

Where's Waldo?

Reflections on copies and authenticity in a digital environment*

DAVID M. LEVY

Introduction

You have probably seen the «Where's Waldo?» children's books. Each double-page spread contains drawings of hundreds of cartoon figures. Your job is to find Waldo, a character who is always dressed in a red-and-white striped woolen cap and shirt and is wearing glasses. Often there are characters who look a lot like him, but if you look closely you can see that some detail or other is wrong (*e. g.*, it is a woman, the cap is solid red). In other words, only one of the figures on the page is the *real* Waldo; the rest are impostors, look-alikes, or close matches. «Pay attention» – these drawings seem to say – «Appearances can be deceiving.»

Waldo presents the problem of authenticity in graphical form. Although a number of the cartoon figures *seem* to be Waldo, only one *is* the authentic Waldo. Being authentic in this case means being who or what you seem or claim to be. In Waldo's case, there can only be one right answer, since we are talking about a unique individual. But in other cases, there may be more than one right answer. This happens when we are concerned with, say, group membership (being a medical doctor) or with types (being a 1956 Chevy). It is only because

* Copyright by the Council on Library and Information Resources, Washington, DC (May 2000). Republished with permission.

Neste número publica-se mais uma comunicação apresentada ao Workshop *Authenticity in a Digital Environment*, sendo as restantes comunicações a publicar nos próximos números das **PÁGINAS**.

Relativamente ao artigo de Charles Cullen, publicado no n.º 6, há a seguinte correção a fazer: na p. 67, linha 21, onde se lê «*sonic*» deve ler-se «*solar*».

we live in a world of multiplicity – where several people or things may appear to be the same – that duplicity is possible. Judgments of authenticity, as I understand it, allow us to navigate through a world by distinguishing genuine multiplicity from duplicity.

In the realm of written forms – in the world of paper and other tangible media – we have, over the centuries, developed elaborate procedures for identifying authentic documents and for ferreting out impostors. In the digital realm, we have barely begun to do this, and there are many technical and social challenges to be met. One challenge comes from the fact that the digital realm produces copies on an unprecedented scale. It is a realm in which, as far as I can tell, there are no originals (only copies – lots and lots of them) and no enduring objects (at least not yet). This makes assessing authenticity a challenge.

What are documents?

I use the word *document* where others might use *text*, *record*, *information-bearing artifact*, or *written form*. «Document» is a cover term for a large group of artifacts, including textual materials, whether handwritten or mechanically realized; graphics and photographs; and audiovisual presentations. But by what criterion do all these things fit into a single, coherent category?

I have come to understand documents by analogy with human beings. Documents are surrogates for people. They are bits of the material world (stone, clay, wood pulp, and now silicon) that we create to speak for us and take on jobs for us. A receipt bears witness to and thereby validates a financial transaction; a restaurant menu speaks for the establishment, the restaurant; a novel tells a story; a political flyer speaks for a candidate or political organization; and so on¹. By saying that documents «speak», I do not mean to limit them to textual or verbal materials. Pictures, drawings, diagrams, moving images, and other conventional forms of communication also speak in the metaphorical way in which I am using the term: they communicate, they tell us things about the world. And when I say documents «take on jobs», I am referring to the way we tailor their form and content to particular tasks and contexts. Genre (whether a

¹ There are great complexities and ambiguities regarding who is speaking in or through a document. In literature, for example, distinctions have been made between the narrator, the implied author, the «real» author, etc. Such complexities and ambiguities also exist, however, when a human being is speaking.

receipt, a menu, a novel or a flyer) is, in effect, the clothing of conventional content to do particular tasks in the world (to witness a financial transaction, recite the dishes available and their prices, etc.) (LEVY 1999).

For a document, speaking *per se* is not enough. It also must be able to speak *reliably*: We depend on documents to carry messages through space and time. In many cases, this reliability is achieved through fixity: letterforms inked on paper can survive for long periods of time. But with newer media, such as video, this reliability is achieved not by fixity but by repeatability. The moving images on a video screen are by their very nature transient. I will never be able to see those very images again. But I can play the tape repeatedly, each time seeing a performance that, for all practical purposes, is «the same as» the one I saw the first time.

If documents are meant to be reliable surrogates for human beings, then it makes perfect sense that we would be critically concerned with their authenticity. Steven Shapin (1994), a sociologist, argues that human social order – that human life itself – is fundamentally based on trust, *i. e.*, on our ability to rely on one another.

