraries are centrally concerned with copies. Consider, for example, the dilemma of a librarian with limited resources addressing the trade-offs between adding additional titles (types), adding extra, duplicative copies of books in high demand (tokens) and other claims on limited resources such as reference service and opening hours (Buckland, 1979).

5.5 Philology

Philology once "reigned as king of the sciences" and can be understood as the source of the modern humanities (Turner, 2014). The roots of information science can also be traced to philology. Compare for example, the definition of philology by the German philologist August Boeckh (1785-1867): "Die Erkenntnis des Erkannten," the knowledge of what is and has been known (cited in McGann, 2014, p. 4), with the definitions of information science provided above. Both concern knowing what is known rather than knowledge of a particular type. "The domain of information science," Saracevic suggests, concerns "the transmission of the universe of human knowledge in recorded form…rather than knowing information" (Saracevic, 2009). Philologists study the documentary record with the assumption that documents carry the evidence of, in the words of the late D. F. McKenzie, "the history of their own making" (McGann, 2014, p. 4) and with the belief that documents can reveal how we know how we know. This is to say the documentary record, which consists of extant copies of documents, has been a shared central concern of both nineteenth-century philologists and twentieth-century information scientists.

As already noted, both intentional changes and unintended errors are created when a copy is made and these changes will ordinarily be inherited when the copy is itself copied. Since transcription changes and errors do not occur entirely at random, one can, with sufficient understanding of transcription behaviors, make inferences from close examination of a copy concerning its source. A famous example of this kind of analysis used to assert beliefs about antecedents of ancient texts is a study by the philologist Karl Lachmann (1793–1851). He inferred in 1850 that the three main surviving manuscripts of the Latin author Lucretius were all derived from a lost earlier manuscript written, he surmised, on 302 pages with 26 lines per page. Also, he concluded that this lost manuscript was in turn copied from an earlier lost manuscript probably written in minuscule handwriting, which in its turn was copied from an even earlier lost manuscript of the fifth century probably written in rustic capitals (Timpanaro, 2005). Nowadays, this kind of analysis would be computer-aided and called artificial intelligence.

5.6 Bibliography

In ancient Greece and in Europe until the seventeenth century, the word bibliography (literally "book writing") referred to the writing out (transcribing) of texts. Then, starting in France, bibliography gradually changed to mean writing *about* texts (Nestler, 2005).

Bibliography, understood as a diverse group of scholarly practices focused on creating and accounting for copies, enables us to trace branches of information science's family tree back into pasts that predate philology. Bibliography also allows us to identify cousins and other contemporary relations in the modern humanities. Although bibliography is now usually associated with printed materials, especially early European printed material, expert bibliographers (e.g. Paul Otlet, Suzanne Briet, Jerome McGann, and especially Donald F. McKenzie (1999)) have insisted that bibliography, like information science, should extend beyond printed books to include all physical objects regarded as actually or potentially meaningful. Bibliography can be seen as having four, interrelated facets: enumeration, description, analysis, and critique. All concern copies and copying.

1. Enumeration is the process of listing copies. The word *list* was originally a term for the edging around and enclosing, a piece of textile. Hence a bibliographical listing determines what texts are included and what are not. Enumeration often identifies tokens of a type.

2. Description records selected features of listed items to reveal their characteristics. Cataloging and metadata creation are description and descriptive detail may be extended indefinitely until sufficient to differentiate similar entities such as different authors with the same name. In textual studies and elsewhere, description is often done in order to show relationships between enumerated items. The powers of bibliographical description are elegantly, if very differently, examined by Fredson Bowers (1949) and Patrick Wilson (1968).

3. Analysis tests enumerative and descriptive assumptions in order to make statements about the nature of any copy. The aim of these analytical tests is to examine how a copy has been contained by the categorical assertions of lists and the summarizing assertions of descriptions. Analysis enables an interpretive understanding of how differences and similarities have been drawn between particular copies, as well as the relationship between a particular copy and the category of object it is understood to copy.

4. Critique refers to the making of evaluative judgements based on more or less rigorous attention to enumeration, description, and analysis. Bibliographical critique concerns, for example, judgements about what should be copied again into new productions such as critical editions. In library contexts, bibliographical critique concerns judgements about collection development, for example, what to include in catalog records, what needs to be added, and how the collection and the records should be arranged.

