Indexing: Philosophy of

André De Tienne

Indiana University Purdue University Indianapolis, adetien@iupui.edu

Follow this and additional works at: http://digitalcommons.unl.edu/docedit

Part of the Digital Humanities Commons, Other Arts and Humanities Commons, Reading and Language Commons, and the Technical and Professional Writing Commons
Indexing: Philosophy of

André De Tienne

Critical editions of great thinkers and writers need excellent, comprehensive indexes. The denser the text, the deeper the index. Having indexed several volumes of the writings of the polymathic, dense-and-deep philosopher Charles S. Peirce, I have had occasion to reflect numerous times both upon the art of indexing and upon its logic. This essay will discuss less the art of it (or its mechanics) than its logic—and, by the same token, its ethics. I have good reasons to do so: first, Peirce is the American founder of the logic of signs (also known as semiotics), and one of the major types of signs that Peirce identified and analyzed is called “index;” second, Peirce is the first thinker to have demonstrated that logic rests on ethics: the desire for truth rests on the desire for the good, not the reverse.

The immediate function of most book indexes is to help readers find the pages within the volume that are most likely to carry a particular idea, thought, or piece of information they are curious about. As such the index provides column after column of answers to the question “does the author discuss A or explain B, refer to C or mention D, and if so, where?” It behooves the indexer to imagine as many questions as possible, and to provide answers in a form most likely to satisfy the imaginary questioner. Although it would be tempting to theorize that any dense book could generate an infinite amount of such questions, that is fortunately not the case, and it would be easy to demonstrate that the set of such questions is always smaller than the set of words within the book. Both sets are finite, and since any text comes with a large number of words that common sense would not want to ask about (e.g., definite or indefinite articles, prepositions, conjunctions), that leaves enough room for those ideas that, though left unexpressed, are palpably present between the lines, and thus indexable.

A related consideration is that the amount of space allocated to a back-of-the-book index is generally a small percentage of the text, and this economic convention forces the indexer to select out many entries, that is, to refuse answering many possible legitimate questions a reader might have. Among consequences this convention entails is that readers might be misled into thinking that certain As, Bs, Cs, or Ds are not to be found within the
text, or might shake their head in frustration when chancing across an A, B, C, or D that the index failed to point out. The anonymous indexer could get cursed at for incompetence or arbitrariness, or even for being a liar. The careful indexer is aware that the word “good” in “good index” carries a moral connotation. A good index is not only one that answers most questions where they might be expected to be raised (hence the importance of judiciously crafted entries and cross-references), but also one that shows fairness to the object of its representation, the text. It should not misrepresent that text by neglecting the valuable ideas, skewing the fruitful thoughts, or hiding the significant pieces of information it contains.

The index as a whole, and each of its elements, act as a medium between the text and its readers or users. What kind of medium? Evidently, we are dealing with a particular kind of triadic relation. It so happens that Peirce defined a sign in general as a medium for the communication of a form. A sign, he wrote at the end of 1908, is anything that is so determined by its object, and so determines an effect, called an interpretant, upon a mind, that that effect is thereby mediately determined by the object.¹ Accordingly, any index entry is a sign intended to create an effect, called an interpretant, in a reader’s mind, that is such that the object of the entry (a particular passage in the text that determines the entry to point to it) mediately determines the reader’s interpretant to take note of it or be otherwise modified by it.

What kind of sign is an individual index entry? One answer is so obvious that it is hard to resist the temptation of stating it: an index entry is an index. It points to the location of a string of words within the text in a manner similar to someone’s pointing an index finger to something. “There it is, there you will find it!” “There” is meant by page number so or so, and “it” is meant by the entry (or subentry) that precedes said page number. Hence, a book index would merely be an alphabetized collection of index signs that keep answering “there it is” to the question “where is A, B, C, or D?” That this rudimentary semiotic explanation is an oversimplification of a more complex process, and thus a temptation that should be resisted, will be shown presently.

All signs are determined by their objects. This means that it is by virtue of some kind of influence exerted by the object that the sign acts as such and tries to bring that object (or some portion of it) to the attention of some other

logical entity (the interpretant). Let us use an example. Take the following indexable sentence chosen at random from a text by Peirce:

A symbol is a sign naturally fit to declare that the set of objects, which is denoted by whatever set of indices may be in certain ways attached to it, is represented by an icon associated with it.²

Let us now enter the working mind of an indexer coming across this sentence, and let us agree, for the sake of discussion, that the space allocated to the index is generous. Two possibilities should be considered. Possibility (1) is that the indexer, professional though he or she may be, does not understand the sentence, not being an expert in Peirce's logic of signs. Possibility (2) is that the indexer is both a specialist of Peirce's semiotics and an editor of the Writings of Charles Sanders Peirce. For short, let us call them P1 and P2. Both have in common a pragmatic goal: as their eyes peruse the text line by line, they are searching for objects that are worth creating signs for in the index because such objects are for one reason or another worth bringing to the attention of the reader—thus worth generating interpretants (further signs in the reader's mind) about the object pointed to.

