
ARTICLE

THE COPYRIGHTABILITY OF NEW WORKS OF AUTHORSHIP: “XML SCHEMAS” AS AN EXAMPLE

*Trotter Hardy**

TABLE OF CONTENTS

I.	INTRODUCTION: HOUSTON–E-COMMERCE AND I-PROPERTY	856
II.	COPYRIGHT AND NEW SUBJECT MATTER: DIFFERENT APPROACHES	858
A.	<i>Rationales</i>	861
1.	<i>Dictionary Definition Approach</i>	861
2.	<i>Standards and Compatibility Approach</i>	866
3.	<i>Merger Approach</i>	868
4.	<i>Policy-Oriented Approach</i>	871
B.	<i>Unnecessary Issues</i>	873
1.	<i>The “Need to Compete”</i>	873
2.	<i>Standards and Compatibility</i>	874
3.	<i>Official Standards</i>	876
4.	<i>Antitrust and Monopolies</i>	879
5.	<i>Antitrust Tying</i>	881
C.	<i>An Incentives-Based Analysis of New Subject Matter</i>	884
D.	<i>Framework Applied to Past Cases</i>	887
1.	<i>Synercom Technologies</i>	887
2.	<i>Lotus v. Borland</i>	889
3.	<i>Sony v. Connectix</i>	890

* Associate Dean of Technology and Professor of Law at William & Mary School of Law. The Houston article nags know something about meandering-around law faculty in far off lands: Keep soliciting.

4. Bender v. West.....	892
III. WHAT IS XML?.....	897
A. XML.....	905
B. Schemas.....	906
IV. ARE XML SCHEMAS COPYRIGHTABLE?	908
A. Dictionary Definition Approach.....	911
B. Merger Approach.....	912
C. Standards and Compatibility Approach.....	913
D. This Article's Approach.....	914
V. CONCLUSION.....	918

I. INTRODUCTION: HOUSTON–E-COMMERCE AND I-PROPERTY

The technical news these days often refers to something called “XML” and “XML schemas.” For example, Microsoft announced in the spring of 2001 that it would combine a variety of its services, including its MSN online service and its “Passport” Internet authentication service, into a combined “Hailstorm” initiative.¹ A crucial part of this initiative was said to be a reliance on “XML schemas.”²

For e-commerce lawyers, any new means for conducting business should be a matter of great interest. For the computer industry and for intellectual property lawyers, anything Microsoft announces raises questions about proprietary rights. For all these groups, “XML” and “XML schemas” have now become a thing to be understood.

Part II.A of this Article reviews several different approaches that courts have taken in the past to “new subject matter” cases. Some courts, for example, have relied heavily on a “dictionary definition” approach, inquiring into the meaning of various words in the copyright statute and their applicability to the facts at hand, but not explicitly inquiring into policy issues, congressional intent, or decisional consequences as a guide to their interpretation. Other courts have invoked the need for standards and compatibility in construing the scope of a plaintiff’s copyright protection. A market leader—such as Lotus 1-2-3 which was at

1. See Press Release, Microsoft Corp., Microsoft Announces “HailStorm,” a New Set of XML Web Services Designed to Give Users Greater Control (March 19, 2001) (outlining the company’s new technology allowing for integration of existing Web-based services), <http://microsoft.com/presspass/features/2001/mar01/03-19hailstorm.asp>.

2. See *id.*

one time in the spreadsheet market—may be given a narrower scope of protection if the court believes that doing so will further a desirable standardization in the spreadsheet market. The familiar doctrine of “merger” is another common form of analysis in many subject matter cases. If a would-be new form of copyrightable subject matter appears to have its expression dictated by its underlying ideas, as was true of the accounting method in the famous *Baker v. Selden*³ case of 1879, the court will of course deny copyrightability. Finally, at least one court in a 1902 case drew almost entirely on policy concerns, even in the absence of statutory authority, creating an analysis that was the very opposite of the “dictionary definition” approach: all policy, no definition whatsoever.⁴

Part II.B of this Article begins to set up a proposed new analytical framework for better addressing “new subject matter” cases by first disposing of several non-issues. In particular, I show that it is not necessary or desirable for courts to take account of many of the things that they have in the past, such as a need for standards and compatibility, in considering the potential copyrightability of some new subject matter. The need for standards, and its close cousins—the need for competition and the desire to avoid monopolies—are central concerns of antitrust law. Copyright law is unnecessarily complicated by importing these antitrust doctrines into the question of what should and should not be treated as copyrightable subject matter. Moreover, court discussions of competition in copyright subject matter cases can easily lapse into a misleading over-simplification, namely that “competition” is an all-or-nothing thing, whereas it is, in fact, a matter of degree—another reason for courts to avoid antitrust and competition doctrine when it is unnecessary to a decision in the copyright case at hand.

Part II.C of this Article sets out the suggested new approach. This approach essentially asks whether the public would be better or worse off with creation of a copyright right. Making that determination requires consideration the incentive effects of copyright’s protection, a consideration that sounds utterly natural for a copyright case, but one that nevertheless surprisingly few courts have undertaken explicitly. A consideration of incentives leads me to propose an analysis that rests on three factors: whether the application of subject matter protection to a new kind of work will substantially increase the quantity of such works; whether the new kind of work will

3. 101 U.S. 99 (1879).

4. *Nat’l Tel. News Co. v. Western Union Tel. Co.*, 119 F. 294 (7th Cir. 1902).

exhibit a potential range of variation in quality; and if so, then whether the potential consumers of such works would be likely to care about those qualitative differences. The remainder of this part of the Article applies the suggested approach to several past subject matter cases, comparing the outcome of that analysis with the actual outcome in each case.

Parts III and IV try to give the general issue some focus by looking at a particular type of new subject matter: "XML schemas." The technical literature is already full of discussions of these schemas, as well as of "XML" more generally. One can guess that it is only a matter of time before these technologies give rise to litigation, and in particular, to questions whether XML schemas are themselves a type of copyrightable subject matter. Part III of this Article therefore explains what these terms and technologies mean.

Part IV first tries to answer the question whether XML schemas would be found copyrightable under the approaches traditionally used by courts (the dictionary definition approach, the merger approach, etc.). The rest of Part IV then applies the three-part analysis proposed in this Article and concludes that, although the analysis ultimately turns on factual findings that can only be made in actual litigation, it is very likely that those findings will support a conclusion that XML schemas are indeed copyrightable.

II. COPYRIGHT AND NEW SUBJECT MATTER: DIFFERENT APPROACHES

Subject matter questions have, in the past, often arisen in relation to a new technology. These cases tend to fall into two broad types, those in which the issue primarily has to do with a work's *medium of fixation*,⁵ and those in which the medium of fixation is unimportant and the issue is more whether the *type of work* is considered copyrightable.⁶

5. See, e.g., *Williams Elec., Inc. v. Artic Int'l, Inc.*, 685 F.2d 870, 873-74 (3d Cir. 1982) (rejecting an argument that the "attract" and "play" modes of a video game were not entitled to copyright protection because they allegedly failed to meet the "fixation" requirement of § 102(a) of the Copyright Act). Deciding that "[t]he fixation requirement is met whenever the work is 'sufficiently permanent or stable to permit it to be . . . reproduced, or otherwise communicated' for more than a transitory period," the *Williams* court found that the images in the videogame satisfied the "fixation" requirement, because "the original audiovisual features of the . . . game repeat themselves, over and over." *Id.* at 874.

6. See, e.g., *Walt Disney Prods. v. Air Pirates*, 581 F.2d 751, 754 (9th Cir. 1978) (holding that cartoon characters are copyrightable).

We tend to think of the former type as more common before the adoption of the 1976 Copyright Act. The 1909 Act and earlier versions seemed to put more focus on whether a particular medium of fixation was copyrightable; the subject matter of the Act was expressed in terms of tangible media like “books” or “prints” or “photo-engravings.”⁷ The 1976 Copyright Act, still in force, tried to turn the focus away from the media of fixation and onto the notion that a copyrightable “work” is really an abstract thing, a thing that can be fixed in any sort of medium, whether “now known or later developed.”⁸

Today, for example, we would have no trouble in saying that a photograph is a work of “visual art” more or less akin to a painting or a drawing. To be sure, the mechanics of creating a photograph are quite different from those used in other two-dimensional arts. But we would be comfortable in including an extensive range of mechanical technologies as all being a part of two-dimensional “art”—charcoal sketches, watercolors, oil paintings, tempera, pastels, engravings, frescos, and the like. “Photography” would not seem to modern-day copyright lawyers to be a strikingly different “art” from these, so much as it would seem to be an art that uses strikingly different tools in its creation. But in the mid-nineteenth century, when photography was invented,⁹ photography seemed very different indeed from the traditional media for two-dimensional representation. The copyright act of the day¹⁰ was likewise interpreted as being much more dependent on the medium of fixation than is true under today’s 1976 Act.¹¹ Consequently, whether a “photograph” could be copyrighted was a serious issue for the court that first entertained such a case,¹² and one that resulted in the answer “no”—photographs were not part of copyright’s subject matter¹³

7. Copyright Act of 1909, ch. 320, § 5, 35 Stat. 1075, 1076–77.

8. Copyright Act of 1976, 17 U.S.C. § 102(a) (1994 & Supp. 2000).

9. See BARBARA LONDON & JOHN UPTON, PHOTOGRAPHY 341–42 (Priscilla McGeehon ed., 6th ed. 1998).

10. Copyright Act of 1831, Ch. 16, 4 Stat. 436.

11. See, e.g., *White-Smith Music Publ’g Co. v. Apollo Co.*, 209 U.S. 1, 17 (1908) (deciding that a musical composition “is not susceptible of being copied until it has been put in a form which others can see and read,” and denying copyright protection to music recorded on perforated rolls). See also Trotter Hardy, *The Internet and the Law: Copyright and “New-Use” Technologies*, 23 NOVA L. REV. 657, 663–66 (1999) (reviewing the gradual abstraction of copyright’s subject matter away from the medium of fixation during the last two centuries).

12. See *Wood v. Abbott*, 30 F. Cas. 424 (C.C.S.D.N.Y. 1866) (No. 17,938). Refer to notes 27–38 *infra* and accompanying text (discussing *Wood* in greater detail).

13. *Wood*, 30 F. Cas. at 425 (deciding that photography did not qualify as a “print” within the meaning of the Copyright Act of 1831, and thus was not entitled to copyright protection under that Act).

(a decision rendered moot by a congressional amendment adding “photography” to copyright’s subject matter in 1865¹⁴).

In spite of the general success of the 1976 Act in generalizing copyright’s subject matter away from particular media, courts deciding cases under the current Act still often face subject matter questions in the wake of new technological developments.¹⁵ The question today is not so often about media of fixation, however, but more commonly about whether some new “thing” is a proper part of the abstract category of “works of authorship,” even if we do not care about the thing’s medium of fixation.

Notable cases like this in the past several decades have addressed questions about “input data formats,”¹⁶ “operating system programs” in ROM,¹⁷ “non-literal elements” of a computer program,¹⁸ “menu command hierarchies,”¹⁹ and case reporter “pagination.”²⁰ In addition, other cases less directly concerned with copyright’s subject matter can nevertheless be viewed as though they were subject matter cases. Among this line of cases are those involving software “emulators.”²¹

14. Act of March 3, 1865, ch. 126, 13 Stat. 540. See CONG. GLOBE, 38th Cong., 2nd Sess. 1337 (1865) (noting the passage of the Act).

15. See, e.g., *Whelan Assocs. v. Jaslow Dental Lab.*, 797 F.2d 1222, 1233–42, 1248 (3d Cir. 1986) (extending copyright protection beyond the literal code of computer programs, holding that it encompassed a program’s “structure, sequence and organization”); *Apple Computer Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1249–54 (3d Cir. 1983) (holding that computer programs expressed in object code or source code, which are embedded in a ROM chip are copyrightable, and that computer operating systems are not uncopyrightable per se).

16. *Synercom Tech., Inc. v. Univ. Computing Co.*, 462 F. Supp. 1003, 1011–14 (N.D. Tex. 1978) (deciding that input formats used with a computer program could be copyrightable “if the ideas they express are separable from their expression”).

17. *Apple*, 714 F.2d at 1249–54 (concluding that computer operating system programs are not per se precluded from copyright).

18. *Whelan Assocs.*, 797 F.2d at 1233–41 (extending copyright protection to “non-literal” elements of a computer program such as its structure, sequence, and organization).

19. *Lotus Dev. Corp. v. Borland Int’l, Inc.*, 49 F.3d 807, 815–18 (1st Cir. 1995) (holding that the “menu command hierarchy” employed in the Lotus 1-2-3 spreadsheet program constituted an “uncopyrightable method of operation”), *aff’d by an equally divided Court*, 516 U.S. 233 (1996).

20. *Matthew Bender & Co., Inc. v. West Publ’g Co.*, 158 F.3d 693, 698–700 (2d Cir. 1998) (deciding that the internal pagination of West Publishing Company’s case reporters were not protected by West’s compilation copyright). “Pagination” was not exactly a “high technology” question, but was interesting nonetheless.

21. See, e.g., *Sony Computer Entm’t, Inc. v. Connectix Corp.*, 203 F.3d 596, 601–08 (9th Cir. 2000) (reversing a preliminary injunction which barred Connectix from selling its Virtual Game Station—a Sony PlayStation emulator—on the grounds that Connectix’s reverse-engineering of Sony’s copyrighted basic input-output system (BIOS) constituted permissible fair use under the Copyright Act), *cert. denied*, 531 U.S. 871 (2000).

A. Rationales

These different subject matter cases often emphasize different rationales for their outcomes. Among the different rationales—not necessarily mutually exclusive—have been whether: (1) the new thing is not copyrightable by definition because it is a “system” or “method of operation”;²² (2) the defendant has a legitimate need to be compatible with the plaintiff’s product²³ (an argument often associated with a consideration of whether the plaintiff’s product has established such a “standard” in the industry that the public benefit from using the standard outweighs the plaintiff’s interest in ownership²⁴); (3) any “expression” in the new thing is essentially dictated by, and hence merged with, the thing’s underlying functional features;²⁵ and (4) America’s general policy favoring marketplace transactions over governmental allocation argues for the creation of some sort of legal property interest like that of copyright.

1. *Dictionary Definition Approach.* Some courts try only to see whether the Copyright Act’s definitions are met by the subject matter at issue.²⁶ Certainly, the first step to take in any question of statutory application is to see how the categories defined in the statute apply to the facts at hand. I refer here, however, to the fact that some courts—even in what they themselves consider to be hard

22. See, e.g., *Lotus*, 49 F.3d at 815 (holding a menu command hierarchy for computer spreadsheet program to be an uncopyrightable “method of operation”); *Brief English Systems, Inc. v. Owen*, 48 F.2d 555, 555 (2d Cir. 1931) (holding speedwriting system uncopyrightable).

23. See, e.g., *Lotus*, 49 F.3d at 820 (Boudin, J., concurring) (approving a competitor’s adoption of “the Lotus [spreadsheet] menu . . . as a fallback option for those users already committed to it by prior experience”).

24. See, e.g., *id.* at 821 (Boudin, J., concurring) (noting that the Lotus 1-2-3 spreadsheet at one time “represented the *de facto* standard for electronic spreadsheet commands,” but concluding that this fact did not justify keeping “customers who have learned the Lotus menu . . . captives of Lotus” if “a better spreadsheet comes along”).

25. See, e.g., *The Yankee Candle Co. v. Bridgewater Candle Co.*, 2001 U.S. App. LEXIS 17389, *15–18 (1st Cir. 2001) (holding that a scented candle manufacturer was not entitled to copyright protection on the subject matter of the photographs printed on its candle labels because of the merger doctrine); *Kepner-Tregoe, Inc. v. Leadership Software, Inc.*, 12 F.3d 527, 533 (5th Cir. 1994) (“[W]hen an idea can be expressed in very few ways, copyright law does not protect that expression, because doing so would confer a *de facto* monopoly over the idea. In such cases idea and expression are said to be merged.”).

26. See, e.g., *Verney Corp. v. Rose Fabric Converters Corp.*, 87 F. Supp. 802, 804 (S.D.N.Y. 1949) (analyzing whether a fabric design fits into the statutory definition of print or label); *Wood v. Abbott*, 30 F. Cas. 424, 425 (C.C.S.D.N.Y. 1866) (No. 17,938) (holding that photographs did not fit within statutory list of copyrightable works because photographs did not exist at the time the Copyright Act of 1831 was passed).

cases—do not undertake any significant policy analysis or discussion. They rely primarily on a “literal minded” dictionary definition approach without much discussion of why Congress might have set up the definitional categories it did or what purposes it might have intended those categories to achieve.