5.7 The Humanities

In Europe, copies were for centuries the basis for developing skill in both rhetoric and art. The student would begin by copying a master (imitation) and then seek to adapt or improve it (emulation) (Fransen & Reinhart, 2019). In East Asia, copying and imitation were also of central importance. For example, what became known as the classics—*Classic of Changes (Yijing)*, *Classic of Documents* (*Shujing*), *Classic of Poetry* (*Shijing*), the *Rites* or the *Record of Rites (Liji)*, and the *Spring and Autumn Annals (Chunqiu)*—were copied and memorized for what might be thought of as rhetorical and artistic purpose but even more centrally for social advancement. The classics served as the basis for state exams that decided who might enter government and gain power and wealth. In addition to these historical roles, different kinds of copies (e.g. adaptations and parodies) are, of course, important to contemporary imaginative media (Beck & Ramtke, 2016; Boon, 2010; Schwartz, 1996).

There are diverse and long-standing debates in the humanities about the nature and force of copies. These discourses often concern representation, reproduction, simulacra, mimesis (imitation in literature), and power, to name a few key terms. As the bibliophile, translator, and cultural critic Walter Benjamin (1936) pointed out in his famous essay, "The Work of Art in the Age of Mechanical Reproduction," technological changes that alter the way that we copy can significantly affect the kinds of art we create, the kinds of communities we inhabit, and the kinds of politics we practice. Humanities discourse about copies is far too extensive and complex to be discussed here beyond what we have touched upon here and in our discussion of bibliography and philology.

6 DISCUSSION

6.1 Relatedness and what is not a copy

Our discussion thus far has been unified by an underlying assumption of relatedness through similarity. Document B can be considered a copy of document A when it has some useful similarity to A and when we are indifferent to differences between B and A. Status as a copy is a situationally understood relatedness arising from an end best achieved by an assumption of sufficient sameness. Document B can also be considered a copy of A when it shares usefully similar properties but its differences from A make it usefully relevant for some purpose. The infrared-based copy of the censored text in figure 2 and the transcription characteristics leading to Lachman's inferred genealogy of manuscripts are good examples. This too is situationally understood relatedness, a kind of relevance.

If document B does not share some usefully similar property with document A, then we would not regard it as a "copy" even though it may be related in other ways including being derived from A. Many possible relationships between objects are possible. We use three to suggest the diversity of objects that are not copies.

- *Transformation*: Copies are always particular formations (distinct in time and space) but not all formations can be considered copies, i.e. usefully similar. One can, for example, analyze a poetic text to examine the relative frequency of each combination of adjacent words, draw representations of what is found, and even materialize the representations of the poem using 3D printers (Yun, 2017). Here the 3D prints, which can be considered documents, are derived from and related to the historical, letterpress-printed poem. The 3D-printed objects are documentary formations derived from a printed poem, but we would hesitate to call them copies of the letterpress poem in most contexts. They may be useful but are not similar.
- *Exchange:* Usefully similar documents may be usefully interchangeable with each other when any one may serve the purpose at hand. A dollar bill in your pocket is usefully similar to the one in your friend's pocket. But an exchange of some bills for a cup of coffee does not make the coffee a copy of the money. A photocopy of the cash would be a copy, of course, and may serve the purpose of representing the shape and size of currency, but it is unlikely to be acceptable for a purchase because it is not usefully similar enough to be legal tender.
- *Semiotics*: Two objects can be related through a shared perceived meaning. Imagine a rock formation that looks like an elephant. In what sense is it an elephant? It is not literally an elephant, nor anything derived from an elephant, but it is perceived as resembling an elephant. Because it looks like an elephant, it has a shared property and can be said to be related to, hence relevant to, elephants, but one would not say that it is a copy of an elephant. In semiotic terms it is an iconic sign of an elephant. Semiotic relationships can contribute to a document's status as a copy but are not, in most cases, themselves copies.