Both P1 and P2 will see that the sentence says something about the Peircean concept of "symbol." Both are thus, at a minimum, likely to create the following entry: "symbol, 17." By paying a little more attention, P1 might even be stirred into creating the entry/subentry combination "symbol: definition of, 17," while P2 might aim for greater precision and state "symbol: its declarative function, 17." Despite a deep love for indexing books of all kinds, or perhaps because of it, P1 will probably move on to the next sentence with a sense of mission accomplished. P2, on the other hand, is a critical editor who has been working on the text for several years, has reconstituted the original manuscript, has edited it, and knows it and many others inside out. P2 fully appreciates the density of the sentence and is ready to index it thoroughly. His prime motivation is not a primal desire for over-promoting the thought of the beloved author, but the awareness that the sentence in question cannot be treated as an isolated atom. The sentence is a proposition that is a dynamic part of a long logical analysis, it is the conclusion to a sophisticated argument and is poised to become a major premise in several ensuing

arguments. P2 also knows that the book contains many different claims about symbols, and that these claims evolve from text to text in ways that the full entry under “symbol” ought to capture through a platoon of well-arranged, well-formulated, subentries and cross-references. P2 fully appreciates the fact that his mission is to provide his demanding readers with a reliable map that faithfully reproduces the diversity of treatment the concept “symbol” has received from the author through hundreds of pages. Hence, P2 understands two basic facts of his work: no sentence in the book can be treated merely atomistically since it is always part of a longer chain of reasoning, and index entries ought to be designed and formulated so that they offer, each one of them, a cohesive map rather than a collection of discrete, disconnected pointers to atomistic textual elements of the text. P2 knows that the many sections in the book offer at least a dozen statements that are variations of more or less exact definitions of the symbol, while there are many other statements of the sort just quoted that play complementary roles that are worth dissociating for the benefit of the readers.

Knowing in advance that “symbol” will be a long and complex entry, P2 will carefully design the set of subentries so that, so to speak, each subentry be aware of its neighbors. A good subentry will not try to steal the steam of others that had been created before it. Each subentry must make sure to carry a meaning that no other was already championing. When space is not too much of a factor, small to tiny differences of meaning may thus become dissociated into different subentries. If space is restricted, subentries will need to be brought to a higher level of semantic generalization, which means that certain differences will be erased for the sake of economy and, therefore, pointers to different passages will be lumped together. Indiscriminate lumping together of semantic differences entails an ethical cost. Not only will the reader need to spend more time sifting through the series of page numbers associated with the lumped subentry—something which admittedly does not rank high on the sin scale—but, appreciably worse and thus higher on that scale, is the risk that the subtlety and sophistication of the author’s thought will get shortchanged.

Indeed, the indexer’s primary mission is not so much to tell readers where to find As, Bs, Cs, or Ds—that is the essence of the secondary mission—than to tell them what are the As, Bs, Cs, or Ds that make an appearance within the text. That primary mission is not usually accomplished by a book’s table of contents. A table of contents provides an idea of the structure or organization of the book, it may even reveal to a large extent the evol-
tion of its conceptual plot, but it cannot do what only a good index can: identify fragmentary pieces of thought one after the other, and establish conceptual connections throughout the book that bypass the structure established by the table of contents at multiple levels. A good index is sufficiently reliable that it also, through its silences, tells the readers what the book does not contain. Does the author mention Socrates anywhere? If the index is silent, the reader is allowed to think that there is no such mention. More careful consideration will make the reader think that, if Socrates' name was actually in print somewhere inside the book, it was in a manner that the pragmatic indexer duly noticed, thought through, and concluded that the manner of its appearance was too sterile to bear any conceivable fruit (or interpretant) if pointed out in the index. Say the book does contain Socrates’ name twice on a particular page, in the following fashion: “All men are mortal; Socrates is a man; Socrates is mortal.” Fully aware that the author was discussing the form of a syllogism known as Darii and not saying anything about Socrates himself, P2 decides against creating an entry for Socrates. P1, though, much impressed by the aura of such a famous name, may find it more difficult to ignore it, and may create an entry, if only to cover all bases and be done with it.