Wood v. Abbott,²⁷ a district court case from 1866, is a good example. In *Wood*, the essential question was whether a “photograph” was copyrightable.²⁸ Photography at that time was still quite a new technology,²⁹ and the issue was one of first impression.³⁰ The statute at that time provided that: “any person . . . who shall . . . invent, design, etch, engrave, work . . . from his own design, any print or engraving . . . shall have the sole right . . . of printing, reprinting, publishing and vending such . . . print, cut, or engraving, in whole or in part, for the term of twenty-eight years”³¹

As the court defined it, the issue was “whether or not these photographs are prints, cuts or engravings, and, therefore, protected by the statute,”³² with a particular focus on whether photographs fit the statutory definition of “prints.”³³

The analysis of this question was based on asking what Congress meant in the literal sense by the word “print.”³⁴ The court concluded that Congress must have meant “print” in a non-photographic sense, because the art of photography had not been invented in 1831 when the then-current statute³⁵ had been enacted:

[Photography] is a new and beautiful art, discovered long after the statute in question was enacted. It is not a development of the art of making prints or engravings which existed at the date of the act. Then, a print was defined to be “a mark or form made by impression or printing; anything printed; that which, being impressed, leaves its form, as a butter print, a cut in wood or metal to be impressed on paper; the impression made; a picture; a stamp; the letters in a printed book,” &c. As used in this statute, it was synonymous with the term “engraving,” with

27. *Wood*, 30 F. Cas. at 424.

28. *See id.* at 424–25.

29. *See* LONDON & UPTON, *supra* note 9, at 342 (noting that the “first public announcement of the invention of a successful photographic process” occurred in 1839).

30. *See Wood*, 30 F. Cas. at 424 (“The case presented . . . is a novel one, at least in this court . . .”).

31. Act of Feb. 3, 1831, ch. 16, 4 Stat. 436 (1831).

32. *Wood*, 30 F. Cas. at 425.

33. *Id.*

34. *Id.*

35. Act of Feb. 3, 1831, ch. 16, 4 Stat. 436 (1831).

which it is connected in the act, which means, in this relation, “an engraved plate; an impression from an engraved plate.” This new art of photography, and all its kindred processes, is an entirely original and independent mode of taking pictures of material objects, and multiplying copies of such pictures at pleasure. That combination of creative or imitative power and mechanical skill by which the artist works out his own conception, or the embodied conception of another, in a fixed form, the fruits of which the law was intended to protect, is not brought into play. No block, plate, or stone is engraved. No figure is drawn, etched, raised, or worked on any surface from which copies are to be produced by impression or printed.³⁶

In short, if photography did not exist at the time Congress enacted the statute, then the statute could not apply to photography.³⁷ And indeed, that is what the court ultimately held.³⁸

Justice Holmes’s opinion in the famous *Bleistein v. Donaldson Lithographing*³⁹ case briefly invoked a dictionary definition approach—albeit in connection with a much more significant policy argument. At issue was whether a commercial advertising poster for a circus was copyrightable subject matter.⁴⁰ At the time, there was a strong current of thought that copyright law was only for the protection of “art” in the sense of “fine art” like painting and drawing, music, drama, and so on.⁴¹ The Court rejected that view, concluding that copyright was intended for any sort of print or drawing, whether created for pure intellectual contemplation or to promote sales.⁴² The policy argument rested on the Court’s reluctance to have judges engage in the assessment of artistic quality to determine that “this” work

36. *Wood*, 30 F. Cas. at 425.

37. *See id.*

38. *Id.* at 425 (holding that Congress could not have intended to include photography as copyrightable work under the Act of February 3, 1831, because photography had not yet been invented at the time of the Act’s passage). In fact, Congress had expressly added “photography” to the statute’s subject matter retroactively, so the court itself understood the limited role its decision would have in the scheme of things. *See id.*

39. *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239 (1903).

40. *Id.* at 248.

41. *See, e.g., J.L. Mott Iron Works v. Clow*, 82 F. 316, 318 (7th Cir. 1897) (refusing to extend copyright protection to drawings in an appliance catalogue, finding that the subjects of the drawings—a washbowl, sink, sponge holder, brush holder, and a robe hook—lacked ornamentation, and could not have been “the subject of artistic treatment”).

42. *See Bleistein*, 188 U.S. at 251 (“A picture is . . . none the less a subject of copyright, that it is used for an advertisement.”).

was “fine art,” but “that” work was something lesser.⁴³ But, along with that policy argument, Holmes mentioned that the wording of the statute did not seem confined to “fine art” in any event, despite the statute’s use of that term for another purpose.⁴⁴

One recent court asked whether the proposed new subject matter in issue was a “system” or “method of operation,” a particular instance of this same “dictionary definition” approach.⁴⁵ In section 102(b), the Copyright Act says the following:

In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.⁴⁶

The majority’s opinion in *Lotus v. Borland* approached the case in that fashion. The Lotus Corporation had developed the widely used spreadsheet software known as “Lotus 1-2-3,” which is:

[A] spreadsheet program that enables users to perform accounting functions electronically on a computer. Users manipulate and control the program via a series of menu commands, such as “Copy,” “Print,” and “Quit.” Users choose commands either by highlighting them on the screen or by typing their first letter. In all, Lotus 1-2-3 has 469 commands arranged into more than 50 menus and submenus.⁴⁷

A Competing software company, Borland, developed its own spreadsheet program, “Quattro,” which could make use of the same menu commands—indeed, Quattro implemented “a

43. See *id.* at 251–52 (stating that the practice of persons trained only in the law to become judges of artistic quality would be “a dangerous undertaking”).

44. Justice Holmes stated:

We assume that the construction of Rev. Stat. § 4952 . . . allowing a copyright to the “author, designer, or proprietor . . . of any engraving, cut, print . . . [or] chromo” is affected by the act of 1874 That section provides that, “in the construction of this act, the words ‘engraving,’ ‘cut,’ and ‘print’ shall be applied only to pictorial illustrations or works connected with the fine arts.” We see no reason for taking the words “connected with the fine arts” as qualifying anything except the word “works,” but it would not change our decision if we should assume further that they also qualified “pictorial illustrations,” as the defendant contends.

Id. at 250.

45. *Lotus Dev. Corp. v. Borland Int’l, Inc.*, 49 F.3d 807, 815 (1st Cir. 1995) (finding the Lotus menu hierarchy to be an uncopyrightable method of operation).

46. 17 U.S.C. § 102(b) (1994) (emphasis added).

47. *Lotus*, 49 F.3d at 809.

virtually identical copy of the 1-2-3 menu tree,"⁴⁸ though with a different on-screen appearance.⁴⁹ Lotus sued Borland, arguing that Borland had infringed Lotus's copyright in the hierarchy of menu commands.⁵⁰ The trial court agreed with Lotus.⁵¹ On appeal, the First Circuit disagreed and concluded that the menu hierarchy was not copyrightable.⁵² The case went to the Supreme Court, which affirmed the First Circuit on a split four-to-four vote, with one justice recusing himself.⁵³

The First Circuit's opinion therefore stands as the authoritative word. The court analyzed the issue by asking whether the hierarchy of commands was a "system" or "method of operation,"⁵⁴ neither of which, by the Act's own definitions, is subject to copyright protection.⁵⁵ The court concluded that the hierarchy was in fact a "method of operation," and hence not copyrightable:

We hold that the Lotus menu command hierarchy is an uncopyrightable "method of operation." The Lotus menu command hierarchy provides the means by which users control and operate Lotus 1-2-3 Users must use the command terms to tell the computer what to do. Without the menu command hierarchy, users would not be able to access and control, or indeed make use of, Lotus 1-2-3's functional capabilities.

The Lotus menu command hierarchy does not merely explain and present Lotus 1-2-3's functional capabilities to the user; it also serves as the method by which the program is operated and controlled.⁵⁶

Having defined the menu hierarchy as a "method of operation," the court was then obliged to find it uncopyrightable by definition.⁵⁷

48. Lotus Dev. Corp. v. Borland Int'l, Inc., 831 F. Supp. 202, 212 (D. Mass. 1993).

49. *Id.*

50. *Id.* at 209.

51. Lotus, 49 F.3d at 809 (summarizing the ultimate outcome of the four previous proceedings before the trial court).

52. *Id.* at 815; refer to notes 54–57 *infra* and accompanying text (discussing the court's reasoning).

53. Lotus Dev. Corp. v. Borland Int'l, Inc., 516 U.S. 223, 233 (1996). Justice Stevens took no part in the decision. *Id.*

54. See Lotus, 49 F.3d at 815–16.

55. 17 U.S.C. § 102 (1994) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery . . .").

56. Lotus, 49 F.3d at 815.

57. See *id.* at 815–16.

Dictionary definitions can, of course, be used flexibly. In one of the first cases dealing with the copyrightability of a motion picture, the court applied a copyright statute that had included “photography” as copyrightable subject matter for years, but did not yet mention “motion pictures.”⁵⁸ The court found that a reel of motion picture film was essentially a very long photograph and hence was copyrightable under that heading.⁵⁹

2. *Standards and Compatibility Approach.* Courts finding a new work not copyrightable often support their conclusions by asserting that copyright protection would unduly hinder competition.⁶⁰ This general statement is bolstered either by a reference to the need for competitors to be able to adopt widely used, de facto standards, or by a reference to the advantages to customers of using a new product that is “compatible with” previous products and that accordingly does not require retraining, reprogramming, or new data preparation techniques.

The First Circuit pointed to these factors in support of its conclusion in *Lotus*, looking favorably on the value of users being able to apply their skills and familiarity with the Lotus 1-2-3 commands to other programs like Borland’s Quattro:

That the Lotus menu command hierarchy is a “method of operation” becomes clearer *when one considers program compatibility*. Under Lotus’s theory, if a user uses several different programs, he or she must learn how to perform the same operation in a different way for each program used. For example, if the user wanted the computer to print material, then the user would have to learn not just one method of operating the computer such that it prints, but many different methods. We find this absurd.⁶¹

In a lengthy concurring opinion, Judge Boudin similarly referred to the role of software standards:

58. *Edison v. Lubin*, 122 F. 240, 241 (3d Cir. 1903) (applying Act of March 3, 1865, ch. 126, 13 Stat. 540 (1865)).

59. *Id.* at 242 (“[T]hat [the reel] is, in substance, a single photograph, is shown by the fact that its value consists in its protection as a whole or unit, and the injury to copyright protection consists not in pirating one picture, but in appropriating it in its entirety.”).

60. *See, e.g., Kern River Gas Transmission Co. v. Coastal Corp.*, 899 F.2d 1458, 1464 (5th Cir. 1990) (“The . . . maps consisted of lines representing the proposed location of the pipeline To extend protection to the lines would be to grant Kern River a monopoly of the idea for locating a proposed pipeline in the chosen corridor, a foreclosure of competition that Congress could not have intended to sanction through copyright law”).

61. *Lotus*, 49 F.3d at 817–18 (emphasis added).

Apparently, for a period Lotus 1-2-3 has had such sway in the market that it has represented the *de facto standard* for electronic spreadsheet commands. So long as Lotus is the superior spreadsheet—either in quality or in price—there may be nothing wrong with this advantage. But if a better spreadsheet comes along, it is hard to see why customers who have learned the Lotus menu and devised macros for it should remain captives of Lotus because of an investment in learning made by the users and not by Lotus.⁶²

A similar point was made in *Synercom Technology, Inc. v. University Computing Company*.⁶³ In *Synercom*, two companies (among several others) competed in the market for engineering software to run on main-frame computers.⁶⁴ The data necessary for the software's calculations had to be entered in the form of "punch cards," sometimes known as "IBM cards" for their original close association with that company.⁶⁵ Quite a bit of data had to be punched onto these cards, and it all had to be put in the correct columns for the computer to understand which numbers were to indicate which type of data.⁶⁶ For example, columns one through six might represent "width," while columns seven through twelve might represent "height," and so on.⁶⁷

One of Synercom's competitors, plaintiff EDI, determined that many customers already would have key-punched a lot of their data into the format required by Synercom's program, and were necessarily familiar with key-punching new input data in the same format.⁶⁸ So EDI produced an engineering program that used the same punch-card format as Synercom's program.⁶⁹ Synercom sued EDI for copyright infringement and lost.⁷⁰

In the process of concluding that the plaintiff had no copyright interest in its input formats, the court compared the input formats at issue to the "H" pattern of an automobile gearshift.⁷¹ With gearshift patterns, the court noted that the "use

62. *Id.* at 821 (Boudin, J., concurring) (emphasis added).

63. 462 F. Supp. 1003 (N.D. Tex. 1978).

64. *Id.* at 1006.

65. See generally EMERSON W. PUGH, BUILDING IBM: SHAPING AN INDUSTRY AND ITS TECHNOLOGY 48-52 (1995) (explaining IBM's development of a punch card with capacity for holding significantly more data than any other card).

66. See *id.*

67. I use "width," etc., as examples without in fact knowing what sorts of data were actually entered into the software; the software was designed to perform "structural analysis."

68. *Synercom*, 462 F. Supp. at 1008.

69. *Id.* at 1009.

70. *Id.* at 1011 (denying Synercom copyright protection for its input formats).

71. *Id.* at 1013 (offering the analogy to "illuminate the idea versus expression

of the same pattern [by car manufacturers other than whichever one originally came up with the pattern] might be socially desirable, as it would *reduce the retraining of drivers*.⁷²

Somewhat in the same vein was *Bender v. West*,⁷³ a declaratory judgment action in which the plaintiff had produced a CD-ROM of court opinions that included references to the pagination of each such case as it appeared in the defendant's West Reporter series.⁷⁴ The court found that West did not have a copyright in the pagination of its Reporters, observing at one point that "the primary use of West's pagination in plaintiffs' products is to allow the user to refer to the location of a particular text within the West case reporters *as has become standard practice* in the legal community."⁷⁵ Although it is unlikely that the standardization issue alone heavily influenced the court,⁷⁶ it is also unlikely that the court would have mentioned it if it had played no role in the court's thinking.

3. *Merger Approach*. Many cases of new technologies and copyright discuss the concept of "merger."⁷⁷ In copyright law, ideas are not copyrightable, but the particular expression of an idea is.⁷⁸ For example, an "idea" in literature might be that two neighborhood groups or family clans are violently opposed to each other, that a young man from one of these groups and a young woman from the other fall in love, that both groups are appalled at this prospect, and that the groups can only become reconciled to this relationship and to each other upon the death of the two lovers. That idea could be expressed in the form of the play *Romeo and Juliet*, or in the musical *West Side Story*. Both plays use the same "idea," but reflect quite different "expressions" of that idea.⁷⁹ In such a case, both plays

controversy," which refers to the fact that Synercom's formats would be "copyrightable if the ideas they express are separable from their expression").

72. *Id.* (emphasis added).

73. *Matthew Bender & Co., Inc. v. West Publishing Co.*, 158 F.3d 693 (2d Cir. 1998).

74. *Id.* at 697.

75. *Id.* at 706.

76. The court based its decision on the lack of copyrightable originality in the West reporters' pagination. *Id.* at 699 ("Because the internal pagination of West's case reporters does not entail even a modicum of creativity, the volume and page numbers are not original components of West's compilations and are not themselves protected by West's compilation copyright.").

77. See, e.g., *Kregos v. Associated Press*, 937 F.2d 700, 705 (2d Cir. 1991) (applying the merger doctrine in determining the copyrightability of baseball pitching forms); see also *CCC Info. Servs., Inc. v. Maclean Hunter Mkt. Reports*, 44 F.3d 61, 72 (2d Cir. 1994) (stating that the trial court was incorrect in applying the merger doctrine in a case concerning whether copyright protection of a book listing used car valuations prohibits use of data by another for a computer database).

78. *Mazer v. Stein*, 347 U.S. 201, 217 (1954).

79. See 1 HOWARD B. ABRAMS, *THE LAW OF COPYRIGHT* § 1.02(D)(2)(a) (2000)

can be copyrighted (assuming that *Romeo and Juliet* were written today).⁸⁰

At times, though, the expression of an idea is almost identical to the underlying idea. For example, an author might develop the idea for a particular type of double-entry bookkeeping. One way in which that “idea” can be “expressed” is in the form of ruled lines and columns on a piece of paper—the same columns and lines that constitute the idea for this method of doing accounting. This very issue arose in the Supreme Court’s *Baker v. Selden*⁸¹ case of 1879. Selden had published a book describing his new accounting method, and the book included pages of blank forms with the necessary ruled lines and columns.⁸² Baker reproduced some of the forms, although not the book’s textual description of the accounting method.⁸³ Selden sued for copyright infringement.⁸⁴

The Court was faced with either protecting Selden’s “expression” comprising the blank forms with ruled lines, and thereby simultaneously protecting the underlying accounting idea, or denying protection to the underlying idea, and simultaneously denying protection to the expression.⁸⁵ Given that choice, the Court opted for denying protection to both, rather than giving copyright protection to Selden’s underlying idea for a method of accounting.⁸⁶ The case is now described as exemplifying “merger”—the merger of underlying idea and expression.⁸⁷

The *Synercom* case mentioned earlier ultimately turned on the court’s view that a similar “merger” had taken place with the subject matter at issue. The court concluded that the input formats were inseparable from—that is, “merged with”—any possible underlying idea.⁸⁸ As the court put it, “[i]f sequencing

(explaining the distinction between “idea” and “expression” in the context of copyright protection).

80. See 17 U.S.C. § 102 (1994) (allowing copyright of “expressions” but not of “ideas”).

81. 101 U.S. 99 (1879).

82. *Id.* at 100.

83. *Id.* at 100–01.

84. *Id.* at 100.

85. *Id.* at 101.