6.2 Context and authority

Existing theory in knowledge organization and information science is greatly concerned with overcoming the effects of time and space through the dissemination and transmission of documents. This is why printing and digital transmission have been considered so important. A single copy of a printed edition can be considered the same as all the other copies for most

purposes. Similarly, a photocopy or digital scan would be acceptable for most tasks. Mass communication is concerned with the delivery of the same message to many people. The preservation of recorded knowledge through time and space is central to information science. Bibliography is very much concerned with determining the material similarity (or not) of two documents. Shannon's communication theory is concerned with producing copies, and assessing and reducing unwelcome corruption ("noise") in transmission. So is analytical bibliography (Greg, 1933; Bowers, 1949).

Paul Otlet's drawings illustrate well a common but naïve presumption of identical sameness of "content" through multiple diverse transformations (Otlet, 1934: 40-42). But this convenient view of how documents and, especially, multiple copies of them, provide uniformity across space and time is simplistic. The difficulty lies in the role of contexts when establishing what counts as usefully similar. There are multiple kinds of context for any document. One is the cultural context of the creator. Another is the cultural and epistemological context of a user who, often at a distance in space and time from where a document was created, is making use of it. What the user already knows or believes will affect how a document is understood and used. Although there has been some good work on this topic (e.g. Dourish, 2004; Lee, 2011; Agarwal, 2018), attention to the kinds of context created by users has been widely ignored. The ability of users to define context, and hence what might be usefully similar, is generally overlooked by researchers. An analysis of the canonical literature of library and information science by Allan Konrad concluded that the user's prior knowledge was simply not considered in nearly 90% of the writings examined and that half of the rest contradicted its relevance (Konrad, 2007, p. 508). There are also what might be called material contexts that contribute to making objects usefully similar. From the ways that books are positioned on shelves to the physical location of a searcher in relation to what can be retrieved, these material contexts are centrally important for much work in information science. This is to acknowledge the authority of material situations without anthropomorphizing books or the bookshelves on which they sit. But the problem is infrequently considered this way. Copies are presumed to be the same or similar without accounting for how material contexts contribute to making them so. Two pictures of an American flag appearing on a computer screen might be considered copies. Two cloth flags hoisted up adjacent flagpoles might also be considered copies. But, in many contexts, the image of a flag on the computer screen would not be considered usefully similar enough to the cloth flag to be considered a copy.

6.3 Decontextualization and its difficulties

Complicating matters as they concern context, a copy, as a particular documentary formation in space and time, is always at some remove from the context of the object it can be understood to copy. Individual copies cannot preserve the context of what they copy. In this sense, they are decontextualized. At the same time, copies, as we describe above, can provide context for one another. Indeed, the ability of copies to contextualize other copies is one of their powers. The contexts provided by copies enable many kinds of description, analysis, and critique. Indeed, they often contextualize how we know by establishing many of our primary frames of reference. They coordinate, for example, our sense of historical continuity and the context it provides by suggesting similarities between objects separated spatially and/or temporally. The many extant historical copies of Shakespeare's poems and plays ensure that we know Shakespeare and that he and his works have been brought to us from another place and time.

As examples of how copies decontextualize, consider Suzanne Briet's three famous examples. A photograph of a star documents the view. It is a copy in that the image is usefully similar to the sky it reproduces. What is not mentioned is that, while the image is usefully similar in the moment, the ways that it is similar and the usefulness of those similarities change rapidly. The image represents a view from a specific point in space at a fixed point in time while the universe moves dynamically on. A stone is removed from a stream; it is irreversibly changed by dissection and then moved to the very different environment of a museum display where comparison with other rock specimens enables us to understand it differently because it has been recontextualized. Its new context facilitates changes in the type-token relationships the stone can suggest as a document by reformulating what might be considered useful similarities (and differences) with rocks from distant streams. Its decontextualization helps to confirm the stone's status as a copy but also the variability of this status. The same can be said of the unlucky antelope that is detached from its natural habitat and from its social group to become a lonely prisoner in an unfamiliar environment (Briet, 2006, p. 10). All three examples are of irreversible detachment, of physical change associated with temporal and spatial separation into alternate contexts.

It is only in the formal realms of mathematics, logic, and computational theory that time, place, and context are abstracted away, that two objects can be considered logically the same, and that processes can be considered reversible. And even in these realms logicians and computational theorists run quickly into the quirky paradoxes of proof in Kurt Gödel's incompleteness theorems and corollaries such as Alan Turing's halting problem, both of which resist logical solution in the absence of context and/or external authority. Some may wish for information science to be in such a formal realm, but that would remove information science from the physical realities of information and the informative things we associate with human knowing and knowledge.