Returning to the random sentence taken from Peirce, it is clear that P2, in his moral desire to do justice to it, cannot be satisfied with affixing only one entry to it. Much more is being said that needs to be reported to make sure appropriate interpretants get generated. Besides the symbol, the sentence more than alludes to two other types of signs: indices (plural form of our familiar index) and icons. P2 cannot but be moved by the fact that new pieces of information are here provided about those two important entities that are of course also discussed extensively elsewhere in the same work. Both are key elements in the function played by the symbol: icons and indices are signs capable of being brought together by the symbol so that the latter can perform a declaration that neither indices nor icons can utter on their own. Thus P2 will add new signs to his collection: “index: its auxiliary role within the symbol, 17” and “icon: object of symbolic declaration, 17.” Whether other components of the sentence are worth indexing, only the context will tell. Let us imagine that, beginning on page 25, Peirce starts discussing the conception of “natural fitness” and devotes to it a series of momentous paragraphs. Where P1 might be unmoved, P2 may feel pangs of anxiety if he hides the fact that in the random sentence of page 17 that very conception surfaces in a fashion that is actually not as innocent as it appears.
One has only to remember the habitual association of the conventional with the symbolical to wonder about the symbol's "natural fitness." Hence, P2 would feel morally better if, to the entry "natural fitness," he added the subentry "of symbols, 17." If, on the other hand, Peirce says nothing about natural fitness anywhere, P2 will probably follow the same course as P1 did, but for less despicable reasons.

Following both his train of thought and Peirce's own, P2 begins to wonder whether the structure of an index entry is actually as indexical as one would uncritically surmise. Considering again the wisely crafted entry "symbol: its declarative function, 17", P2 reflects first on the page number "17". What would make it an index? Well, there happens to be a page, not far from the beginning of the book, that displays the exact same pair of digits in its upper-right corner. It also happens that no other upper-right corner in the book displays exactly and only that same odd number. That number is a particular kind of sign that functions as it does because, through universal numeric convention, it is always found after the pair of digits 16 and before that of 18 within the teens. The formal rule that governs such a sign's construction and placement is extremely rigid and not likely to be changed by any arbitrary mind. This allows it to be used as part of a very convenient, simple and powerful, retrieval system. P2 next reflects that the number "17" has several characteristics or properties that are actually not relevant to its functioning in a book index. The fact that it is an odd number is barely relevant: it at most indicates that the page that bears it will be a recto side and not a verso side—but that is a kind of knowledge that is not indispensable to its use; it may help speed up a search, but not significantly so. That it happens to be a prime number is fully irrelevant, as is the fact that the addition of its two digits yields 8, or that it is itself the sum of 8 and 9 among other possibilities. Those numeric properties do not matter. The only one that does is its singular placement within an ordered series.

The number "17" would be, on that account, an indexical sign equivalent to a very precise "over there": "over there, on the seventeenth page of this book, not on the sixteenth, not on the eighteenth, but between those two." We cannot deny the sheer demonstrative power of that "17." But what makes it so demonstrative? We cannot claim that this "17" acts like a finger pointing straight to something in front of it. The actual page 17 is not in front of it...
the “17,” and “17” in the entry cannot be said to be acted upon by, or to be in some kind of contiguity with (as is required of genuine indices), page 17 itself. But “17” in the entry, in the first place, is a sign that looks like the sign in the upper-right corner of page 17. It is formally identical with it, and as such it serves, not as an index, but as an icon of it. A child who could read but not count could indeed use “17” as an iconic clue to find the page with a similar sign in the top margin. A child who could count, however, would see that “17” points to a particular place: “between 16 and 18.” That is clearly indexical.

If this is the case, then “17” in the entry is a sign that combines both an icon and an index, although it seems to be primarily iconic. It represents iconically any occurrence of itself in some other place, and it indexically denotes that place by virtue of the counting rule it instantiates. This is not true of the number 17 found in the upper-right corner of page 17: that number 17 is not primarily an icon but an index. It is a mark attached to that page and to no other within the book, and it happens to be upon it and so pointing at it by virtue of that page being exactly where it is within the ordered numerical series of pages inside the book. It is that page’s unique identifier, and it is in physical contiguity with it. It is a singular sign pointing at a singular page, and that is exactly what an indexical sign is expected to be doing. This does not mean that in such a circumstance 17 is a pure index (after all it embodies a counting rule), but that the semiotic function it plays most prominently on the page is indexical, not iconic, and not symbolic.