«How could coordinated activity of any kind be possible if people could not rely upon others' undertakings? No goods would be handed over without payment, and no payment without goods in hand. There would be no point in keeping engagements, nor any reason to make engagements with people who could not be expected to honor their commitments», he writes. Much as we rely on one another, we also have come to rely on documents in the making and maintaining of a shared, stable, social order. So it is no accident that words such as *trust*, *reliability*, and *truthfulness*, which are fundamentally social, would apply to documents as much as to people. It is likewise no accident that documents, as surrogates for us, would be accountable in the same terms.

What is a copy?

I worked for Xerox for a number of years, so it should hardly be surprising if some of my thinking and my examples come from the world of photocopying. In that world, «to make a copy» means to put one or more pieces of paper on the photocopier platen or in the RDH (recirculating document handler) and push the Big Green Button. What comes out at the other end of the machine is a «copy». In this context, a copy is something that is the result of a *process* of copying. It says

nothing about whether the result is a good copy or a bad copy, or whether or not it is useful.

But there is a second notion of copy, which has more to do with the *product* than the process. To be a copy in this sense is to stand in a certain relation to an original, that is, to its origin. To be a copy in this sense is to be faithful to the original. The definition of «faithful», however, depends on the circumstances in which the copy is being made and on the uses to which it will be put. The context of use, in other words, determines which properties of the original must be preserved in the copy. Does it matter that I have just made a photocopy of a signed will? It depends on what I intend to do with it. If it is for informational purposes (to show you what my will says), then it is an adequate copy; for some legal purposes, however, it won't do.

The point is, a document can be *identical* only with itself, if «identical» is taken to mean «the same in every respect». When we say that something is «the same», we generally mean one of two things. We either mean that it is «the very same» thing (as in «This is the same car I drove yesterday») or that it is «of the same type» as something else («I read that same book last year»). It is this second notion of sameness – sameness of type, sameness in virtue of sharing certain properties – that is at issue in copying (LEVY 1992).

Even an extremely high-fidelity copy will be different from the original in innumerable ways, because to copy is to *transform*. The copy will be on a different piece of paper that has its own unique properties. The process of photocopying will make letterforms thicker or thinner than those on the original, and will make images lighter or darker; it will add noise or remove it; it will change tones, shapes, aspect ratios, and so on. Differences will always be introduced in copying; the trick is to regulate the process sufficiently so that the resulting differences are of little or no consequence and that the properties of greatest consequence are shared. Determinations of which properties matter are made in the context of purpose and use.

Copying without an original

I have presented a simple and straightforward notion of copying. Although I have used the photocopier to illustrate how it works, this notion is not dependent on any particular technology. Making a copy by hand embodies the same idea. Moreover, although I have talked about making a single copy, one can obviously make multiple copies of an original – an indefinite number, in fact. It is common for someone

to create a «master» document and to produce any number of copies from it. What is crucial in this scheme is that there is an *original* from which the copies are made.

But there is another scheme – one that does not require an original. It is a manufacturing technique, a means of producing a large number of artifacts from a single source. If you want to make coins, for example, you can create a mold and pour molten metal into it to cast the coins. This is also the way the printing press works. You create a set of printing plates that are used to produce inked pieces of paper.

The reason I say there is no «original» in this technique is that the *source*² from which the copies are made (the mold or the printing plate) is a very different kind of thing than the copies³. You cannot spend the mold (although you may be able to mint more coins); you would not normally choose to read the text on the printing plate. This means that the word *copy* is being used in a somewhat different sense. It perhaps harks back to the root meaning of the word (copious, plentiful). But there is another sense in which the artifacts produced in this way are copies: they are copies *of one another*. Indeed, to a large extent, the purpose of this technique is to manufacture a set of «identical» artifacts – artifacts that are all «the same», that is, of the same type. These artifacts are identical in the sense that they are interchangeable with one another for certain purposes.

The examples I have given so far involve the production of enduring physical artifacts, or things. But this method of copying from a source also works for producing activities or events, which by their very nature are transient. Consider the case of a play, where a script (the source) serves as the basis for a number of performances (the copies) or an audio or videotape (the source), which leads to the realization of sounds or visual images, or both.