86. See *id.* at 107 (“[T]he mere copyright of Selden’s book did not confer upon him the exclusive right to make and use account-books, ruled and arranged as designated by him and described and illustrated in said book.”).

87. See Dale P. Olson, *The Uneasy Legacy of Baker v. Selden*, 43 S.D. L. REV. 604, 604 (1998) (“*Baker v. Selden* remains an enduring authority for its pre-eminent role in arbitrating between protectable expression and unprotectable ideas in copyrighted works . . .”).

88. See *Synercom Tech., Inc. v. Univ. Computing Co.*, 462 F. Supp. 1003, 1014 (N.D.

and ordering [of the data input fields] is expression, what separable idea is expressed?"⁸⁹ The implicit answer was "none."⁹⁰

The defendant in *Apple v. Franklin*⁹¹ similarly argued that the Apple computer's operating system programs had merged with their underlying ideas, or that the idea of producing a computer compatible with the Apple computer permitted only one type of expression.⁹² The Third Circuit rejected that argument on the grounds that the idea of "compatibility" was not the idea that needed to be examined for purposes of "merger."⁹³ Rather, it was the idea of performing a particular operating system function like reading and processing a particular programming language,⁹⁴ for which many different expressions were possible. In any event, the court did consider merger as a doctrine of relevance to the copyright-and-new-technology case before it, although one of the requirements of merger was not satisfied by the facts at hand.⁹⁵

Another computer program case, *Whelan v. Jaslow*,⁹⁶ helps to clarify the concept of merger by illustrating its opposite. At issue were two computer programs with the same overall purpose—management of a dental laboratory.⁹⁷ One author sued

Tex. 1978) ("Here if order and sequence is the expression, the skilled effort is not separable, for the form, arrangement, and combination is itself the intellectual conception involved.").

89. *Id.* at 1013.

90. The court stopped just short of explicitly reaching that finding. Instead, it said in one paragraph that a gear shift pattern cannot be copyrighted (even though descriptions or photographs of it might be), and in the very next that "Synercom's argument that the order and sequence of data was the expression, not the idea, has been rejected." *Id.*

91. *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240 (3d Cir. 1983).

92. *Id.* at 1252–53 (addressing defendant Franklin's argument that there were a "limited number of ways to arrange operating systems to enable a computer to run the vast body of Apple-compatible software" and holding that computer operating system programs written in object code and embedded on ROM are copyrightable).

93. *Id.* at 1253 (noting that achieving total compatibility with Apple programs is a mere "commercial and competitive objective," not a factor for determining whether a merger has occurred).

94. *Id.* ("The idea which may merge with the expression . . . is the idea which is the subject of the expression. The idea of one of the operating system programs is, for example, how to translate source code into object code. If other methods of expressing that idea are not foreclosed as a practical matter, then there is no merger.").

95. *Id.* (analyzing the merger issue, but refusing to decide it because "the district court made no findings as to whether some or all of Apple's operating programs represent the only means of expression of the idea underlying them").

96. *Whelan Assocs., Inc. v. Jaslow Dental Lab., Inc.*, 797 F.2d 1222 (3d Cir. 1986).

97. *Id.* at 1224–25. Elaine Whelan, who later established Whelan Associates, developed the Dentalab program to manage Jaslow Dental Laboratory. *Id.* at 1225–26. Later, Rand Jaslow, an officer of Jaslow Dental, and others formed Dentcom to market their re-write of the Dentalab program. *Id.* at 1226–27.

the other for copyright infringement.⁹⁸ The second program was not a direct copy of the first, but was in some sense a “re-write” of the first in a different computer programming language.⁹⁹ The court defined the “idea” of the programs in a broad and general enough way, such that there was no question of “merger” whatsoever. “[T]he purpose or function of a utilitarian work would be the work’s idea, and everything that is not necessary to that purpose or function would be part of the expression of the idea.”¹⁰⁰ In other words, the idea was “managing a dental lab,” which is something that can obviously be expressed in a great many different ways, including manual procedures as well as computer programs.¹⁰¹ With that broad and abstract a definition of “idea,” there was no question of any merger having taken place.¹⁰²

4. *Policy-Oriented Approach.* An intriguing case from 1902 illustrates the application of a pure policy approach.¹⁰³ The facts were quite similar to the Supreme Court’s later and better known *INS* case.¹⁰⁴ The plaintiff Western Union had initiated a kind of news service, distributed by means of telegraph wires.¹⁰⁵ The news was sent over the wires to be printed out by automatic machines known as “tickers” on long strips of paper.¹⁰⁶ (Although the court does not use the term “ticker tape,” one presumes that this technology was indeed the origin of that term.) These printed news tapes were available to the public in most places where Western Union maintained a telegraph office—which was nearly everywhere.¹⁰⁷ The defendant operated a similar news ticker service, but apparently saved itself the time and trouble of writing its own news by routinely gathering that news from

98. *Id.* at 1227.

99. *Id.* at 1226. The original program was written in EDL, while the infringing program was written in BASIC. *Id.*

100. *Id.* at 1236 (emphasis in original).

101. *Id.* at 1236 n.28.

102. *Id.* at 1239 (concluding that the “detailed structure of the Dentalab program is part of the expression, not the idea, of that program”).

103. *Nat’l Tel. News Co. v. Western Union Tel. Co.*, 119 F. 294 (7th Cir. 1902).

104. *Int’l News Serv. v. Associated Press*, 248 U.S. 215 (1918). International News Service, a news collection business and a competitor of Associated Press, copied “news from [Associated Press] bulletin boards and from early editions of [Associated Press] newspapers.” *Id.* at 231. International News Service then sold the copied news as its own. *Id.* An injunction was issued against the “bodily taking of the words or substance of [Associated Press]’s news until its commercial value as news had passed away.” *Id.* at 232.

105. *Western Union*, 119 F. at 295.

106. *Id.*

107. *Id.* (“The appellee, the Western Union Telegraph Company, does a general telegraphing business, having offices in every state, village, hamlet and railroad station in the country, and wires connecting the same with central offices through the country.”).

Western Union's tickers and redistributing it to its own customers.¹⁰⁸ Plaintiff sued defendant, at least for copyright infringement and perhaps also for common law remedies not clearly specified in the opinion.¹⁰⁹

The court began its analysis by observing that copyright protection had grown more expansive over the years:

. . . so that it now includes books that the old guild of authors would have disdained; catalogues, mathematical tables, statistics, designs, guide-books, directories, and other works of similar character. Nothing, it would seem, evincing, in its makeup, that there has been underneath it, in some substantial way, the mind of a creator or originator, is now excluded.¹¹⁰

And yet, surprisingly with such a prelude, the court continued to say that that copyright protection did have limits—it applied only to the product of authorship and did *not* extend to a plain recitation of facts such as appeared on the news ticker tapes in the case at hand.¹¹¹ The plaintiff could not, therefore, maintain its action for copyright infringement.¹¹²

108. *Id.* (noting that the National Telegraph News Company's copying of Western Union's news was not disputed).

109. *Id.* at 296. The court framed the issue as follows:

The general question raised by [National Telegraph News Company's] contention, then, is this: Is the printed tape, coming out of [Western Union's] tickers, a book or article within the meaning of the copyright laws of the United States, and especially of section 4956 (U.S. Comp. St. 1901, p. 3407), and if not a book or article within the meaning of the copyright law, is there any remedy that will protect this feature of [Western Union's] business against the kind of piracy shown?

Id.

110. *Id.* at 297.

111. *Id.* at 297–98.

[W]e may fix the confines at the point where authorship proper ends, and mere annals begin. Nor is this line easily drawn. Generally speaking, authorship implies that there has been put into the production something meritorious from the author's own mind; that the product embodies the thought of the author, as well as the thought of others; and would not have found existence in the form presented, but for the distinctive individuality of mind from which it sprang. A mere annal, on the contrary, is the reduction to copy of an event that others, in a like situation, would have observed; and its statement in the substantial form that people generally would have adopted. A catalogue, or a table of statistics, or business publications generally, may thus belong to either one or the other of these classes. If, in their makeup, there is evinced some peculiar mental endowment—the grasp of mind, say in a table of statistics, that can gather in all that is needful, the discrimination that adjusts their proportions—there may be authorship within the meaning of the copyright grant as interpreted by the courts. But if, on the contrary, such writings are a mere notation of the figures at which stocks or cereals have sold, or of the result of a horse race, or baseball game, they cannot be said to bear the impress of individuality, and fail, therefore, to rise to the plane of authorship. In authorship, the product has some likeness to the mind underneath it; in a work of mere notation, the mind is guide only to the fingers that make the notation. One is the product of originality; the other

Nevertheless, the court thought it obvious that without legal protection for its efforts, the plaintiff's business could be destroyed.¹¹³ On that basis, the court found as a matter of equity that property rights protection should extend to the plaintiff's news service and that the defendant should therefore be enjoined.¹¹⁴ In short, the court determined to provide copyright-like protection—albeit of a very short duration¹¹⁵—for underlying policy reasons even where copyright law did not apply!¹¹⁶

B. Unnecessary Issues

Many copyright opinions mention or even rely on factors that either are not, strictly speaking, necessary for a decision about copyrightable subject matter, or are simplifications that could possibly be misleading. My analytical framework does not refer to these factors, so I feel some need to explain why I think that they are of less relevance than may first be imagined.

1. *The “Need to Compete.”* Courts in intellectual property cases sometimes refer to the necessity of competition. A variety of key phrases come to mind in this connection, such as: “the need to compete,”¹¹⁷ that “competitors must be free to enter the market,”¹¹⁸

the product of opportunity.

112. *Id.* at 298. The court determined the news to be “an exchange merely over wider area, of ordinary sightseeing,” and that, therefore, it was not “the fruit of intellectual labor, and would not, if actually copyrighted, be protected by the courts.” *Id.*

113. *Id.* at 296 (recognizing the cost National Telegraph News Company would avoid if it could appropriate Western Union's news lawfully).

114. *Id.* at 299–301 (finding that “though the immediate thing to be acted upon by the injunction is not itself, alone considered property, it is enough that the act complained of will result, even though somewhat remotely, in injury to property”).

115. The court affirmed the district court's grant of protection for a total of sixty minutes for each news item. *Id.* at 299, 301. After that time, the defendant was free to reproduce the news. *Id.*

116. *Id.* at 300–01 (“Is the enterprise of the great news agencies . . . to be denied appeal to the courts . . . for no other reason than that the law, fashioned hither to fit the relations of authors and the public, cannot be made to fit[?] . . . We choose, rather, to make precedent—one from which is eliminated, as immaterial, the law grown up around authorship—and we see no better way to start this precedent . . . than by affirming the order appealed from.”).

117. *See, e.g.,* *Intergraph Corp. v. Intel Corp.*, 3 F. Supp. 2d 1255, 1278–79 (N.D. Ala. 1998) (“Intel is unlawfully using its monopoly in the high-performance CPU relevant market to foreclose or restrain competition by Integraph in the graphic subsystem relevant market. Intel has already entered that market and has clearly announced plans to expand in that market while at the same time denying Intergraph access to the CPUs and technical information it needs to compete. This will unnecessarily ‘handicap’ or restrain Intergraph's ability to compete in that market.”), *vacated*, 195 F.3d 1346 (Fed. Cir. 1999).

118. *See, e.g.,* *Lotus Dev. Corp. v. Borland Int'l, Inc.*, 49 F.3d 807, 821 (1st Cir. 1995) (questioning why Borland should not be free to attract Lotus customers if the Borland

or that a company cannot “effectively compete” if it is not allowed to make use of some claimed form of intellectual property.

The trouble with these formulations is that they imply that “competition” (or the “ability to compete,” or the “ability to compete effectively”) is a binary quality—one is led to think that a company either “can” or “cannot” engage in the described behavior.¹¹⁹ But surely “competition” is a matter of degree, something that falls on a continuum. It would be perfectly sensible to refer, for example, to the fact that some company can “better compete” if certain conditions are met, such as that their product take on some feature its competitor already possesses, or to note that the price of a product might be lowered by ten percent to provide more effective competition to a rival. Indeed, the fact that products are sometimes put on sale implies that competition must be a matter of degree; otherwise, a given price would either “compete” or “not compete,” and there would be little point to temporarily lowering a price as a competitive tactic.

Moreover, “competition” is always possible on *some* market level. Under no conceivable circumstances would a software company like Lotus be able to stop Borland from competing in the business world by selling running shoes. Because “money” is fungible, in some sense, *all* sellers of any products are competing for buyers’ dollars. On the other hand, competition is always impossible on some other level. For example, Lotus could certainly have stopped Borland from outright duplication and sale of the Lotus 1-2-3 software installation disks.¹²⁰ Despite what courts say, the question is almost never whether we want “competition or not.” Stated that baldly, the answer must be “yes, of course we want competition.” The questions are rather: “what sort of competition,” “how much of it,” and “at what market level?” These are questions that this Article answers with its approach to subject matter questions based on copyright’s policy of incentives.

2. *Standards and Compatibility.* How important is it in a copyrightable subject matter case to take account of “standards” and

product is superior).

119. One might think initially that the ability to “compete effectively” relies on the notion of “effectiveness,” and that “effectiveness” in turn implies varying degrees of effectiveness and hence that competition is a matter of degree. But in fact, any reference to “effective competition” is functionally equivalent to a reference to “competition”: if one cannot compete “effectively,” then one cannot compete, period. No company would bother to enter a market in which it could “compete,” but only “*ineffectively*.”

120. See *Lotus*, 49 F.3d at 816 (“It is now well settled that the literal elements of computer programs, i.e., their source and object codes are the subject of copyright protection.”) (quoting *Computer Assocs. Int’l v. Altai, Inc.*, 982 F.2d 693 (2d Cir. 1992)).

“compatibility”? Many companies would like to be able to have their products become so entrenched with the consuming public that they become standards. Apart from whatever penalties antitrust law might impose, should these companies pay a price for their success in the form of a loss of any copyrights they might have enjoyed in the absence of that success?

The issue is sometimes thought to be captured in the notion of “timing.”¹²¹ When a newcomer to a field first develops a work, as when Lotus first developed the user interface for its spreadsheet, the range of possible expressive choices is very wide. As the district court in that case observed, any number of different words could be used for the individual spreadsheet commands, and they could be arranged into menus in a huge variety of ways.¹²² The court further noted this was also true for the second comer, Borland, if compatibility with Lotus 1-2-3 were not a goal:

A very satisfactory spreadsheet menu tree can be constructed using different commands and a different command structure from those of Lotus 1-2-3. In fact, Borland has constructed just such an alternate tree for use in Quattro Pro’s native mode. Even if one holds the arrangement of menu commands constant, it is possible to generate literally millions of satisfactory menu trees by varying the menu commands employed.¹²³

But when the second comer like Borland enters the field with the express purpose of taking customers away from the field’s leader like Lotus, the second comer can much more easily do so if it can make use of the same set of menu commands originated by the leader. In that light, the range of expressive choices open to Borland is much narrower at this later time than it was for Lotus at an earlier time. Similar observations could be made about the *Synercom* case.¹²⁴ The second comer had determined to take customers away from the leading producer of software. At the time of that decision, doing so was far easier if the second comer could make use of the leading company’s input

121. See II PAUL GOLDSTEIN, COPYRIGHT, § 8.5.1.2 nn.56 & 60 (2d ed. 1998).

122. Lotus Dev. Corp. v. Borland Int’l, Inc., 799 F. Supp. 203, 217–18 (D. Mass. 1992) (noting that using synonyms for menu command names results in many combinations of commands in the main menu and that, because changes in submenus “increase the number of possible menu hierarchies geometrically,” the final number of possible menu hierarchies is “extremely large”), *reversed*, 49 F.3d 807 (1st Cir. 1995).

123. *Id.* at 217.

124. Synercom Tech., Inc. v. Univ. Computing Co., 462 F. Supp. 1003 (N.D. Tex. 1978).

data formats.¹²⁵ Hence, goes the argument, the range of expressive choices for those formats was far less at this later time.

This characterization of the issue as one of timing, however, tends to obscure a crucial predicate—the decision by the second comer to capture customers from the leader. Rather than observing that a second comer who wants to take away the leader’s customers has a limited range of expressive choices, we would do better to ask: to what extent should copyright law be interpreted to defer to a second comer’s desire to take customers away from the leader? A second comer may, after all, enter a field without necessarily trying to capture existing customers from the leader. In the *Synercom* case, for example, at the time the second-comer decided to use the leader’s formats, the market for engineering calculations of the sort at issue already included some fifteen competitors.¹²⁶ Other than the defendant, *none* of these other competitors was using the same input data format that the leader used.¹²⁷ If a market is large enough to feature many different customers using many different input formats (or user interfaces, or whatever is at issue), then denying copyright protection for any one of those formats will actually reduce competition over those formats, as competitors will now lose an incentive to develop new ones.

3. *Official Standards.* The standards issue also arises in a different, but related, copyright context. Sometimes a legislature will adopt a building, electrical, or other code as law.¹²⁸ The code was written by a private organization, which claims copyright in the document.¹²⁹ These kinds of cases do not raise questions concerning a new type of work of authorship and its potential copyrightability—a building code is a garden-variety literary work that plainly can be original and expressive enough to be copyrighted.¹³⁰ But they do

125. *Id.* at 1008–09 (recognizing that Synercom spent nearly \$500,000 to train customers to use its software, and that EDI, the second comer, could “simply pluck the fruit of Synercom’s labors and risks”).