Yet, something else is at play that we cannot ignore. The number “17” in the entry is also a sign that spells out a rule of conduct to the index user. What rule? There are several ways of formulating it, depending on the index user’s competence. One candidate, assuming some level of competence, would be “if you jump toward the beginning of the book and flip rapidly through the pages, you will find one with an identical pair of digits in its upper-right corner, and the text on that page will tell you something about the entry this number is attached to.” This rule is here put in the second person because that is how the indexer is speaking to the would-be reader. But the reader encountering “17” in the index entry would turn it into the first person: “If I were to flip the pages toward the beginning of the book, I should be able, according to the anonymous indexer (or according to this sign here), to find a page, bearing the same pair of digits, that will tell me something about the entry this number is attached to.” According to Peirce, this conditional proposition, or rule, is a form that the sign “17,” by virtue of
its appearance in a back-of-the-book index (thus, by virtue of its context), embodies and communicates to the interpretant, and what fuels that process is the fact that the form is actually generated by the object of the sign (and not by the sign-maker).

The “17” in the entry at the back of the book, from this standpoint, is not primarily iconical or indexical. The fact that it spells out a rule of conduct, Peirce has shown, is a function that only a symbol can fulfill (this is not the place for a demonstration). Spelling out a rule of conduct is a kind of declaration. It declares that the “set of objects” denoted (indexically) by the number 17 printed on them (there is only one such object in a book, though), is truthfully represented by the (iconic) occurrence of the identical-looking “17” in the entry, and that readers should therefore conform their conduct to that reported state of things. That declaration does not reside, to be sure, in the iconic occurrence of “17” itself printed after a comma next to the entry, but in the fact that that occurrence is taking place at the end of the book in a place called “index,” and that placement adds an extra meaning to that “17”–it is superimposed onto that “17.” A contextualized icon is indeed never a mere icon. That such a declaration is the function of a symbol is precisely what the random sentence quoted earlier happens to state serendipitously.

This brings P2 to ponder the semiotic nature of the entry itself: “symbol: its declarative function.” Could that be an index? Surely not purely so, whether one considers it separately from its context in the back of the book and thus from the “, 17” that follows it, or not. Structurally, it consists of a complex symbol that links the property “declarative function” to the concept “symbol” by way of a subordinating colon. Much thought went into the production of that entry—a kind of thought that no automated indexing tool could match, if only because the word “function” is not physically present in the original sentence, although it is operatively present, but its detection requires inferential and thus interpretational power that transcends binary sniffing. The thought that produced the entry was a thought busy reading, analyzing, sorting out, the sentence on page 17 of the book. The formulated entry constitutes one of the interpretants that the sentence-sign elicited within the mind of the sign-maker (or indexer). Many interpretants may get elicited, and it is the indexer’s task to select that one that seems at once best to represent what is going on in the text and most likely to be a form the

But as one would expect from any embodied symbol, such an entry is also at once iconic (it provides a diagram, through the colon, of a conceptual relation that seeks to mimic a relation P2 detected within the random sentence) and indexical (it was solicited by a relation really present in the text). As a contextualized symbol, the entry also declares that the actual relation detected within the random sentence is fittingly represented iconically by the colon-mediated subordination between “symbol” and “declarative function.”

140  Documentary Editing  28(3) Fall 2006
reader will either be looking for or be able to correctly understand. That process is evidently primarily symbolical. By contrast, a computer-generated concordance is not a list of interpretants but an alphabetically sorted list of strings of characters fished from the text, sometimes according to a sophisticated algorithm calculating string frequency, matching strings against a pre-loaded list (or thesaurus), and playing the Boolean game. No matter how thoughtful its programmer may be, to an algorithm a text is a blind input, and the list of strings it generates a blind output. Readers can often tell the difference between mind-made and machine-made indexes. An algorithm can easily generate “symbol, 17” and many other such strings or combinations of strings, and to some extent this may be useful, but never “helpful,” for algorithms are amoral. A good index—an ethical one—needs to tell readers what the text talks about, here and there, everywhere, not what strings of characters the text is, or not, made of. While the electronic base of a search engine consists only of soulless pointers to character strings, a conceptual index consists of soulful “declarations” about the identity and whereabouts of thoughts.