At the same time, the physical realities of information and informative things are themselves informed by conceptual abstractions, as we suggested above. Copies remain particulars, meaning that each is a unique material object (on paper, microfilm, electronic, or whatever) in time and space. But they are also regarded as abstract objects in types (classes) formulated by similarities and difference. Either way they inhabit and constitute contexts. If they are removed from one physical and/or conceptual context, they inevitably arrive in another and are, in consequence, differently positioned in space, time, and cognitive perspective. As Derrida said, there is no being out of context ("Il n'y a pas de hors-texte") (Derrida, 1967, p. 227). Copies as documents cannot be context-less even if they decontextualize what they copy.

Context is difficult to delimit when discussing copies because we are forced to shift back and forth between the material particulars of copies and the ways those particulars are formulated conceptually as part of a group of similar objects. Despite these difficulties, context is important to our discussion of copies because contexts determine what counts as a copy. This is to say that context determines how copies can be enumerated as well as the ways that they count as usefully similar objects. And if how context alters the status of objects has been largely overlooked in information science, it has been noticed elsewhere. The film maker Lev Kuleshov showed, for example, in the 1910s and 1920s that an expressionless face would appear to acquire different expressions if adjacent shots are varied (Kuleshov effect, 2020). It is a cognitive phenomenon by which viewers derive more meaning from the interaction of two sequential shots than from a single shot in isolation. Ludwik Fleck denounced the reliability of reference works which

emphasized brevity at the expense of context (Fleck 1935/1979). Canadian sociologist Dorothy Smith (1984) identified the significance of context for meaning. Communication theorist Wilbur Schramm addressed the differing contextual "frames of reference" of sender and receiver (Schramm, 1972, p. 31). The political theorist Anthony Giddens wrote, "Besides expanding the level of time-space distanciation, writing also opens the way for those divergences of interpretation which in modern historiography have come to be called 'hermeneutics.' Writing gives rise to texts that enjoy an 'objectified' existence independent of the sustaining oral traditions in daily social practices. The 'conflict of interpretations' engendered by texts is very closely related to ideology . . ." (Giddens, 1981, p. 150). The implication is that information scientists have been looking in the wrong direction, or, at least, casting a gaze that is importantly incomplete. There is some research on the information seeker and the seeker's context, but we also need a complementary research program on the document and its context and can note Bernd Frohmann's attention to this problem in a chapter entitled "The Materiality of Mme Briet's Antelope" (Frohmann, 2012).

7 CONCLUSIONS

The history and theory of copying has received relatively little attention in information science. Instead, much attention has been given to writing, printing, telecommunication, and digital computing. This is ironic since it is their ability to make copies that make these technologies important. The ability to copy appears to inspire a desire to create more copies. Our modern ability to copy has vastly increased the number and variety of extant copies, suggesting the etymological roots of the word "copy," *copia*, meaning abundance.

"Copy" can be understood narrowly as the outcome of a copying process and we may think of a copy as being the same as what it copies, but, strictly speaking, no two material objects can ever be entirely the same since each copy is a unique, material, particular. The source of a copy is usually called its "original." But in a broader sense no document is ever totally original and all documents are in some part derivative. In consequence, there is no fundamental theoretical separation between copies and other documents in terms of originality or sameness. It is a matter of degree that depends on purpose and context. So "copy" is also understood broadly for documents that are not derived directly from the same original but are usefully similar for some purpose in some situation. In practice, "the same" suggests a categorial similarity that makes objects acceptable substitutes, tokens of some type, specimens of some species, or members of some set. This same categorical similarity enables difference between usefully similar things to be informative and useful for forensic analysis, inferencing, and other purposes. Perceived status as a copy is only one of many kinds of relationships between documents. Others include transformation, fungibility (exchange value), and, in semiotics, sign relations.

Theorizing copies provides a basis for a more complete and unified view of information science. Three classic definitions of information science make limited sense unless interpreted to mean that information science is based on and concerned with physical copies of documents. Shannon's communication theory would be more aptly named copy theory. Search in information retrieval systems can be viewed as seeking copies. Cataloging models such as FRBR can be understood as ways that information scientists articulate useful similarities among documents based upon their opinions and preferences, as well as abstractions such as "work" and "original text," for the purpose of cataloging. Philology, the principal source of modern

humanities, is centrally concerned with copies and, in that sense, can be seen as anticipating information science. Bibliography can be understood as being concerned with the enumeration, description, analysis, and critique of copies. Diverse and long-standing discourses in the humanities concern the nature and force of copies. Future research on copy theory in information science can productively engage these discourses.