126. *Id.* at 1007.

127. *Id.* (“There are hundreds of programs available for structural analysis, and at least fifteen are competitive with [plaintiff’s product] STRAN. All but [defendant] EDI and [plaintiff] Synercom have different input formats.”)

128. *See, e.g.*, 63 Fed. Reg. 8454, 8454–55 (Feb. 19, 1998) (Office of Management and Budget Notice of Final Revision of Circular A-119) (directing federal agencies to adopt privately developed standards “whenever practicable and appropriate” to “eliminate . . . the cost to the Government of developing its own standards”).

129. *See, e.g.*, *Veeck v. Southern Bldg. Code Cong. Int’l Inc.*, 241 F.3d 398, 403 (5th Cir. 2001) (finding that adoption of a model building code written by a non-profit organization did not invalidate the organization’s copyright).

130. *Id.* at 402–03 (“To establish copyright infringement, the plaintiff must prove a

raise the question whether adoption as a standard in some sense “revokes” what would otherwise be copyrightable about the document adopted.¹³¹ The straightforward argument is that once a work becomes a legally enforceable document, it is the equivalent of a law—or *is* a law—and as such, it should be free of any restrictions on further reproduction.

Here, though, as in the “antitrust” issues discussed immediately below, this analysis confuses two things: (1) the nature of a property right on the one hand; and (2) some other legal doctrine that deals with property rights on the other. To deny a copyright to a work otherwise clearly copyrightable, on the grounds that the work has become a law, misses the point: how does a clearly copyrightable work become a law in the first place? After all, most laws never exist in some prior copyrightable state. They are written by legislators as part of their governmental duties and are in the public domain from their very first fixation right up to the point of enactment.¹³² What makes a copyrighted code turn into a law is a particular action by a legislature—the adoption of a pre-existing writing as the official policy of the legislature. In other words, we have a case of some governmental body taking action to turn what was otherwise an item of private property into a public one, for sound reasons of public benefit. We have a substantial extant body of law that deals with this fact pattern under the heading of “eminent domain” and the “takings” clause of the Constitution.¹³³ Governments have many reasons for using private property for public purposes; in some instances, the imposition on the property owner is slight enough that it is not considered a “taking” at all.¹³⁴ In that case, just compensation to the owner is

valid copyright and copying by the defendant of constituent elements of the work that are original. Here, there is no question that [the plaintiff] holds valid copyrights to the building codes and that [the defendant] copied the codes by placing them on the Internet.”).

131. See *Bldg. Officials & Code Adm’rs Int’l, Inc. v. Code Tech., Inc.*, 628 F.2d 730, 736 (1st Cir. 1980) (questioning whether “the public’s essential due process right of free access to the law” can accommodate a non-profit organization’s copyright in a model building code adopted by the Commonwealth of Massachusetts).

132. See, e.g., *Davidson v. Wheelock*, 27 F. 61, 62 (C.C. D. Minn. 1866) (holding that when Minnesota awarded copyright to the publisher of statutes of Minnesota, the publisher acquired the exclusive right to publish the edited statutes with head and marginal notes and references, but not the statutes themselves, which were “open to the world”).

133. “[N]or shall private property be taken for public use, without just compensation.” U.S. Const. amend. V. See *generally* *Pennsylvania Coal Co. v. Mahon*, 260 U.S. 393, 413 (1922) (ruling that when a regulation causes a diminution of “a certain magnitude” in existing property rights, “there must be an exercise of eminent domain and compensation to sustain the act”).

134. *Penn Cent. Transp. Co. v. City of New York*, 438 U.S. 104, 136–38 (1978)

not necessary.¹³⁵ In other cases, the imposition is found to be substantial and is consequently considered a “taking,” for which both due process and just compensation are necessary.¹³⁶

In *none* of the takings cases, however, has a court concluded that if the government has a need for the use of some form of private property, then the property must no longer be “property” at all.¹³⁷ Rather, the cases start with the assumption that something is indeed “property,” and then go on to inquire as to the degree and substantiality of the imposition on the owner’s rights in the property at issue.¹³⁸ Likewise, courts could do the same in copyright and standardization cases. If, for example, a building code is written by a private author and is otherwise sufficiently original and expressive to qualify for copyright protection, then those facts do not change because of any action by a legislative body. Indeed, the text of the Copyright Act itself prohibits such a change in copyright status, and arguably prohibits any action by a legislative body that would invalidate a copyright even with due process and just compensation. Section 201(e) of the Copyright Act specifies that:

When an individual author’s ownership of a copyright, or of any of the exclusive rights under a copyright, has not previously been transferred voluntarily by that individual author, no action by any governmental body . . . purporting to seize, expropriate, transfer, or exercise rights of ownership with respect to the copyright, or any of the exclusive rights under a copyright, shall be given effect

(holding that the application of a historic landmark regulation that prevented the development of a multistory office building above Grand Central Terminal was not a taking requiring compensation because, inter alia, the regulation does not interfere with the present use of the Terminal or with plaintiff’s ability to “obtain a ‘reasonable return’ on its investment”); see also *Pennsylvania Coal*, 260 U.S. at 413 (“Government hardly could go on if, to some extent, values incident to property could not be diminished without paying for every such change in the general law.”).

135. *Penn Cent. Transp.*, 438 U.S. at 136–38.

136. See *Dolan v. City of Tigard*, 512 U.S. 374 (1994) (holding that the City of Tigard could not condition issuance of a building permit on the dedication of a bicycle path and greenbelt by the owner of the land without compensation when the City has made no “individualized determination that the required dedication is related both in nature and extent to the impact of the proposed development”).

137. See, e.g., *Loretto v. Teleprompter Manhattan CATV Corp.*, 458 U.S. 419, 434–35 (1982) (holding that when the “character of the governmental action is a permanent physical occupation of property,” a taking has occurred requiring compensation, even when the occupation is merely the installation of a cable on a landlord’s roof for cable TV) (internal citation omitted).

138. See, e.g., *Penn Cent. Transp.*, 438 U.S. at 130, 136–37 (acknowledging “air rights” as property, but holding that certain restrictions on the use of those air rights did not constitute a taking).

under this title¹³⁹

A government can institute eminent domain proceedings—or the author of a code adopted by a legislature can institute a “reverse condemnation” law suit—to enable a proper resolution of whether the author retains copyright ownership, or whether that ownership is transferred to public hands. Together, the Copyright Act’s section 201(e), and the takings clause along with its related eminent domain power, amply deal with cases in which the copyrightability of an officially adopted standard is at issue. Copyright law is needlessly complicated if these “takings” issues are confused with the question of what is properly a part of copyright’s subject matter.

4. *Antitrust and Monopolies.* Copyright is a form of property right, and property rights are essential to competition in a market economy. Buyers cannot compete with other buyers, or sellers with other sellers, if none of these parties can own something that can be the subject of a sale. That kind of “ownership” is what we mean by “property rights.” But any form of property rights can be broad or narrow. Our legal system could be structured, for example, to grant a “property right” in something as broad as “the sale of any type of computer software.” For that matter, we could create a property right in “the right to sell anything.” Companies could bid for that right in, say, a government auction; the winner would have the right to sell . . . whatever. No other company could sell anything. That would be a broad property right indeed. When property rights become this broad, we usually do not call them “property rights” at all, but rather say that the owner of that right has obtained a “monopoly.” But the different terminologies are just two sides of the same coin. A “monopoly” is the equivalent of a “property right” that is so broad as to be unwise or harmful. Property rights can be narrow as well. One could grant property rights in shoe laces used on left shoes, and a separate right in laces used on right shoes. That would seem an unduly narrow subdivision of rights.

The common law’s response to all these things was generally to create fairly broad rights, and let owners subdivide them as they thought best. That is, of course, what property law’s “bundle of sticks” metaphor means. And nominally, our copyright law does something similar. Section 106 defines a quite broad set of authors’ rights, which can be divided up in arbitrarily many

139. 17 U.S.C. § 201(e). See also I PAUL GOLDSTEIN, COPYRIGHT § 4.4.4.1 n.55 (2d ed. 1998) (“Section 201(e) departs from the approach of [a previous Senate bill] by invalidating *domestic as well as foreign* involuntary transfers.”) (emphasis added); *id.* § 4.4.4.1 (“Takings of copyright under the government’s eminent domain power . . . fall squarely within the terms of section 201(e).”).

smaller ways.¹⁴⁰

Rights can also be aggregated. Although the Copyright Act allots copyright ownership to the initial author of a work, nothing in the Act prevents a single individual, not necessarily even an author, from buying one or more separate copyrights and owning a collection of them. Book publishers do this, for example, whenever they acquire more than one copyright from authors.

How does the law prevent rights from being subdivided into sticks that are “too small” to be practical or wise, or from being aggregated into collections that are “too big” to be practical or wise? The short answer is that the law does not prevent the former at all, but it does prevent the latter under the rubric of antitrust law.

Normal market forces will prevent a too small subdivision of rights—businesses will find it impractical to deal with “left shoe laces” and “right shoe laces” as separate property rights. They will therefore have ample monetary incentives to aggregate these smaller rights as necessary to form a cost-effective larger unit. On the other hand, anti-trust law is designed to prevent undue “bigness,” at least if it tends toward monopoly. The Sherman Antitrust Act since 1890 outlaws monopolization.¹⁴¹ An industry may also have dis-economies of scale that will provide an upper limit on “bigness.”

With normal market pressures preventing a too-small division of property rights, and antitrust law as well as dis-economies of scale preventing a too-large aggregation of rights, there is little need for a separate set of “property ownership” rules to accomplish the same goals. We do not design a special “law of shoe lace ownership” to prevent people from trying to own left and right laces separately. Similarly, we rely on antitrust law and do not design a special “law of car ownership” to prevent an automobile manufacturer or dealer from buying up all other manufacturers or dealers. In the same way, we do not need to design special laws of intangible copyright property ownership to

140. See 17 U.S.C. § 106 (2001) (delineating the exclusive rights belonging to the owner of a copyright); see also H.R. REP. NO. 94-1476:

General Scope of Copyright. The five fundamental rights that the bill gives to copyright owners—the exclusive rights of reproduction, adaptation, publication, performance, and display—are stated generally in section 106 [this section]. These exclusive rights, which comprise the so-called “bundle of rights” that is a copyright, are cumulative and may overlap in some cases. Each of the five enumerated rights may be subdivided indefinitely and, as discussed below in connection with section 201 [of this title], each subdivision of an exclusive right may be owned and enforced separately.

Id.

141. 15 U.S.C. § 2 (1998).

bring about the right “size” and subdivision of copyrights. Copyright owners will find it uneconomical to deal in too-small or in too-large units, and antitrust law will help to ensure the latter result as well.

Consequently, courts faced with questions of potentially new copyrightable subject matter need not be more concerned about potential monopoly effects than they would be in facing any question of property ownership.¹⁴²

5. *Antitrust Tying*. In many interesting copyright cases, the defendant has sought in some way to achieve compatibility with some product or component produced by the plaintiff. In *Sony v. Connectix*,¹⁴³ the defendant created an “emulator” that would be compatible with all of the different video game “units” (which could be game “cartridges,” or in the case of *Connectix*, game CD-ROMs, or any other media) produced for the plaintiff Sony’s PlayStation game console.¹⁴⁴ Each game unit contains computer instructions that, in connection with a live game player’s interactions, will cause a certain sequence of visual displays on the game console. Connectix sought to create an emulator that would produce precisely the same displays in response to precisely the same game unit instructions interacting with precisely the same actions of the game player. *Lotus*¹⁴⁵ featured a defendant likewise creating a kind of “emulator” that would produce the same on-screen (or at least the same numerical) results in response to the same user actions that Lotus 1-2-3 would produce.¹⁴⁶ *Synercom Technologies*¹⁴⁷ featured a defendant who created a kind of software “emulator” that would produce the same engineering analysis results in response to the same data inputs the plaintiff’s software would produce.¹⁴⁸ The defendant in *Apple v. Franklin*¹⁴⁹ similarly sought to produce a computer that would process other software programs in the same

142. On the advisability of adding antitrust doctrine to copyright doctrine, please refer to note 161 *infra*.

143. *Sony Computer Entm’t, Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000), *cert. denied*, 531 U.S. 871 (2000).

144. *Id.* at 598. A player could use the emulator, called the Virtual Game Station, to play Sony PlayStation game CD’s on his or her personal computer. *Id.*

145. *Lotus Dev. Corp. v. Borland Int’l, Inc.*, 49 F.3d 807 (1st Cir. 1995).

146. *Id.* at 810 (noting that Borland offered the “Lotus Emulation Interface,” which would allow users to “see the Lotus menu commands on their screens” and “interact with Quattro or Quattro Pro as if using Lotus 1-2-3”).

147. *Synercom Tech., Inc. v. Univ. Computing Co.*, 462 F. Supp. 1003 (N.D. Tex. 1978).

148. *Id.* at 1009. The defendant sought to “emulate” the Synercom software by copying the data input format so that Synercom customers would not have to reformat their data cards. *Id.*

149. *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240 (3d Cir. 1983).

way that the Apple computer would process them.¹⁵⁰ That is, the defendant sought to create a computer (including both software and hardware) that would “emulate” the Apple computer.¹⁵¹

These cases produced different results. The defendants in *Connectix*, *Lotus*, and *Synercom* were allowed to emulate the plaintiff’s response to data input, while the defendant in *Apple* was not.¹⁵² Clearly, the first three cases involved defendants who undertook a great deal of independent and original work to develop their software.¹⁵³ In fact, the defendants in those cases wrote their own underlying software “from scratch” without copying the plaintiff’s software.¹⁵⁴ In contrast, in the *Apple* case, the defendant gave up that effort as too time-consuming and unreliable, and instead copied the Apple software outright.¹⁵⁵

Very likely, the question of independent creation of the underlying software was influential in the various courts’ approaches to all of these cases. But the crucial issue in these cases is not so much whether the software at hand is

150. *Id.* at 1243. The defendant, Franklin, accomplished producing an Apple compatible computer, in part by literally copying fourteen copyrighted Apple operating system programs. *Id.* at 1244–45.

151. “The [Franklin] ACE 100 was designed to be ‘Apple compatible,’ so that peripheral equipment and software developed for use with the Apple II computer could be used in conjunction with the ACE 100.” *Id.* at 1243.

152. *Sony Computer Entm’t, Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000) (holding that Connectix’s copying of Sony’s copyrighted PlayStation BIOS software while designing the Connectix Virtual Game Station emulator by reverse engineering was fair use when the emulator did not contain any of Sony’s copyrighted material), *cert. denied*, 531 U.S. 871 (2000); *Lotus Dev. Corp. v. Borland Int’l, Inc.*, 49 F.3d 807, 815 (1st Cir. 1995) (holding that “the Lotus menu command hierarchy is an uncopyrightable method of operation,” and therefore Borland’s copying of the menu command hierarchy to create the Lotus Emulation Interface for Borland’s Quattro software was not enjoined); *Synercom*, 462 F. Supp. at 1013; *Apple*, 714 F.2d at 1249, 1253–54 (holding that Franklin infringed when it incorporated virtually identical copies of fourteen Apple operating system programs in Franklin computers, because computer operating system programs written in object code and embedded on ROM are copyrightable).

153. *See Sony*, 203 F.3d at 601 (noting that Connectix engineers developed the Virtual Game Station emulator over a six month period by various reverse engineering processes). The emulator software was written by Connectix and did not contain any of the Sony PlayStation BIOS software code. *Id.* at 600–01. *See also Lotus*, 49 F.3d at 810 (“Borland released its first Quattro program to the public in 1987, after Borland’s engineers had labored over its development for nearly three years.”); *Synercom*, 462 F. Supp. at 1008–09 (stating that the defendant EDI had developed an entirely new structural analysis program, SACS II, and that EDI “wrote a preprocessing program that made SACS II wholly compatible with [Synercom’s] input formats”).

154. Refer to note 153, *supra*.

155. *See Apple*, 714 F.2d at 1245 (“Franklin made no attempt to rewrite any of the programs prior to the lawsuit except for Copy Apple introduced evidence that Franklin could have rewritten programs, including the Autostart ROM program, and that there are in existence operating programs written by third parties which are compatible with Apple II.”).

copyrighted,¹⁵⁶ but whether the plaintiff can assert some sort of right to sell a combination of things, one of which happened to be software, and the other of which was something else. In effect, the plaintiffs sought to protect a higher level grouping: in *Apple*, it was computer-plus-operating software, along with the application programs that ran on that computer; in *Lotus*, it was spreadsheet software, along with the menu of commands associated with it; in *Connectix*, it was game consoles, along with individual game units; and finally, in *Synercom*, it was engineering calculation software, along with a particular set of input data formats to be used with that software.