Mind-made conceptual indexes sometimes attract scholars’ scorn. Their main objection is that the indexer, whether of the P1 or the P2 kind, is imposing on them an interpretational grid that they would rather do without. They tend to praise sophisticated search engines because the latter allow them to conduct thorough Boolean searches that are fast and bias-free. This is undeniable: binary searches, being asemiotic, are of course bias-free since they are thoughtless. But their success depends entirely on their users’ knowledgeability and competence. Specialists of Peirce, for instance, already know a great deal about his writings and are therefore far more able to squeeze well-targeted textual passages out of a search engine. The non-specialist who has a fainter idea of what Peirce talks about will be limited to searching words more or less at random, hoping for a few meaningful hits. The fact is, conceptual indexes and search engines fulfill very different functions, and trying to compare them in a competitive spirit is to commit a category mistake—a logical temptation hard to resist, but a fallacy nonetheless.

Scholarly (or P2-) produced indexes are sophisticated research tools produced for the benefit not only of lay-readers but also of P2’s peers. A well-made P2-index constitutes a genuine intellectual contribution to the indexer’s academic field, and thus a service to the profession, precisely because of the earnest interpretational effort that went into its production, an effort based on P2’s own competence in the field. Scholars who use a P2-index are just as free to dissent with the indexer’s formulation and selection.

_Documentary Editing_ 28(3) Fall 2006 141
of entries as they are with any paper of their peers discussing the text. These same scholars, however, are also likely to benefit from the indexer's special insight, for P2 will, just as any serious interpreter, identify connections that would have escaped the reader. One momentous reason is that P2, through a systematic and thorough reading, is bound to see certain conceptual patterns emerge from the text that would escape the attention of the non-indexer—patterns that even the author of the text may not have been aware of. Indexing a text requires indeed a particular mindset, one that most readers are not trained and have no reason to adopt themselves, and one that is distinct from the author's.

A P2-index of a several-hundred-page dense-and-deep book takes several weeks of concentrated work. This is not a labor that can be rushed through a tight deadline. Each sentence needs to be submitted to attentive scrutiny, one that will wonder about two types of things: what significant concepts are playing a role in that sentence, and what role that sentence itself is playing in its immediate context and in the wider scheme of the book. Reading both the lines (direct reading) and between the lines (meta-reading) will frequently force P2 to backtrack in the index, rephrase or readjust existing entries, and create new ones for pages passed, because the author's developing argument may bring P2 to realize that he had earlier failed to grasp the importance (and thus indexability) of a particular concept, or had somehow misrepresented it. It will also put him on alert and sharpen his analysis of subsequent passages.

P2 listens attentively to the admonitions of his conscience. P2 knows his fallibility and strives to lessen its impact by exercising critical common sense. He has the power to draw attention to passages that otherwise might be neglected, and vice versa. No matter how some index users may be suspicious of P2's intentions, though, one important fact remains: it is impossible for P2 (perhaps not for P3 if P3 is defined as an immoral indexer) to be willfully arbitrary. Not only does P2's moral sense prevent him from it, but so does also the logical (or semiotic) nature of indexing. Entries are not dictated by P2's detached mind but by the text's influence on P2's inner sign-making agency, an influence that may get as deep as P2's past conceptual experience is broad. The interpretants that emerge at the tip of P2's pen do so on their own, without soliciting P2's consent. P2's meta-mind will sort them out, select and shape them based on a variety of considerations, but as long a P2 remains of good faith, his entries will merely represent a faithful report of the text's influence on his indexing mind (not to be confounded, for instance, with
P2's *philosophizing* mind, which constitutes another kind of mindset.

The fact that no two P2-indexers would ever produce an identical index for the same book is no argument against the freedom from arbitrariness just advocated. No two P2-indexers share an identical background, range of intellectual interest, or habits of writing, and therefore one cannot expect the text’s influence on their indexing mind to be identical. Both, however, will go through a comparable semiotic process, the logic of which will be identical, so that their indexes, though different, will actually be comparable as well. Everything else being equal (including space restrictions and general competence), the degree of comparability will be inversely proportional to that of arbitrariness. True, any given P2 will fail to index a few entries another P2 would not have ignored, but the fruits of no P2’s labor should fuel the resentment of disgruntled scholarly readers. For such readers, too, have a moral responsibility, which is to accept P2’s fallibility as much as their own, to recognize that no P2-produced index may relieve them from the duty of reading the book and making up their own mind about the text, and to welcome P2 as a valuable partner amidst their inner circle.

After going through all these reflections, P2, finally satisfied with his analysis of the "random sentence," is ready to move on to the rest of page 17. And P1? Ah, P1 is happy, too: publisher’s check in hand, P1 is already busy indexing another book.