ACKNOWLEDGEMENTS

REFERENCES

Agarwal, N. K. (2018). *Exploring context in information behavior: Seeker, situation, surroundings, and shared identities*. Morgan and Claypool.

Association for Information Science and Technology. (2020). *What is information science?* https://www.asist.org/about/what-is-information-science/

Badia, A. (2019). *The information manifold*. MIT Press.

Beck, A. & Ramtke, N. (Eds.) (2016). *Imitat, Zitat, Plagiat und Original in Literatur und Kultur der frühen Neuzeit*. Peter Lang Edition.

Bendikson, L. (1932). Phototechnical problems, *Library Journal 57*, 789–94.

Benjamin, W. (1936/2005). *The work of art in the age of mechanical reproduction*. https://www.marxists.org/reference/subject/philosophy/works/ge/benjamin.htm

Boon, M. (2010). *In praise of copying*. Harvard University Press.

Borko, H. (1968). Information science: What is it? *American Documentation*, *19*, 3–5. https://doi.org/10.1002/asi.5090190103

Bowers, F. T. (1949). *Principles of bibliographical description*. Princeton University Press.

Briet, S. (2006). *What is documentation?* Scarecrow Press.

Buckland, M. K. (2017). Document theory. In B. Hjorland & C. Gnoli (Eds.), *ISKO encyclopedia of knowledge organization*. http://www.isko.org/cyclo/document

Buckland, M. K. (2006). *Emanuel Goldberg and his knowledge machine*. Libraries Unlimited.

Buckland, M. K. (1991). Information as thing. *Journal of the American Society of Information Science 42*(5), 351–360. http://people.ischool.berkeley.edu/~buckland/thing.html

Buckland, M. K. (1979). On types of search and the allocation of library resources. *Journal of the American Society for Information Science 30*(3), 143–47.

Buckland, M. K. (2021). Photocopiers. In A. Blair, P. Duguid, A.-S. Goeing, & A. Grafton (Eds.), *Information: A historical companion* (pp. 655-58). Princeton University Press.

Derrida, J. (1967). *De la grammatologie*. Editions de Minuit.

Dourish, Paul. 2004. What do we talk about when we talk about context? *Personal and ubiquitous computing 8*:19-30.

Fleck, L. (1935/1979). *Genesis and development of a scientific fact*. Chicago University Press.

Frangoul, A. (2018, March 14). *With over 1 billion users, here's how YouTube is keeping pace with change*. CNBC. https://www.cnbc.com/2018/03/14/with-over-1-billion-users-heres-how-youtube-is-keeping-pace-with-change.html

Fransen, S. & Reinhart, K. M. (2019). The practice of copying in making knowledge in Early Modern Europe: an introduction. *Word & Image: A Journal of Verbal/Visual Enquiry 35*(3), 211–222. https://doi.org/10.1080/02666286.2019.1628611

Frohmann, B. (2011). The documentality of Mme Briet's antelope. In J. Packer & S. B Crofts Wiley (Eds.), *Communication matters: Materialist approaches to media, mobility, and networks* (pp. 173–82). Routledge.

Giddens, A. (1981). *A contemporary critique of historical materialism*. Macmillan.

Greg, W. W. (1933). The function of bibliography in literary criticism illustrated in a study of the text of "King Lear". *Neophilologicus 18*, 241–262. https://doi.org/10.1007/BF01514481

Harré, R. & Llored, J. (2018). Procedures, products and pictures. *Philosophy 93*(2), 167–186. https://doi.org/10.1017/S0031819117000535

Hayes, P. (2011). On being the same as [Keynote address]. International UDC Seminar 2011, Classification & Ontology: Formal Approaches and Access to Knowledge, The Hague (Netherlands). UDC Consortium. http://www.udcds.com/seminar/2011/media/slides/UDCSeminar2011_PatrickHayes.pdf