This view of the “emulation” and “compatibility” cases as an attempt to assert rights in a combination of things has clear echoes of antitrust tying cases—when can a seller condition the sale of a processing unit on the purchase or use of the input data to be processed?¹⁵⁷ Or vice-versa—when can a seller condition the sale of the input data units on the purchase or use of a particular processing unit for that input data? In commerce generally one sees attempts to do this sort of thing all the time in daily life: the instruction booklet for my vacuum cleaner at home cautions me to use only “authorized” bags from the same manufacturer; the label on my dishwasher urges me to use only a named brand of dishwashing powder, from a company presumably owned or contractually committed to the dishwasher manufacturer; the owner’s manual for my car assures me that the use of any other manufacturer’s parts may void my warranty; my color printer’s instructions warn that the best output can only come from using the manufacturer’s own brand of special paper—and heaven forbid that I might actually use another manufacturer’s *inks*; and so on.

If these sorts of product pairings amount to antitrust tying violations, then they are antitrust tying violations. If they are mere “suggestions” only because they are unenforceable as a matter of contract or warranty law, or would be antitrust violations if they were any stronger than that, then they are unenforceable or otherwise just short of unlawful. But in no

156. The copyrightability of Apple’s operating system software, permanently “etched” in the read-only-memory of the computer, was nominally an issue in *Apple v. Franklin*. The court in fact decided the issue quite quickly, finding it to be not an especially difficult question. *Id.* at 1249.

157. See, e.g., *Eastman Kodak Co. v. Image Tech. Servs., Inc.*, 504 U.S. 451, 463–64 (1992) (finding that a requirement conditioning the sale of replacement parts for photocopiers on the purchase of service from Kodak constituting a tying arrangement); *United States v. Microsoft Corp.*, 253 F.3d 34, 64–66 (D.C. Cir. 2001) (finding that technologically binding an Internet browser to an operating system had an anti-competitive effect).

event are these pairings issues of “personal property” law. Likewise, ownership of copyrighted property like computer software might be used to violate contract¹⁵⁸ or warranty law,¹⁵⁹ or to commit an antitrust tying violation,¹⁶⁰ but in none of these cases would the results tell us much about the nature of intellectual property rights per se. Courts can therefore ask whether it makes sense from a copyright law perspective to treat a collection of things (or a single “thing” at a higher level of abstraction) as a copyrightable entity, without regard to whether such a collection might violate some other body of law.¹⁶¹

C. *An Incentives-Based Analysis of New Subject Matter*

It’s common to observe that a copyright prevents competition,¹⁶² so that copyright protection should only apply to those works that would not have been created without the existence of that incentive. If a given work would have been created with or without copyright, then copyright protection simply prevents competition in the reproduction and distribution

158. See, e.g., *Nat’l Car Rental Sys., Inc. v. Computer Assocs. Int’l, Inc.*, 991 F.2d 426, 431–32 (8th Cir. 1993) (rejecting the argument that contract claims are pre-empted by copyright infringement claims).

159. See, e.g., *Christopher v. Cavallo*, 662 F.2d 1082, 1083 (4th Cir. 1981) (affirming that fraudulently representing a work as not infringing upon a copyright can constitute a breach of warranty of good title).

160. See, e.g., *Microsoft*, 253 F.2d at 62–63 (refuting the argument that the use of a copyright cannot give rise to antitrust liability).

161. Some courts have interpreted copyright law in the light of antitrust principles by developing a doctrine of “copyright misuse.” See, e.g., *Lasercomb America, Inc. v. Reynolds*, 911 F.2d 970, 979 (4th Cir. 1990) (finding that a party’s inclusion of “anticompetitive clauses in its standard licensing agreement constitute[s] misuse of copyright,” and holding that such misuse prevented the party from suing for copyright infringement). Although one can question what the source of this federal common law power is, I do not do so here. I do think, however, that interpreting property ownership rules in light of antitrust doctrine is no more necessary or appropriate with the law of intangible property than it would be with the law of real or personal property. Second, I think that meshing antitrust (or other legal principles like warranty or contract) and copyright principles together creates a much more complex body of copyright law than would exist if the two remained separate. Finally, I think that neither the public nor the judicial system obtains any significant gain from the introduction of that greater complexity. Indeed, my own view is that copyright law is vastly more complex already than is good for us, and that on that ground alone, courts ought to be very hesitant to add even more complexity, such as judicially-created rules about copyright misuse, to deal with policy concerns that can amply be dealt with under existing antitrust laws.

162. See, e.g., Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs*, 84 HARV. L. REV. 281, 281 (1970) (stating that the exclusive rights produced by copyright law prevent competition in book publishing); Matthew P. Larvick, *Questioning the Necessity of Copyright Protection for Software Interfaces*, 1994 U. ILL. L. REV. 187, 199 (1994) (stating that copyright protection conflicts with the free market’s “tenet that copying and improving” upon products leads to greater competition).

of the work and keeps the price higher than it otherwise would have been. In a close case on an issue of copyrightability, then, an important factor for courts to consider is whether an incentive to creation is needed for the type of work at issue.

But it is naïve to think only in terms of whether “an incentive” is “needed or not needed.” Undoubtedly, some novelists would write novels out of a love for writing or a need for self-expression even if novels were not copyrighted; some composers would compose; some programmers would program; and so forth. If one thinks of incentives only in terms of individuals, one might therefore conclude that for novelist “A,” a copyright incentive is needed, but for novelist “B,” love of writing is enough, and so on. One might then be tempted to craft an incentive policy by imagining incentives as binary—needed, or not needed. But a policy like copyright law that applies to hundreds of millions of citizens must be thought about in more statistical terms, in terms of aggregate effects and probabilities. A more accurate inquiry into the policy of incentives is therefore to ask whether the incentive of copyright ownership would be likely to make a significant difference in the total output of copyrighted works.

Even the notion of the “output” of creative works has to be examined. We care about both the creation and the dissemination or use of works. If copyright’s incentive structure brought forth only secret diaries that no one but the author ever read or saw, or wonderful computer programs that were never used, we might consider the policy to be a failure. We want an incentive policy, in short, that will make a difference.

Finally, our thinking about incentives must acknowledge quality as well as quantity. Creativity—certainly in a corporate environment, but perhaps also for the lone author working in a garage—comes in degrees of investment. Once the initial decision is made to create something, there is often still a huge range of effort (money, time, priority, etc.) that can be put into the creation. In general, more effort will bring about a better result and less effort a worse one. Investment incentives can make a difference, in short, both to the quantity of works that are produced and disseminated, and also to the quality of those works.

In a case like *Lotus*,¹⁶³ this difference is crucial. If the only question were whether a copyright incentive would be likely to affect the total output of menu command hierarchies, or more broadly, “user interfaces,” the answer would be “no.” With the

163. *Lotus Dev. Corp. v. Borland Int’l, Inc.*, 49 F.3d 807 (1st Cir. 1995).

state of computer technology at the time (and still true today, though one hesitates to predict much beyond tomorrow), a piece of software like a spreadsheet must have some means for users to interact with it. The very fact of a company's decision to invest in the creation of a commercial software product—a product that is itself protected by a copyright in the “code” written by programmers—necessarily entails a decision to create the means for user interaction. In that sense, the copyright on the underlying program code is enough by itself to ensure that a menu command hierarchy or other user interface will also be created. Otherwise the program cannot be used. Moreover, even in the absence of copyright protection, the market will ensure that a user interface cannot be *too* poorly designed, else the product will not be as successful as it could be.

Apart from any copyright protection on interfaces themselves, then, we can see that there are two other incentives that play an important role in software interface development. One is an incentive to create some sort of interface at all, in order that the underlying program can be used.¹⁶⁴ The second is an incentive to make that interface “reasonably” good, or perhaps we should say “good enough,” or else the product will sell less well than it might.¹⁶⁵

But here is the point: if there is a range of quality in user interfaces *beyond* the “good enough” level, then an incentive tailored to user interfaces as such—different from the incentive needed to induce development of the underlying software—would be likely to make a significant difference. So a crucial question for courts facing the copyrightability of something like a user interface, as the *Lotus* court faced, is whether a significant gap lies between the “good enough” interface that will be produced in any event, and the “very good” or “excellent” interface that requires an additional incentive. If there is little gap, then the need for copyright's incentive for the interface is correspondingly weak.¹⁶⁶ If the gap is large, then the incentive of copyright protection is more important.¹⁶⁷

164. See, e.g., *id.* at 815 (explaining that a menu command hierarchy was necessary for the software to be used at all).

165. Refer to note 178 *infra* and accompanying text (noting that Microsoft researched software usage in order to produce a better product). See also *Lotus*, 49 F.3d at 809–10 (noting that “macros” had been added in order to decrease the time required for the user to set up and operate the program).

166. The benefit of promoting development of the “improved” interface must be weighed against the cost of decreased competition in order to determine whether the protection afforded by copyright law is appropriate. Larvick, *supra* note 162, at 206.

167. If there is little copyright incentive, firms may be less inclined to innovate. *Id.*

A corollary to this point is that the relevant consumers must care about the range of quality differences. It is at least theoretically possible that a wide range of quality levels might exist in some sort of work, like a software user interface, but that consumers would not care about those differences. If that were true, then the application of copyright protection to higher quality levels would be irrelevant.¹⁶⁸ If consumers did not care about quality differences, then interface authors would have no reason to invest in creating better quality interfaces, or in advertising and promoting them, whether they were copyrightable or not.

Depending on the particular facts, factual findings in copyright cases involving some new subject matter would therefore be necessary to establish several things: (1) whether the incentive of copyright would be likely to make a significant difference in the output and dissemination of works of the subject matter at issue; (2) whether a range of different quality levels exists for the subject matter beyond the normal market incentives to make that subject matter usable and “good enough”; and (3) whether that range of difference is likely to matter to consumers.

D. Framework Applied to Past Cases

1. *Synercom Technologies*. We can apply this general framework to other cases as well, such as *Synercom Technologies*.¹⁶⁹ I noted earlier that denying copyright protection to the input data formats in that case would result in a loss of competition over the development of different varieties of input formats.¹⁷⁰ But denying copyright protection to a given type of subject matter will always, by definition, result in a loss of competition over the development of new varieties of that same subject matter. Yet we do not extend copyright to every possible subject matter, so there must be some offsetting values to the value of “competition.” How should courts make the assessment in relation to input data formats particularly? Again, I suggest a consideration of three factors: (1) whether the incentive of copyright will likely make a significant difference in the quantity of input data formats produced and disseminated; (2) whether there is a range of quality offerings in input formats, a range extending beyond what is “good enough” to bring the overall

168. If the copyright incentive is large, firms may be more inclined to innovate, but there is a risk of decreased competition. *Id.*

169. *Synercom Tech., Inc. v. Univ. Computing Co.*, 462 F. Supp. 1003 (N.D. Tex. 1978).

170. Refer to note 127 *supra* and accompanying text.

software product to market;¹⁷¹ and (3) whether that range of quality differences is likely to matter to the product's consumers.¹⁷²

If the answer to all these questions is "no," then any additional copyright incentive on the input formats would be superfluous and should not be granted.¹⁷³ First, in the *Synercom* case in particular, some sort of input data format is essential for the underlying software product to be usable at all.¹⁷⁴ If software is developed that processes data, there must be some way for the data to be provided as input to the process. Granting copyright protection to those formats does not seem likely to make much of a difference in the quantity of such formats produced.

Equally important here is the quality issue. Yet, in the *Synercom* case, it seems intuitively likely that expert testimony would have shown a very narrow range of quality variations in the different input data formats. That is, one imagines that the different formats were roughly comparable in ease of use, accuracy, or whatever other indicia of quality one might come up with. Finally, it seems equally likely—based again on nothing more than intuition—that customers were not particularly sensitive to these differences for their own sake.

If all of those observations were true—and "truth" here is a matter for factual findings in the individual case—then the *Synercom* court's conclusion that the input data formats were not separately copyrightable was correct. The addition of copyright protection would not add to the developer's incentives and was therefore unnecessary. In short, copyright law in a case like *Synercom* should recognize that the relevant competition is over the underlying software product's capabilities and features, not over the input data formats per se.¹⁷⁵ As long as the second comer has independently written its own underlying software, it should be allowed to make use of the leader's input data formats

171. Refer to notes 166–68 *supra* and accompanying text (describing the requirement that there be a significant gap between the quality of works that would be available with the copyright incentive and the quality of works that would be available without the copyright incentive).

172. Refer to note 169 *supra* and accompanying text (describing the requirement that the consumers actually care about any gap in quality caused by a copyright incentive).

173. Refer to note 169 *supra* and accompanying text.

174. *Synercom*, 462 F. Supp. at 1005 ("In using a program one must have a format for input so that the input of data and the instruction to the computer are compatible with its program.").

175. See *id.* at 1008 (stating that EDI was in a position to "pluck the fruit of Synercom's labors and risks," as long as EDI's product was as good as or better than Synercom's product).

precisely for the purpose of trying to take customers away from that leader.¹⁷⁶

2. *Lotus v. Borland*.¹⁷⁷ The *Lotus* case is quite different from *Synercom* in this regard. With all due allowance for the fact that experts must in fact establish the point, we can guess that a set of four hundred and sixty-nine user commands and their arrangement in menus, sub-menus, sub-sub-menus, and so on, would offer a range of quality well beyond whatever is “good enough” to bring the product to market. In part, I base this intuition on personal experience—I often find myself frustrated in using software by not being able to figure out where in a menu hierarchy is located the command that I want to use. Sometimes this results in my pulling down each of the top-level menus of a software package, one at a time, scanning each drop-down list for the command I want. From that experience, I conclude that some menu arrangements might make more sense than others—that some arrangements would in particular be easier to remember than others. And in turn, that means that a software developer could spend less or more time and research on these commands and produce a worse or better product accordingly.¹⁷⁸

All these things suggest both the likelihood that an additional copyright incentive would produce more effort in the creation of a user interface from a software developer than just the minimum needed to bring out a product on the market, and that customers of the product would care about the resulting quality differences. If we hypothesize that expert testimony would have confirmed these intuitions in the actual *Lotus* case, then the district court was right and the appellate court was wrong—copyright protection should have been extended to the user interface itself because the resulting incentives would have mattered to the quality of different user interfaces, customers would likely care about such differences, and Borland should accordingly have been found to be a copyright infringer.

176. *Id.* at 1013 n.5 (“EDI’s preparation of a FORTRAN preprocessor program from the descriptions contained in the manuals cannot constitute an infringing derivative use provided this was done without copying of the plaintiff’s FORTRAN program, as it was.”).

177. *Lotus Dev. Corp. v. Borland Int’l, Inc.*, 49 F.3d 807 (1st Cir. 1995).

178. I also know that at least some companies do undertake research on these things: some years ago, I (presumably one of many people) received a request from Microsoft to install a special “instrumented” version of Microsoft Word. The stated purpose of this version of Word was to record, over a period of months, the different Word commands I used and their order and frequency.

3. *Sony v. Connectix*.¹⁷⁹ *Sony v. Connectix* is not about user interfaces or copyrightable subject matter, but it can be analyzed in this same way. In the actual case, plaintiff Sony did not argue that the Connectix software emulator infringed Sony's operating system program in the PlayStation console.¹⁸⁰ The emulator was independently written by Connectix.¹⁸¹ Sony's argument was that in order to develop the emulator, Connectix had to reproduce the Sony programs a number of times for analysis and dissection—a part of its reverse engineering of the console's technology.¹⁸² It was this "intermediate" copying that Sony objected to on copyright grounds.¹⁸³ The Ninth Circuit adhered to its own close precedent in *Sega Enterprises v. Accolade*,¹⁸⁴ concluding and quoting *Sega* that "where disassembly is the *only way to gain access to the ideas and functional elements embodied in a copyrighted computer program* and where there is a legitimate reason for seeking such access, disassembly is a fair use of the copyrighted work, as a matter of law."¹⁸⁵

But it is possible to examine the case in a different light, as involving a subject matter issue. Sony could be said to have asserted copyright ownership in a higher level of abstraction than the console program code. It sought, in effect, to assert a copyright in "the technology needed to respond to the combination of game unit and player inputs in a specified way," namely, the way that the PlayStation console responded.¹⁸⁶ If it could copyright that "response technology," then Connectix would be an infringer, for it reproduced precisely (or nearly so) that same response technology, albeit by using different software.¹⁸⁷

Sony already has an incentive to make (or to license others to make) good individual games, as each such game is a clearly

179. *Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000), *cert. denied*, 531 U.S. 871 (2000).

180. *Id.* at 598 ("The [Connectix product] Virtual Game Station does not contain any of Sony's copyrighted material.").

181. *Id.*

182. *Id.* at 601 (detailing the reverse engineering performed by Connectix).

183. *Id.* at 598 ("Connectix repeatedly copied Sony's copyrighted BIOS during a process of 'reverse engineering' . . . Sony claimed infringement . . .").

184. *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1514 (9th Cir. 1993) (stating that Sega's video game console was reverse engineered to enable the defendant to create its own game units that would be compatible with the Sega console).

185. *Sony*, 203 F.3d at 602 (quoting *Sega*, 977 F.2d at 1527–28 (emphasis added by the *Sony* court)).

186. *Id.* at 598.