In none of these cases, however, is the source ever enough. Manufacturing the intended artifacts also requires a complex of skills, know-how, and, often, technical equipment. The mold for coins is useless without the right metals and the skill to do casting; a printing plate is useless without a printing press and knowledge of how to use it;

² I will use the word *source* to designate the thing from which copies are made in this method, and the word *original* when I mean something that is of the same kind as the copies.

³ I do not mean to suggest that there can *never* be an original that is used to guide the making of the source. I may print an edition of *Leaves of Grass*, taking the text from the 1891 edition. In this case, some actual printed copy of the 1891 edition is my original. Nevertheless, the production of my new edition is mediated by the printing plates I have created, and these plates are *not* an original.

the script needs a cast of actors; and the videotape needs a video player. In each case, the quality of the product or the performance depends on a skillful and properly executed process of production. The source, in other words, does not and cannot fully specify the properties of the things it is used to make. There is a division of responsibility between the source and the environment in which it operates.

It is worth comparing print with analog audio or video recording before talking about the digital case. In the case of printing, the source is used to produce a definite number of copies, an edition. Each copy in an edition is a stable physical object whose existence is independent of the source. But in the case of the recording, when the tape is defined as the source, there is no notion of a definite number of copies (*e. g.*, replayed performances); rather, once you have the tape and an appropriate player, you can produce a (relatively) unlimited number of copies, or performances. Moreover, unlike the products of print, the copies are completely dependent on the source for their existence. Should the tape be damaged or lost, there will be no more performances. This gives the source a greater importance in the case of recordings. You *have* to preserve it if you want copies in the future. (And, of course, you have to preserve the player, which is the means of making copies from the source.) In the case of printing, by contrast, once the source has done its work, it is no longer needed. (Indeed, the advantage of movable type is that it can be reused, *i. e.*, the elements of the source can be recycled.)

Digital documents

Like printed documents and recorded audio and video performances, digital documents are founded on a distinction between a source and the copies produced from it. The source is a digital representation of some kind, a collection of bits. The copies are the sensible impressions or manifestations – text, graphics, sound, whatever – that appear on paper, on the screen, and in the airwaves. Getting from the source to the copy requires a complex combination of technical and social environment, including an elaborate configuration of hardware and software.

In one sense, digital technologies are very much modeled on the printing press. They allow users to create what amount to digital printing plates from which they can «print» an arbitrary number of copies. The relation with traditional print is particularly strong when the copies produced are textual and graphical in nature, as is so much of the material on the Web today. But digital documents, even those with textual content, share significant features with analog au-

dio and video recordings as well. With audio and video, we tend to think of the source (in this case, the audio or videotape) as more permanent than the copies produced from it (the performances), which are inherently transient. Currently, we seem to be importing this same hierarchy of permanence into the digital domain. We think of the digital source (such as a Microsoft Word file) as more permanent than the text and images that appear on the screen. This makes sense, because we know how to «save» the file. When we have done so, it will typically survive on a hard drive or a floppy despite power loss, whereas the screen image cannot. But as we adopt this way of thinking, we are also coming to treat paper copies (analogous to screen images) as more transient than the source file. We often print out a paper copy to read and then toss it away, confident that we will be able to print out another as long as we have the file. But the truth is, at least for the moment, that paper has a better chance of survival than a digital source.

Indeed, digital entities are generally less stable than their counterparts on paper and other tangible media, and digital production tends to yield much greater variability of product than analog production does. In the case of print, once we have the plate and a press, the amount of variability is limited. Even more so is this the case with an analog recording: once we have the tape and an appropriate player, the amount of variability in performances is typically fairly well constrained. The differences generally are limited to minor variations in quality. For digital copies, however, there is likely to be a much greater range of variability. Some of the variability is intentional and it is a great strength of the technology. We can easily edit digital documents and quickly produce variants. Some variability is unintended and is an unresolved problem: digital copies are extremely sensitive to the technical environment, to the point that features we would like to preserve in subsequent copies may be hard (or impossible) to maintain. Displaying the file on a different computer may lead to font substitutions, different line breaks, and so on. These same sorts of variability may even occur on the same computer if, in the interim, the environment has changed in some crucial way⁴. Consequently, two different viewings of the «same» source may differ in important ways – they may not be «the same».