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1997/2009). *Functional requirements for bibliographic records : final report*. International Federation of Library Associations and Institutions. https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf

Konrad, A. (2007). *On inquiry: Human concept formation and construction of meaning through Library and Information Science intermediation* [Doctoral dissertation, University of California, Berkeley]. http://www.ischool.berkeley.edu/research/publications/2007/konrad/dissertation

Kuleshov effect. (2020, August 2). In Wikipedia. https://en.wikipedia.org/wiki/Kuleshov_effect

Lee, C. A. (2011). A framework for contextual information in digital collections. *Journal of Documentation 67*(1), 96–143. https://doi.org/10.1108/00220411111105470

Lubetzky, S. & Hayes, R. M. (1969). Bibliographic dimensions in information control. *American documentation 20*, 247-52.

Malkan, J. (2005). What is a copy? *Cardozo arts & entertainment law journal 23*, 419-63.

McGann, J. (2014). *A new republic of letters: Memory and scholarship in the age of digital reproduction*. Harvard University Press.

McKenzie, D. F. (1999). *Bibliography and the sociology of texts*. Cambridge University Press.

Nestler, F. (2005). *Einführung in die Bibliographie*. Hiersemann.

Otlet, P. (1934). *Traité de documentation*. Mundaneum.

Owen, D. (2004). *Copies in seconds: how a lone inventor and an unknown company created the biggest communication breakthrough since Gutenberg: Chester Carlson and the birth of the Xerox machine*. Simon & Schuster.

Rettler, B. & Bailey, A. M. (2171). Object. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 ed.). Stanford University. https://plato.stanford.edu/archives/win2017/entries/object/

Saracevic, T. (2009). Information science. In M. J. Bates (Ed.), *Encyclopedia of library and information sciences* (3rd ed.) (pp. 2570–2585). Taylor and Francis.

Schramm, W. (1972). *The process and effects of mass communication*. Rev. ed. University of Illinois Press.

de Fremery and Buckland. Copy theory. *JASIST* 73, no. 3 (2022): 407-418.

Schwartz, H. (1996). *The culture of the copy: Striking likenesses, unreasonable facsimiles.* Zone Books.

Secretary hand. (2020, August 2). In *Wikipedia*. https://en.wikipedia.org/wiki/Secretary_hand

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal 27*, 379–423, 623–656.

Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

Smith, D. (1984). Textually mediated social organization. *International Social Science Journal 36*(1), 59–75.

Svenonius, E. (2000). *The Intellectual Foundations of Information Organization*. MIT Press.

Tanselle, T. (1989). Reproductions and scholarship. *Studies in Bibliography 42*, 25–54. http://xtf.lib.virginia.edu/xtf/view?docId=StudiesInBiblio/uvaBook/tei/sibv042.xml;chunk.id=vol042.02;toc.depth=1;toc.id=vol042.02;brand=default

Thompson, C. (2015). How the photocopier changed the way we worked—and played. *Smithsonian Magazine.* https://www.smithsonianmag.com/history/duplication-nation-3D-printing-rise-180954332/

Tillett, B. (2003, April 4–5). *The FRBR model (functional requirements for bibliographic records)* [Presentation]. ALCTS Institute on Metadata and AACR2, San Jose, CA, United States. https://www.loc.gov/catdir/cpso/frbreng.pdf.

Timpanaro, S. (2005). *The genesis of Lachmann's method*. University of Chicago Press.

Turner, J. (2014). *Philology: The forgotten origins of the modern humanities*. Princeton University Press.

Wetzel, L. (2009). *Types and tokens: On abstract objects*. MIT Press.

Williams, M. E. (1987/1988). Defining information science and the role of ASIS. *Bulletin of the American Society for Information Science*, *14*(2), 17–19.

Wilson, P. (1983). *Second-hand knowledge: an inquiry into cognitive authority*. Greenwood Press.

Wilson, P. (1968). *Two kinds of power: An essay on bibliographical control*. University of California Press.

Yeo, G. (2010). 'Nothing is the same as something else': Significant properties and notions of identity and originality. *Archival science 10*, 85-116.

Yun, S. Y. (2017, March 27). Turning poems into 3D artwork: Foreign professor offers Korean literature as high-tech edutainment. *The Korea Times*. http://www.koreatimes.co.kr/www/culture/2017/03/142_226462.html