187. *Id.* at 599 ("[A] consumer can load [Connectix] Virtual Game Station software onto a computer, load a [Sony] PlayStation Game into the computer's CD-ROM drive, and play the PlayStation game.").

copyrightable computer program,¹⁸⁸ and the benefits of making better game programs will inure directly to Sony in the form of greater sales of the game units. Sony likewise has an incentive to make some sort of “response technology” at all, as that technology is necessary to be able to establish the game playing market.¹⁸⁹ So the PlayStation “response technology” is in many ways similar to the “user interface” of *Lotus* because once the decision is made to produce and sell game units (or spreadsheet software in *Lotus*), there is also a built-in incentive to produce and sell the necessary response technology (or the user interface of menu commands in *Lotus*).¹⁹⁰ As was true in *Lotus*, in *Connectix* normal market forces will ensure that the response technology be “good enough,” or else it will sell less well than it otherwise might.¹⁹¹

So once again, we must focus on the quality issue: whether there is a range of different quality levels in the response technologies, beyond what is “good enough” to sell the underlying console and games in the first place, and whether customers care about any such different quality levels.¹⁹²

Again, factual findings would be called for in an actual case, because the answers to these questions turn on the facts of the technologies and markets at issue. But we can indulge in speculation. Mine is that a game response technology probably has to be “pretty darned good” in order to be good enough to bring it to market at all. Those who enjoy games of this sort seem to be demanding and critical.¹⁹³ Responses must be quite quick: most

188. See, e.g., *Computer Mgmt. Assistance Corp. v. Robert F. DeCastro, Inc.*, 220 F.3d 396, 400 (5th Cir. 2000) (“Computer programs are entitled to copyright protection.”); *Sony*, 203 F.3d at 602 (stating that the object code of a computer program can be copyrighted as expression under 17 U.S.C. § 102(a)); *Fonar Corp. v. Domenick*, 105 F.3d 99, 104 (2d Cir. 1997) (stating that source and object codes of computer programs are subject to copyright protection); *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1247–48 (3d Cir. 1983) (explaining that the 1980 amendments to the copyright code clearly indicate that computer programs are copyrightable).

189. I am assuming that it would not be feasible to produce the game units first and then wait around until someone reverse engineered them to develop game consoles. That is, I am assuming that *Connectix* would not initially develop a “console” emulator unless the market was already pretty far along based on a console produced first by Sony.

190. *Compare Lotus Dev. Corp. v. Borland Int'l, Inc.*, 49 F.3d 807, 815 (1st Cir. 1995) (stating that without the menu command hierarchy, users would not be able to make use of *Lotus*'s software), *with Sony*, 203 F.3d at 598 (stating that Sony PlayStation console was the device marketed by Sony to permit consumers to play Sony PlayStation games).

191. Refer to note 166 *supra* and accompanying text (explaining that if a product is not “good enough,” it will fail to reach its full sales potential).

192. Refer to notes 167–69 *supra* and accompanying text (explaining the requirements that a significant gap in possible quality must exist and that consumers must care about the gap).

193. Most of my speculation on these points is based on casual reading of a variety of

of the PlayStation and similar games are intended to operate in “real time,” with human or exotic creatures running, climbing, jumping, and so on, with at least as much speed as they would exhibit were they “real.” Moreover, faster response technologies would enable even faster or more realistic games to be developed, which would in turn cause the consoles to sell in greater volume.¹⁹⁴ Both of those desirable qualities—playing existing games fast enough, and offering the ability to play even more complex games fast enough in the future as those games are developed—seem to me to have ample sales incentives *without* the need for additional copyright protection either on “the technology needed to respond to game and player inputs in a specified way.”

Conceivably, an expert witness could show the contrary: that a range of quality levels does exist in “the technology needed to respond to game and player inputs in a specified way” and that customers care about those quality levels. But, to the extent that expert testimony supported my speculations about the video game market,¹⁹⁵ the Ninth Circuit’s conclusion in *Connectix* that Sony had no right to prevent Connectix from marketing a PlayStation emulator was in fact the proper one.¹⁹⁶

4. *Bender v. West*.¹⁹⁷ *Bender v. West* is a delightful case because the essential subject matter question, dealing with printed page layouts and pagination, has so little to do with modern technology.¹⁹⁸ Still, it was modern technology that brought the issue to the point of litigation. The development of low-cost CD-ROM production (not to mention the general computerization of court opinions, word processing, etc.) enabled the challenger HyperLaw to enter the business of computerized court opinion sales.¹⁹⁹ The

computer industry magazines over a period of years. Those readers who have children who are “gamers” (or are gamers themselves) can verify my findings independently.

194. Refer to note 166 *supra* and accompanying text (noting that an inferior interface will not sell as well as a superior interface).

195. Refer to note 194 *supra* and accompanying text (speculating that faster response technologies provide sufficient sales incentive without the need for additional copyright protection).

196. *Sony Computer Entm’t, Inc. v. Connectix Corp.*, 203 F.3d 596, 609–10 (9th Cir. 2000) (denying Sony’s request for injunctive relief to prevent Connectix from marketing their Virtual Game Station), *cert. denied*, 531 U.S. 871 (2000).

197. *Matthew Bender & Co., Inc. v. West Pub’g*, 158 F.3d 693 (2d Cir. 1998).

198. *Id.* at 695 (noting that the suit arose due to plaintiff Matthew Bender’s desire for a declaratory judgment that its insertion—into its own “compilations of judicial opinions” stored on CD-ROMs—of citations to corresponding pages in “West’s printed version of the opinion,” did not infringe West’s copyrights).

199. *Id.* at 697 n.4 (highlighting greater efficiency of CD-ROM publications when compared to books).

century-old print medium of the West company had created a *de facto*—and for many jurisdictions, the *de jure*—standard for case citation.²⁰⁰ Much as Connectix produced something that was compatible with a widely used standard—the PlayStation game units²⁰¹—and Borland produced something compatible with a widely used standard—the Lotus 1-2-3 menu tree²⁰²—and Franklin produced something compatible with a widely used standard—Apple application programs²⁰³—so Bender and HyperLaw had produced something compatible with the widely-used West page citation standard.²⁰⁴

I have already suggested that official adoption of a copyrighted work into law calls for the application of “takings” jurisprudence and the law of eminent domain.²⁰⁵ *Bender* offers another example of the phenomenon of public adoption into law of a privately-created work (at least in those jurisdictions that required West page citations),²⁰⁶ although not as obviously so as the cases involving building codes and the like. But the predicate to the application of takings jurisprudence is that a property owner actually own property.²⁰⁷ To be sure, courts would simplify their analysis and the law by not considering takings questions in the determination of whether a valid copyright exists in the subject matter in issue. But they must still engage in that initial determination. How would that determination be made for the West Reporter pagination with the approach I propose in this Article?

A threshold question is what the particular subject matter is that we should analyze. The narrowest view of that subject matter—in my experience, the one most commonly asserted in

200. *Id.* at 696.

201. *Sony*, 203 F.3d at 598.

202. *Lotus Dev. Corp. v. Borland Int'l, Inc.*, 49 F.3d 807, 810 (1st Cir. 1995).

203. *Apple Computer Corp. v. Franklin Computer Corp.*, 714 F.2d 1240, 1242–43 (3d Cir. 1983).

204. *Bender*, 158 F.3d at 697–98 (noting that “Bender intends to insert . . . a parallel citation . . . to the West case reporter at the beginning of the opinion and a citation to the successive West page numbers at the points in the opinion where page breaks occur in the West volume” and that “Hyperlaw includes parallel citations to West’s case reporters for all cases appearing in [two of West’s reporters], and intends to add star pagination as well”).

205. Refer to note 133 *supra* and accompanying text (explaining that a large body of law—the law of eminent domain and the “Takings” clause of the Constitution—already exists for the same fact pattern as copyright).

206. *Bender*, 158 F.3d at 696 (“One citation guide recommends—and some courts require—citation to the West version of federal appellate and trial court decisions and New York State court decisions.”).

207. Refer to note 138 *supra* and accompanying text (stating that takings cases always start with the assumption that the subject is property).

informal discussions of the case—was that it was “page numbers.”²⁰⁸ West’s primary argument, however, was that copyright protection was appropriate at a higher level of abstraction, namely that of the sequence and arrangement of cases in its Reporter volumes.²⁰⁹ In the actual case, the Second Circuit seemed to accept, at least for purposes of its decision, that West’s overall sequence and arrangement was in fact copyrightable.²¹⁰ The case turned instead on an issue of infringement: had HyperLaw “copied” the West arrangement by introducing star pagination into its cases?²¹¹

Infringement and subject matter questions are often just the flip sides of the same coin, however.²¹² A very narrow definition of “subject matter” will inevitably result in a finding of few infringements;²¹³ a very broad definition of “subject matter” will result in many more. We can therefore usefully analyze the *Bender* facts as a subject matter question, but we must define that subject matter in terms of something that HyperLaw actually did “take” from West. If what HyperLaw did take was copyrightable under the approach of this Article, then this Article implies that HyperLaw was an infringer and the Second Circuit was wrong (and vice-versa).²¹⁴

If we accept the court’s finding that HyperLaw did *not* take West’s sequence and arrangement,²¹⁵ then we should look elsewhere for what was taken. The most likely subject matter is either the very thing so often asserted informally—West’s “page numbers”—or perhaps the slightly higher level of abstraction of “pagination,” which implies a bit more than just “numbers” as the subject matter of interest but a bit less than “sequence and arrangement.”²¹⁶

208. *Cf. Bender*, 158 F.3d at 699 (“At issue here are references to West’s volume and page numbers . . .”).

209. *Id.* at 700 & n.11.

210. *Id.* at 700 n.11 (“[W]e can assume without deciding that West’s case reporters contain an original and copyrightable arrangement.”).

211. *Id.* at 701–02.

212. Refer to note 26 *supra* and accompanying text (noting that courts deciding copyright cases still often face subject matter questions).

213. Refer to notes 28–33 *supra* and accompanying text (discussing *Wood*, in which the application of a narrow dictionary definition of the word “prints” led to the conclusion that photographs were not copyrightable).

214. Refer to notes 171–72 *supra* and accompanying text (listing the factors necessary for copyright infringement under the approach taken by this Article).

215. *Bender*, 158 F.3d at 702 (“[T]he sequence of cases [in *Bender*’s and HyperLaw’s CD-ROM discs] owes nothing to West’s arrangement.”).

216. *Id.* at 695 & n.1 (describing the “star pagination” used by *Bender* and HyperLaw to embed citations).

Question number one under this Article's approach is thus whether the incentive of copyright protection would be likely to make a significant difference in the total output of page numbers or pagination, or in their dissemination to potential users of those numbers or pagination. Expert testimony might prove otherwise, but certainly one's very strong instinct is to conclude "no." Page numbers may be thought of as analogous to the user interface of a software package. If a business is going to produce a book (or a software package), it will almost always have to number the pages (supply a user interface) for the book to be useful. Perhaps the necessity with page numbers is slightly less than it is with a user interface, for a book can be read without page numbers, but the utility of numbering is so great that it seems likely that a copyright on page numbers would not make a significant difference to the total output of those numbers.

Looking at the issue in terms of the slightly more abstract level of "pagination" surprisingly makes the case for copyright protection even weaker. Although a book can be printed and read without page numbers, it is not a "book" (at least not in the usual sense) if it lacks "pagination" of some sort. To be sure, many electronic documents can appear as "scrolls" without particular page breaks, but then presumably the owner of such a work would not attempt to claim a copyright in "pagination" in the first place. We can conclude that it is unlikely that a copyright on "pagination" would result in a significantly greater total output of "pagination" than would be made available to the public otherwise.²¹⁷

The second question is whether there exists a range of quality levels in page numbers or pagination, a range of quality levels that the relevant consumers would be likely to care about. This question is actually slightly more difficult than the previous one, at least in regard to "pagination." I would guess that there is little or no range of quality in "page numbers." Not that there are no choices to be made—a publisher must decide whether to have more than one set of numbers, for example. Perhaps some introductory pages should start at Roman numeral "one," and the main text restart at "one"; perhaps an index should restart at "one" again; or perhaps there should be only a single sequence throughout. But the existence of "some choices" is not the same thing as the existence of a significant range of quality levels in these choices *beyond* what would be called for as "good enough" to

217. Refer to note 165 *supra* and accompanying text (stating the proposition that if an interface is necessary to use the product, an adequate interface will be designed even in the absence of copyright protection).

bring the underlying book to market.²¹⁸ And I very much doubt—though again, expert testimony could always establish the contrary—that readers care significantly about these choices in page numbers.

“Pagination” seems to offer a stronger argument when we concentrate on quality levels as opposed to quantities. Books come in a variety of physical sizes, with different margins, type faces, “leading” (spacing between lines), and so on. There seems to be a considerably greater range of choices in these things, which are intimately related to a book’s “pagination,” than exists with “page numbers” per se. An expert might testify that “pagination” in this sense has different quality levels—some types of pagination might be easier to read, or more aesthetically pleasing, and so on. An expert might also be able to say that these quality differences matter to readers as well, for perhaps whether a book has “many small pages,” “few large pages,” “generous spacing,” or “large type face” affects readers differently.

But the danger of this analysis is that it tends to muddy up what the subject matter in question really is. Do we mean by “pagination” actual physical page size? Do we also mean “type face”? Do we mean “margins”? Fortunately, if we bring the analysis back to the facts of the *Bender* case, we can anchor the discussion much more easily. In *Bender*, none of those things was “taken.” At most, “pagination” on the *Bender* facts means something like “the choice of page breaks alone,” as presumably a CD-ROM would not incorporate West’s type faces, page sizes, etc., as such.²¹⁹ So we really only need to ask: (1) whether there is a range of different quality levels in “the choice of page breaks alone” beyond whatever would be good enough to bring the underlying product (a collection of cases) to market; and (2) whether consumers would be likely to care about those different quality levels.²²⁰ Viewed in this light, the answer continues to appear to be “no,” with appropriate allowances for the fact that expert testimony must establish this fact in any given case.

In the end, analyzing *Bender* as a subject matter case, focusing on what was actually taken as the appropriate subject

218. Refer to notes 167–69 *supra* and accompanying text (discussing the gap between an interface that is “good enough” and one that is “very good” or “excellent”).

219. See *Bender*, 158 F.3d at 705 n.20 (explaining that *Bender*’s product displays the cases in one column with no page breaks and with endnotes, but West’s reporters display cases in two columns and with endnotes).

220. Refer to notes 167–69 *supra* and accompanying text (discussing the requirements that a gap in possible quality must exist and that consumers must care about the gap).

matter, leads us to the conclusion that the case was correctly decided on copyright grounds.²²¹

III. WHAT IS XML?

A major purpose of this Article is to apply its proposed approach for analyzing new subject matter cases to a new form of potentially copyrightable work to see what the approach would predict. I have chosen “XML schemas” as that example. The following section explains both “XML” and “XML schemas” for readers not familiar with those concepts. A good awareness of both concepts is essential to an understanding of the subsequent legal analysis. After this section, Part IV will sketch out how prior courts might approach the question of schema copyrightability, and then try to answer the same question with the approach outlined in this Article. Readers who already know what “XML” and “XML schemas” are can safely skip this section and proceed directly to Part IV—Are XML Schemas Copyrightable? Others are encouraged to read on.

About thirty years ago, the IBM corporation hired a lawyer named Charles Goldfarb to become a computer programmer.²²² IBM assigned him to work on, among other things, typesetting and word processing with computers.²²³ Word processing technology of the day (and even today) primarily identified only one thing for documents in a computer readable medium: the format for printing the document.²²⁴ Typical word processors contained special codes to indicate what parts of the document should be in bold face, what parts in large type sizes, how big the margins should be, and so on.²²⁵ What word processors could not do was indicate, for example, what parts of a document were the “introduction” or even what parts indicated the author of the document. If a document were a court opinion in a word processing format, for example, nothing in the word processing technology could unambiguously indicate that “this paragraph is a citation,” while “that paragraph is a holding,” or that “this sentence indicates procedural history,” while “that one indicates the attorneys’ names.”

221. It is possible that West had a better case under some other body of law, such as unfair competition; I do not analyze such possibilities here. The court of appeals held that Bender and HyperLaw would not infringe on West’s copyright by including star pagination in their CD-ROM versions of judicial opinions. *Bender*, 158 F.3d at 708.

222. Charles F. Goldfarb, *The Roots of SGML—A Personal Recollection*, at <http://www.sgmlsource.com/history/roots.htm> (on file with the Houston Law Review).

223. *Id.*

224. *Id.*

225. *Id.*

True, some specialized systems for these purposes had been developed. Lexis (and later Westlaw), for example, had developed its own methods of doing exactly that. The Lexis database from its earliest days separately identified most of these kinds of “parts” in court opinions, and allowed searching within the parts individually. But the Lexis system, like other specialized systems, was designed for court opinions and other legal documents. It would not be possible to use the Lexis system to contain, for example, articles from medical literature because those articles did not fall naturally into parts called “CITATION” or “COURT” or “SUBSEQUENT HISTORY,” all of which make sense only in the specialized arena of the law.

When Lexis began adding other types of documents, such as news articles, it developed additional specialized search terms that made sense in that new context. But again, the terms used, such as “BY-LINE,” make sense in the context of news, and were not easily adapted for searching altogether different documents—for instance, corporate invoices.