Under such circumstances of radical variability, there does not appear to be anything like a stable document or object. Over time, the digital

⁴ As sound and motion are digitally recorded, issues of uncontrolled variability will increasingly arise here, too.

source may move from server to server. The version that ends up on your local computer may have been copied from a server and will likely have undergone further transformation; for example, your local browser or editor may generate other local, and possibly partial, digital sources in the process of creating something you can actually see. What you do see at any given moment will be the product both of the local digital source and of the complex technical environment (hardware and software), which is itself changing in complex and unpredictable ways. The digital source, the perceptible copies, and the environment are all undergoing change in ways that no one yet knows how to control.

Authenticity in a digital environment

Assessments of authenticity in the world of paper and other stable, physical media rely heavily on the existence of enduring physical objects. If you want to determine whether the document in front of you is the unique individual it purports to be (someone's last will and testament, for example), you can try to determine its history. But you can do this only because it *has* a history, an extended existence in time. If you want to determine the authenticity of something that is one of many (a member of an edition, for example) you can compare it with another copy, a reference copy. And even where the thing in question is transient (such as the performance of a play), you still may be able to make use of a stable reference object (such as the script). In all these cases, either the object in question or a reference object has an enduring, physical existence that helps ground the determination of authenticity⁵.

What happens in the digital case if there are no stable, enduring digital objects? One possibility is that we will find a way to *create* them. In one current view, objects are at least in part socially constructed; they are bounded and stabilized through social interaction (SMITH 1996). Literary works (*e. g.*, *Hamlet*) are a clear example of this. Although we cannot really say what works are, we have nonetheless created a cultural mechanism (copyright and the courts) to help us decide where the boundaries between works lie. Here there can be no question of ultimate, natural answers – only social answers based on law and politics. In the digital domain, I see Jeff Rothenberg's pro-

⁵ How do we know whether to trust the authenticity of reference objects? The whole process recurses. I agree with Clifford Lynch, who suggested in his presentation at this workshop that the process is ultimately grounded in our trust of others. The «buck stops» when we accept someone's (or some institution's) claim that some object in the chain of reasoning is authentic.

posal (in this collection) to stabilize digital environments through emulation as one attempt to create stable digital objects. (I am not sure it is a workable solution, but that is another matter.)

Without the security of stable digital objects, what might we do? One possibility would be to maintain audit trails, indicating the series of transformations that has brought a particular document to the desktop. Such a trail (akin to an object's provenance) could conceivably lead back to the creation of the initial document or, at least, back to a version that we had independent reasons to trust as authentic. Having such an audit trail (and trusting it) would allow us to decide whether any of the transformations performed had violated the document's claimed authenticity. A second possibility would ignore the history of transformations and would instead specify what properties the document in question would have to have to be authentic. This would be akin to using a script or a score to ascertain the authenticity of a performance.

Conclusion

I have no conclusion other than this: Understanding what we *want* to accomplish, and what we *can* accomplish, with regard to authenticity in the digital realm will take considerable effort. If nothing else, this workshop has convinced me of the cultural importance, as well as the difficulty, of the work that lies ahead.

REFERENCES

LEVY, D. M.

- 1992 «What Do You See and What Do You Get? Document Identity and Electronic Media». In *Screening Words: User Interfaces for Text; Proceedings of the Eighth Annual Conference of the UW Centre for the New OED and Text Research*. Waterloo, Ontario: University of Waterloo – Centre for the New Oxford English Dictionary and Text Research
- 1999 «The Universe Is Expanding: Reflections on the Social (and Cosmic) Significance of Documents in a Digital Age». *Bulletin of the American Society for Information Science*. 25:4 (1999) 17-20

SHAPIN, S.

- 1994 *A Social History of Truth: Civility and Science in Seventeenth-Century England*. Chicago: The University of Chicago Press

SMITH, B. C.

- 1996 *On the Origin of Objects*. Boston: MIT Press

DAVID LEVY is an independent consultant in the areas of documents, digital libraries, and publishing. Between 1984 and 1999, he was employed as a researcher by the Xerox Palo Alto Research Center. The focus of his work over the past decade has been the nature of documents and the tools and practices through which they are created and used. His research interests include digital libraries; the reuse of documents; document design, structure, and standards; work practice studies; the combined use of paper and digital media; and the relation between technology and the character of modern life.

Mr Levy holds a Ph.D. degree in computer science from Stanford University and a Diploma in calligraphy and bookbinding from the Roehampton Institute, London. He is currently a member of a National Research Council commission charged with advising the Library of Congress on its technology strategy. He is completing a book, *Scrolling Forward: Making Sense of Documents in a Digital Age*, which will be published by Arcade.