What IBM’s Goldfarb, working with colleagues Ed Mosher and Ray Lorie, sought to develop was a more generalized method of creating “search terms” for any kind of document or, more precisely, for creating the labels that could be used to identify the different parts of any kind of document whatsoever.²²⁶ And they succeeded.²²⁷ They created a “mark-up language” that could be used to label the parts of any document in any fashion that the user of the document chose.²²⁸

The “language” relied on the use of so-called “markup tags” or labels that were interspersed throughout a document.²²⁹ An SGML “tag” consists of a word surrounded by “angle brackets,” so that a tag for OPINION would be written “<OPINION>”.²³⁰ The tags had to come in pairs because they were designed to surround the part of a document that they labeled.²³¹ That meant there had to be both beginning and ending tags.²³² Goldfarb and his colleagues settled on a “slash” character to indicate that a particular tag was an ending tag, like this: “</OPINION>”.²³³

226. *Id.*

227. *Id.*

228. *Id.*

229. ELECTRONIC TEXT CENTER, UNIVERSITY OF VIRGINIA, TEI GUIDELINES FOR ELECTRONIC TEXT ENCODING AND INTERCHANGE, Part 1, para. 2, at <http://etext.Virginia.edu/TEI.html> (last visited July 21, 2001).

230. *See id.* at para. 2.3.1.

231. *Id.*

232. *Id.*

233. *Id.*

Because they had developed a technique that was, by design, not specialized to any particular type of document, industry, or profession, they named it the “Generalized Mark-up Language” or “GML.”²³⁴ Later, after more refinement, this system became known as the “Standard Generalized Markup Language,” or “SGML.”²³⁵

SGML worked by allowing anyone to determine the parts of any document and give them any name, just as Lexis had determined that opinions could be divided into “WRITTENBY” and “COURT” and “OPINION” and similar parts.²³⁶ But the SGML breakthrough was to allow an SGML author to create a separate set of *definitions* of these names and parts, along with a way to indicate their relationship to each other.²³⁷ These definitions could express, for example, that the “WRITTENBY” part must always come before the “OPINION” part. Or that an “OPINION” part could contain multiple “PARAGRAPH” parts,²³⁸ but not the other way around.

By formalizing this separate set of definitions, Goldfarb and his colleagues enabled programmers to write programs that could read a set of SGML definitions first, then go on to read and interpret any document with its parts labeled in accordance with those definitions.²³⁹ The result of this development was that, at least in theory, a single document processing program could search databases of court opinions, medical articles, news articles, or corporate invoices.²⁴⁰

In addition, a “document processing” program need not be confined to searching a database. SGML also enabled such a program to function as a word processing program in the conventional sense. That is, such a program could be instructed

234. Incidentally, “Generalized Markup Language” or “GML” is also the first initial of the founders’ last names: Goldfarb, Mosher, and Lorie. *Id.*

235. *Id.*

236. *Id.* (explaining that SGML operates independently of the semantics of textual elements, and can be tailored to operate with specific applications).

237. *Id.* at para. 2.4.

238. I don’t know that Lexis actually formalizes the notion of “paragraphs” or not, but they make a good example of one way that an “OPINION” could be subdivided.

239. See Adrienne Azaria, *SGML: A Lifesaver in a Sea of Electronic Documents*, NETWORK WORLD, Dec. 12, 1994, at 67, LEXIS, News Library, Computers/ Communication Archive News file (stating that SGML allows programmers to design programs that can identify, manage, and share information in documents that have been defined by document type definitions (DTDs)).

240. See *id.* (noting that, by preserving the content and structure of documents, SGML makes it easier and faster for programs to locate information from different networks and applications).

that "WRITTENBY" should be printed in bold face, or that each "PARAGRAPH" should be preceded by a blank line.²⁴¹

Companies that wanted to make use of SGML to control document printing in this way nonetheless had to write their own set of formatting instructions.²⁴² In other words, SGML provided a way to identify parts of documents, like "TITLE" or "FOOTNOTE," but it did not provide any built-in mechanism to control the appearance of those parts on paper.²⁴³ Companies would write their own processing programs to do the latter task, and each company that wanted to use SGML in this way wrote their own, proprietary set of programs to do so.²⁴⁴

By the mid 1980s, word processing programs and SGML were more or less complementary but opposite.²⁴⁵ Word processing programs provided a way to identify parts of a document as "BOLD FACE" or "DOUBLE SPACED."²⁴⁶ But they provided no built-in way to label parts of a document as "OPINION" or "CITATION."²⁴⁷ On the other hand, SGML allowed an easy identification of document parts as "OPINION" or "CITATION," but had no built-in way of labeling those parts to appear in "BOLD FACE" or "DOUBLE SPACED."²⁴⁸

Some work was eventually done to develop a merger of those two capabilities. The notion was that it ought to be possible to create a "style sheet" that would say, for example, that any document part labeled "OPINION" should be printed "DOUBLE

241. See A GENTLE INTRODUCTION TO SGML: WHAT'S SPECIAL ABOUT SGML? ch. 2.1 (C.M. Sperberg-McQueen & Barnard) (noting that SGML's use of a descriptive markup system allows programmers to write programs that apply different processing instructions to identify parts of the word-processing document), available at <http://www.tei.uic.edu/orgs/tei/sgml/teip3sg/index.html> (last visited July 22, 2001).

242. See *Comp.text.sgml—Frequently Asked Questions* [hereinafter FAQ] (noting that SGML does not say how a document should be processed for printing), at http://www.pitt.edu/~dlsst51/SGML_FAQ.txt (last visited July 25, 2001).

243. See *id.* (noting that SGML and DTDs allow programs to structure information but do not say how that information can be processed for printing).

244. See *id.* (noting that programmers must still write programs that can translate the SGML document into a form that is suitable for word processing).

245. See Ed Tittel, *Write It, Format It—Once; Web Publishing Ventures Prompt Three Pioneers to Rethink How They Create and Distribute Information*, INFOWORLD, October 9, 1995, 1995 WL 10830587, at *7 (noting that SGML describes the organization of a document whereas word processing programs describe the presentation of the content of the document, among other things).

246. See *id.* (noting that word processing programs describe how the content of documents are presented).

247. See *id.* (describing the inability of word processing programs to separate the presentation description of a document from the structure and organization of the document).

248. See FAQ, *supra* note 242, at *2 (noting that SGML assists in the structure and organization of a document but does not address how that document should be printed).

SPACED,” and that any parts labeled “CITATION” should be printed in “BOLD FACE.”²⁴⁹ This work produced a style sheet standard known as “DSSSL.”²⁵⁰

Despite the enormous potential that SGML and DSSSL seemed to offer, the two systems remained highly complex—so complex that any work with SGML by itself, let alone with SGML and DSSSL combined, required a substantial amount of computing power.²⁵¹ “Substantial computing power” at that time meant multi-million dollar main frame or “mini” computers, and necessarily meant a significant commitment and investment.²⁵² Consequently, SGML did not achieve widely popular use, though it was widely known.²⁵³ When a physicist at CERN in Geneva named Tim Berners-Lee began to realize that the Internet could be used to read and display documents online, he turned to the familiar SGML concepts to develop what became known as the “Hypertext Markup Language,” or “HTML.”²⁵⁴

HTML used a limited form of SGML to mark up online documents. Berners-Lee used the SGML method of having “tags” or labels surrounded by angle brackets, combined in pairs to form opening and closing tags that surrounded a part of a text.²⁵⁵

249. See Mark Walter & Arlene Karsh, *SGML Crosses Technology—Adoption Chasm into the Bowling Alley*, THE SEYBOLD REPORT ON PUBLISHING SYSTEMS, January 29, 1996, LEXIS, News Library, Computers/Communication Archive News file, at *4 (noting that SGML supplies the structure of the document (i.e., “OPINION” and “CITATION”) whereas DSSSL—Document Style Semantics and Specification Language—supplies the formatting instructions for the document (i.e., “printed in ‘DOUBLE SPACED’” and “printed in ‘BOLD FACE’”).

250. See *id.* (noting that DSSSL “represents the generic language for writing style sheets and formatting instructions for SGML documents”).

251. See generally Robert J. Boeri & Martin Hensel, *Making the Commitment to XML*, EMEDIA PROFESSIONAL, March 1, 2000, 2000 WL 17730374, at *1 (describing SGML as huge (a 300-page specification), complicated, difficult, and uncompromising).

252. See Richard Blackwell, *University’s Supercomputer Tackles Many Applications*, THE FINANCIAL POST (TORONTO), September 7, 1985, LEXIS, News Library, Fin. Post-Computer and Communication Stories file *1, *1 (estimating that in 1985 total expenses of a supercomputer at a university were approximately \$35–40 million over a five-year period).

253. See Jason Hobby, *Emerging Skills XML*, COMMUNICATIONS WORLD, January 28, 2000, 2000 WL 12245709, at *1 (stating that, although SGML became an international standard in the mid 1980s, it was primarily used by large industries such as aircraft manufacturers and pharmaceuticals).

254. See ARBORTEXT, *SGML: Getting Started*, at 1, 15 (1995) (referring to HTML as an application of SGML and the “format of files published on the World Wide Web”), at http://www.arbortext.com/data/Getting_Started_with_SGML/getting_started_with_sgml.html.

255. See *A Small Course on HTML: The Next Generation* (stating that, in HTML, a “tag” for an element is the name of the element between the “less-than” sign (<) and “greater-than” sign (>), and the contents of the elements (i.e., the text of the paragraph) may be found between the starting and ending tags), at http://www.stack.nl/~optimusb/courses/htm14.0/top_index.html (last visited July 22, 2001).

Thus, HTML used “<P>” as an opening tag to indicate the beginning of a new paragraph and “</P>” to indicate the end of that paragraph.²⁵⁶

So an HTML document on the Web might look, in part, like this:

```
<P>Here is the beginning of a new paragraph. . . .  
And this is the last sentence in that  
paragraph.</P>
```

In a similar way, other tags were defined for headings, tables, lists, and the like. So the various headings in a document about the Galaxy might look like this:

```
<H1>The Galaxy</H1>  
<H2>The Solar System</H2>  
<P> The solar system comprises the sun and the  
planets. . . . </P>  
<H2> The Stars</H2>  
<P> The stars are also a part of our galaxy . . . .  
</P>
```

In this way, HTML became a highly truncated and limited form of SGML. It borrowed its tags from similar-looking tags in SGML-marked documents, though it offered none of what Goldfarb had originally conceived of as the SGML advantages, namely the ability of individual users to define their own “tags” and provide their own, computer-readable, definitions for them.²⁵⁷ Instead, Berners-Lee, and later various World Wide Web standards bodies,²⁵⁸ provided what definitions there were for each HTML tag and left it to those who developed Web browsers to implement those definitions.²⁵⁹ Netscape might display a

256. See Betty Harvey, *Next-Generation Internet Markup Language*, ASSOCIATION MANAGEMENT, June 1, 2001, 2001 WL 12765115, at *3 (noting that “<p>” represents a “paragraph” in HTML language).

257. See David C. Churbuck, *Document Esperanto*, FORBES, June 7, 1993, at 112 (quoting Charles Goldfarb as saying that GML—later known as SGML when it entered the public domain—“changes the paradigm” by making data paramount and allowing the owner to dictate its form).

258. See Marc Ferranti, *Style Sheets to Ensure HTML Interoperability; Consortium Calls for Standard Tags*, INFOWORLD, April 15, 1996, 1996 WL 10011050, at *1 (discussing the proposal made by the members of the World Wide Web Consortium, including Microsoft Corp., Netscape Communications Corp., America Online Inc., and CompuServe Inc., for a common set of HTML commands).

259. See Bob Metcalfe, *From the Ether: Web Father Berners-Lee Shares Next-Generation Vision of The Semantic Web*, INFOWORLD, May 24, 1999, 1999 WL 16660885, at *2 (noting that HTML has standard tags defined by international organizations for the presentation of Web pages while other languages have tags that can be defined by anyone).

heading like “<H1> . . . </H1>” in one size and type of font, while Internet Explorer might display it differently.²⁶⁰

By the mid 1990s, many people perceived the need for a more sophisticated means of marking up Web pages.²⁶¹ Among other things, HTML only provided for the delivery across the Internet of entire Web pages as a unit.²⁶² There was no standard way, for example, to have a program like a browser pull in only a small part of the information contained on another Web page for display.²⁶³ One might want to do this to extract the current temperature from another Web site’s weather page, for example, without having to download the entire page. Or one might want to extract current exchange rates for a particular foreign currency without having to download an entire page from the Federal Reserve. Not only is it inefficient to download so much text just to obtain a small fraction of it, but more significantly, it is very difficult to instruct a computer program as to how it should be able to “find” the temperature or currency rate of interest located somewhere on a page with lots of other unneeded information. Computer programs to do that kind of extraction can be, and have been, written and used, but they only work if the Web pages of interest remain in exactly the same layout, with exactly the same text and wording from day to day. If the owner of such a Web page makes any changes to the page that contains weather forecasts or currency exchange rates, the other program that reads that page has to be rewritten in order to “recognize” the newly designed page.²⁶⁴

260. In fact, HTML browsers have quite a few differences in the way they display HTML documents, a fact that has greatly complicated the development of Web pages that are intended to be used with any browser and substantially increased the cost of such development. See Ferranti, *supra* note 258, at *2 (noting that competing Web browser developers have added extensions to the basic structural tags of their own browsers that have led to incompatibilities with other browsers using HTML, and that these differences between the tags of various browsers worry Web companies which create Web documents because, the more that browsers diverge from one another, “the more difficult it is to maintain a Web site that can be read by everyone”).

261. See PR Newswire Association, *Graphic Communications Association Finally! ‘Extensible HTML’ for Web Publishing!*, PR NEWswire, February 19, 1997, LEXIS, News Library, PR Newswire—Computer and Communication Stories file, at *2 (quoting Norman Scharpf, President of the Graphic Communications Association, as saying that HTML is too simplistic to support the full function publishing on the Web as the amount of online documents continue to grow).

262. See Rajeev Kasturi, *The Tower of XML*, INTELLIGENT ENTERPRISE, October 20, 2000, available through Westlaw at 10/20/00 INTELLENTRP ERP34, at *2-3 (discussing XML’s advantages over HTML).

263. See *id.* (implying that, until XML came along, there were no mark-up languages that allowed browsers to manipulate data for presenting the data the way the user wanted).

264. See Mark McFadden, *Sticking with XML: Tying Disparate Systems Together*,

It was therefore increasingly apparent to Web researchers and standards bodies that life on the Web would be better if the different parts of a Web page were identified unambiguously as to their content, so that the temperature for Chicago would somehow be labeled in a way that a computer program could always understand—as “the temperature for Chicago”—and the exchange rate for Francs to dollars would similarly be labeled as “the exchange rate for Francs to dollars.”²⁶⁵ If Web pages contained content that was carefully marked or labeled in this way, other computer programs or browsers could download only the information they needed and, even better, could always “find” that information on a page regardless of how the page was formatted or arranged.²⁶⁶

This thinking led back to SGML in its original conception: a means of defining an unlimited number of markup tags, along with a computer-readable set of definitions and restrictions on how those tags were used in a given document.²⁶⁷ The SGML markup system offered a ready answer to the problem of identifying parts of Web pages according to content. The creator of a Web page document could create tags for “<CHICAGO-TEMPERATURE> . . . </CHICAGO-TEMPERATURE>; or for “<FRANCS-TO-DOLLARS-RATIO> . . . </FRANCS-TO-DOLLARS-RATIO>.”²⁶⁸ As long as the related information were included inside such tags, a computer program could always read the tags and find the information contained between them.²⁶⁹ Or a computer program like a Web browser could ask another program running on the originating Web site to select only the

ENT, February 23, 2000, LEXIS, News Library, Computer News file, at *1 (noting that HTML “tells a browser how to arrange text, graphics, and links on a screen” but does not have a mechanism for describing the data being presented; therefore, HTML has no way to effectively recognize the data on a Web page unless tags were added).

265. See Ben Gottesman, *Why XML Matters*, PC MAGAZINE, October 6, 1998, 1998 WL 18431544, at *1-3 (discussing the need for a language that provides a meaningful and flexible way to describe information pertaining to a document and stating that there is a need for a meaningful way to describe data because computers must be able to “speak the same language” in today’s Internet-based society).

266. *Id.* at *2 (noting that, “by providing a common language for describing data” on the Web, XML allows for more precise searching and makes data navigating much easier on the Internet).

267. Refer to notes 236–39 *supra* and accompanying text (discussing SGML’s ability to allow its users to create the parts of any document, give them any name, and to create the definitions of these names and parts).

268. See *id.*

269. Refer to notes 229–32 *supra* and accompanying text (noting SGML language relied on “markup tags,” and required a beginning tag and an ending tag that surrounded the part of the document that they labeled in order for the tags to be understood).

information of interest and send only that part of the information back across the Internet.²⁷⁰

A. XML

The original specifications for SGML seemed more complicated than necessary for these purposes, so computer engineer Jon Bosak designed a smaller, but still quite elaborate and sophisticated, subset of SGML that could be used in just this way to greatly extend the widely used HTML on the Web.²⁷¹ In addition to the familiar “<H1>” and “<P>” and similar tags, a Web page designer could add new tags, like “<CHICAGO-TEMPERATURE>.”²⁷² The notion that Web page creators could extend the use of HTML-like tags in almost any way that they chose led to the terminology of an “extensible HTML,” or simply XML, which brought the components of HTML and XML together.²⁷³

XML is thus a reprise of SGML, aided by several additional and crucial factors. XML was simpler, just enough so that it was significantly easier than SGML to understand and implement in computer programs. It came along in the context of the extraordinary explosion of World Wide Web technology and rendered that technology more useful still by an order of magnitude. And it arose at a time when computers themselves were vastly more powerful than they had been in the 1970s and hence were much better able to perform the complex processing necessary to make use of a sophisticated markup language.

270. See McFadden, *supra* note 264, at *2 (stating that, by adding tags with the related information inside the “<” and “>” signs, browsers could recognize “data components of the document” and “make it possible to manipulate or export that data as requested by the user”).

271. See *Careers; Happiness is XML-shaped*, COMPUTING, June 1, 2000, LEXIS, News Library, Computing file, at *1 (noting that Jon Bosak is credited with the original idea of XML and that XML is a “simpler dialect” of SGML that has far exceeded the abilities of HTML).

272. See Betty Harvey, *Next-Generation Internet Markup Language*, ASSOCIATION MANAGEMENT, June 1, 2001, 2001 WL 12765115, at *3-4 (noting that XML allows developers to create the grammar that describes their data and to create specific rules used for their own information).

273. See Kevin Featherly, *W3C Lays Groundwork for Device-Independent Web*, NEWSBYTES, January 27, 2000, 2000 WL 2272462, at *1 (stating that XHTML is comprised of a mixture of authoring tags from HTML and XML which will allow for more flexibility on the Web). XML is thus a step above HTML, a prior deviation from SGML. See Terry Moriarty, *The Naked Truth: With XML, The Metadata Emperor Still Has No Clothes*, INTELLIGENT ENTERPRISE, November 10, 2000, available through Westlaw at 11/10/00 INTELLENTRP 60, at *2-3 (noting that XML is an evolution from the HTML tag language because XML allows users to define their own tags corresponding to their own data).

In addition, attention soon turned again to the problem of style sheets. If XML were to be widely useful, so would a form of style sheets for XML, as DSSSL had been designed for SGML. The XML style sheet, though, should take advantage of significant advances in computer science and offer the ability not only to format with "BOLD FACE" and so on, but also to sort and rearrange text, select some text and discard other text, and so on.

At first, the XML developers incorporated the existing SGML method for creating computer-readable definitions for the XML tags.²⁷⁴ This method relies on a specialized vocabulary for writing tag definitions and specifying the relationships among the tags used in a given document.²⁷⁵ The set of such definitions is called a "Document Type Definition," or "DTD."²⁷⁶ A DTD defining the tags for a court case, for example, might specify that there is a tag (officially called an "element") that identifies each "CASE," and that each such a case must contain tags (or "elements") that identify a "CITATION," a "DATE," a "WRITTENBY" section, and an "OPINION" section, and that the "OPINION" section can contain one or more "PARAGRAPHS." Those tags, like "PARAGRAPH," that contain no further tags within them would usually consist of just "text." If that were the complete definition, the proper DTD write-up would look more or less like this:

```
<!ELEMENT CASE = (CITATION, DATE,
WRITTENBY, OPINION)>
<!ELEMENT CITATION = text>
<!ELEMENT DATE = date>
<!ELEMENT WRITTENBY = text>
<!ELEMENT OPINION = (PARAGRAPH)+>
<!ELEMENT PARAGRAPH = text>277
```

B. Schemas

Notice that the special symbols and syntax of this Document

274. See Ahmad Abualsamid, *XML: A Metastar is Born*, NETWORK COMPUTING, April 16, 2001, 2001 WL 9806562, at *5 (noting that XML used the "long-standing" SGML standard for marking up tags known as Document Type Definitions or DTDs).

275. Refer to note 239 *supra* and accompanying text (discussing how SGML allowed authors to create a set of definitions to describe different parts and names of documents and describe their relationships to each other).

276. See ARBORTEXT, *supra* note 254 (defining a Document Type Definition (DTD) as a "formal definition of the elements, structures, and rules for marking up a given type of SGML document").

277. The "plus sign" is used to mean "must appear at least one time, and may appear more times than that." See Moriarty, *supra* note 273, at *3 (stating that the "plus sign" after AUTHOR_NAME requires one or more author names). In other words, an opinion must have at least one paragraph.

Type Definition are different from the general style and syntax of HTML or XML tags themselves.²⁷⁸ At some point, those involved with XML's development thought that it made more sense to express the definition of a complex document with a language syntax that derived from XML itself, instead of using the completely different syntax represented by Document Type Definitions.²⁷⁹ For one thing, using XML syntax for defining XML documents had a certain conceptual "elegance" to it, notwithstanding that the results are quite verbose in actual appearance, as will be seen.²⁸⁰ For another thing, it had the distinct advantage that software already written to read XML syntax in a *document* would automatically be able to read the syntax of that document's *definition*—that is, what is now called the document's *schema*.²⁸¹

All that was required to do this was to come up with a slight variation on the names of the tags used to indicate the contents and relationship of the different parts of the XML document.²⁸² The trick was simply to find some tag that would identify the resulting XML file as an "XML Schema Definition," instead of an XML document. The tag eventually chosen was simply the abbreviation for "XML Schema Definition," or "xsd."²⁸³ With that preliminary tag identifying what followed as the definition of the parts of a document, rather than a document itself, it was fairly straightforward to round out the basic set of tags that would be required. One would need to define various parts of a document, which as we have seen in the SGML and XML worlds are called "elements."²⁸⁴ So an obvious tag to identify that basic part of a

278. Refer to text following note 256 *supra* (giving an example of what HTML tags look like); refer to note 285 *infra* and accompanying text (giving an example of what an XML document in XML schema format looks like).

279. Solomon H. Simon, *XML Schemas Get the Nod*, INTELLIGENT ENTERPRISE, July 23, 2001, 2001 WL 21117910, at *3 (stating that by using XML Schemas to define complex documents, one does not have to learn a completely new syntax to describe one's grammar because the schemas are already based in XML syntax).

280. Refer to note 285 *infra* and accompanying text (exemplifying a lengthy XML document defined by an XML schema).

281. See Simon, *supra* note 279, at *2 (defining "schema" as "the XML construct used to represent the data elements, attributes, and their relationships as defined in the data model" of an XML document).

282. Jim Rapoza, *A Better Way to Share Information—XML Schema Permits Rich Data Descriptions*, EWEEK, May 7, 2001, LEXIS, News Library, Computer/Communication News, Current file, at *1 (explaining that XSD is a set of shared markup vocabularies that provide a way to describe the different parts of an XML document).

283. *Id.* (explaining that the XML Schema Definition (XSD) allows developers to build a standard language for schemas).

284. Refer to notes 276–77 *supra* and accompanying text (noting that XML incorporated the SGML technique that used specialized vocabulary to identify elements, which in turn identified other parts of the document).

document would be “<xsd:element>.” And so on.

If we were to take the definition of a “case” just shown above in “DTD” format, and convert that to the format of an XML document schema, we would have something that looks like this:

```
<xsd:element name="case">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element name="citation" type="xsd:string"/>
      <xsd:element name="date" type="xsd:date"/>
      <xsd:element name="writtenby" type="xsd:string"/>
      <xsd:element name="opinion">
        <xsd:complexType>
          <xsd:sequence>
            <xsd:element name="paragraph" type="xsd:string"
              minOccurs="0" maxOccurs="unbounded"/>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>285
```

It looks wordy (it *is* wordy), but it works.²⁸⁶ And there you have it: XML schemas.

IV. ARE XML SCHEMAS COPYRIGHTABLE?

How will tomorrow’s courts deal with the issue of a schema’s copyrightability? How might they deal with it? This section starts by quickly suggesting how the analysis might go if future courts continue to use the types of analysis that they have used in the past. It then concludes by applying to the question the approach that this Article recommends instead.

A schema is designed to be used “with” or “for” something else.²⁸⁷ An XML schema enables computers to understand a

285. Refer to notes 276–77 *supra* and accompanying text (providing an example of what a proper DTD format of an XML document for a court case might look like); notes 284–87 *supra* and *infra* and accompanying text (discussing the role that XSD plays in XML schemas and providing an example of what a basic tag in XML schema format would look like).

286. Cf. Jaideep Roy & Anupama Ramanujan, *XML Schema Language: Taking XML to the Next Level*, and what that XML document would look like in XML schema format (Figure 2) (giving an example of an XML document (Figure 1)), at http://ads.computer.org/itpro/homepage/Mar_apr01/roy/index.htm (last visited Aug. 2, 2001).

287. See Simon, *supra* note 279, at *2 (defining schema as “the XML construct used to represent the data elements, attributes, and their relationships as defined in the data model” and noting that XML Schemas use XML syntax to specify metadata of a particular class of documents).

separate collection of data—an XML “database”—and thereby to make use of those data.²⁸⁸ In this way, schemas resemble many of the examples of other new subject matter that we have seen. The use of one work with another can be seen in the input data formats of *Synercom*,²⁸⁹ intended to be used with engineering calculation software,²⁹⁰ the operating system of *Apple v. Franklin*,²⁹¹ intended to be used with application programs,²⁹² the game consoles intended to be used with game units in *Sony v. Connectix*,²⁹³ and the menu command hierarchy intended to be used with spreadsheet software in *Lotus*.²⁹⁴

Differences can be identified, to be sure: with present XML technology, a collection of XML data can be used without an accompanying schema.²⁹⁵ The user of such a collection must know about the structure of the data to make use of it,²⁹⁶ but that is generally the case with collections of data today. The addition of a schema adds a new level of usability and importance to data collections by opening up many new ways for accessing the collection,²⁹⁷ but it is not essential to using such a collection in the old ways at the old levels.²⁹⁸ In many of the other cases we have examined, the new subject matter was essential for a user

288. See Roy & Ramanujan, *supra* note 286, at *3 (noting that one of the benefits of XML Schemas are that they model complex information and make it “easier for computers to process” that information).

289. *Synercom Tech., Inc. v. Univ. Computing Co.*, 462 F. Supp. 1003 (N.D. Tex. 1978).

290. See *id.* at 1006 (noting that Synercom wanted to develop “instructional manuals and input formats” that gave design engineers greater access to the engineering program).

291. *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240 (3d Cir. 1983).

292. *Id.* at 1245 (noting Franklin’s attempt to ensure 100% compatibility between its operating systems program and the “application programs created to run on the Apple computer”).

293. *Sony Computer Entm’t, Inc. v. Connectix Corp.*, 203 F.3d 596, 598 (9th Cir. 2000) (recognizing that one of the purposes of Connectix’s Virtual Game Station, a software program, was to allow computer owners to play Sony PlayStation games on their computers), *cert. denied*, 531 U.S. 871 (2000).

294. *Lotus Dev. Corp. v. Borland Int’l Inc.*, 49 F.3d 807, 810 (1st Cir. 1995) (noting that Borland copied the “words and structure of Lotus’s menu command hierarchy” into their own programs so that spreadsheet users who used Lotus 1-2-3 could now use Borland’s programs without having to learn new commands), *aff’d by an equally divided Court*, 516 U.S. 233 (1996).

295. See Simon, *supra* note 279, at *4 (noting that an XML document can still be used with a DTD or an XML schema).

296. *Id.* at *2 (noting that when a person creates a XML documents, he must follows either a DTD or XML schema format in order to use the data in the document in its intended fashion).

297. *Id.* at *4 (stating that XML shemas allow users or developers to modify, share or combine the document with other documents).

298. *Id.* (noting that an XML document can still be used with a DTD instead of an XML Schema, but the DTDs do not offer the flexibility offered by XML schemas).

to make use of some other component—some sort of input data format had to be created or the engineering analysis software would have had nothing to process.²⁹⁹ Some sort of operating system program had to be developed for the Apple computer, for example, or different application programs could not have been executed.³⁰⁰ This more “optional” nature of schemas may suggest a greater need for incentives for creation and hence a stronger argument for copyrightability.

On the other hand, it is possible that the range of choices open to the designer of a schema is narrower than with some other technologies. The *Synercom* court, for example, calculated that about three million different data input formats for engineering analysis software were possible,³⁰¹ and found that in fact, at least fifteen different ones were in actual use.³⁰² As already noted,³⁰³ the district court in the *Lotus* case similarly found that millions of different arrangements of menu command hierarchies for spreadsheet software were possible. It seems unlikely that the author of an XML schema would have anything comparable to that many expressive options, an argument cutting against the copyrightability of schemas.

No doubt other differences could be identified as well. In any event, we can start the analysis by imagining this fact pattern, extrapolated from current events. A plaintiff organization develops a schema to be used for a certain purpose or in a certain industry. Use of this schema is licensed, either for free or for a fee, but in either event, use is permitted only on certain conditions. Perhaps the license authorizes a user to obtain a copy of the schema and use it “manually” to design an XML database, or perhaps it authorizes a user to refer to the schema “mechanically,” that is, to verify with a computer that a given

299. *Synercom Tech., Inc. v. Univ. Computing Co.*, 462 F. Supp. 1003, 1006 (N.D. Tex. 1978) (noting that the input formats enabled the user to gain easier access to the computer engineering programs).

300. *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1245 (3d Cir. 1983) (noting that the operating system programs needed to be exact in order to run the application programs).

301. *Synercom*, 462 F. Supp. at 1012 (“By varying only the order constituent of the format instruction, the manner of communicating with the computer may be expressed in ten factorial (10-9-8-7-6-5-4-3-2-1); that is 3,628,800 expressions.”).

302. The right number might be thirteen or fourteen, but I read the court’s statement as saying fifteen. *See id.* at 1007 (“There are hundreds of programs available for structural analysis, and at least fifteen are competitive with STRAN. All but EDI and Synercom have different input formats.”)

303. Refer to note 122 *supra* and accompanying text (discussing the district court’s finding in *Lotus Development Corp. v. Borland International Inc.*, 831 F. Supp. 202, 215 (D. Mass. 1993), that the menu tree in Lotus 1-2-3 could be arranged in many different ways because of the different words which could be used for the spreadsheet commands).

XML database is structured in accordance with the schema's definitions. But, in any event, the use of the schema is intended to be through some sort of access "gateway" maintained by the author of the schema.³⁰⁴ Now suppose that a defendant makes use of such a schema in an unauthorized way. Perhaps it obtains a copy of the schema and seeks to redistribute it to the public. Perhaps it provides a copy of the schema for others to verify or validate their own database with reference to the schema. Perhaps it provides a service related to the processing of data that requires organizations to make reference to the schema as it is hosted by the schema's author, but in some way that violates the terms of the license.

Plaintiff schema author now brings a copyright action against this defendant. What result?

A. *Dictionary Definition Approach*

We know that some courts use a dictionary definition approach to issues of new copyrightable subject matter.³⁰⁵ Such a court would analyze the issue under section 102(b),³⁰⁶ asking whether a schema is in fact nothing more than a "system," an "idea," a "process," or a "method of operation," and so forth.³⁰⁷ The conclusion might be that an XML schema is a "system" or "process," inasmuch as it is used at times as part of a check on the compliance of a set of data with the schema's standard.³⁰⁸ Indeed, when a computer checks for that compliance, we might characterize the check as a "systematic process," lending force to the copyright conclusion that a schema is a "system" or a "process" itself and hence not copyrightable.³⁰⁹ The counter argument will be that whatever is systematic is the overall

304. One reason a schema author might want some sort of "gateway" is to extract some sort of "toll." That toll might be a direct charge for use, or it might be the collection of personal data to produce a mailing list, or it might be something else.

305. Refer to notes 26–44 *supra* and accompanying text (discussing the use by several different courts of the "dictionary definition" approach to determine whether copyright protection should be given to certain subject matter).

306. 17 U.S.C. § 102(b) (1994) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.").

307. Refer to notes 54–55 *supra* and accompanying text (noting that the First Circuit in *Lotus* inquired as to whether the subject matter was a "system" or "method of operation" because copyright protection applies to neither).

308. Refer to notes 287–88 *supra* and accompanying text (noting that XML schemas are designed to be used "with" a program and allow that program to understand and use a separate set of data in different XML documents).

309. See 17 U.S.C. § 102(b) (stating that "in no case does copyright protection . . . extend to any . . . system [or] method of operation").

process: it is a series of steps to be taken.³¹⁰ One part of those steps may include using or reading or processing the schema, but the fact that the schema is an instrument or thing used within a system or process does not make the schema itself a “system” or “process.”³¹¹

B. Merger Approach

As has often been true in new technological subject matter debates,³¹² the defendant would point to the mechanical nature of schemas—something designed to be processed by computer—as a kind of functionality that dictated or merged with any possible expressive qualities, robbing them of copyrightability. *Baker v. Selden*³¹³ will of course be cited on this point.³¹⁴ The fact that it is machine readable suggests “functional” and “merger,” doesn’t it?

Timing. The designer of a database has a range of choices in that design. If a schema is written at that same time, then the same wide range of choices is open to the writer of the schema. If the database design is taken as a given, and later a schema is written, there are still some choices open to the schema author,

310. Refer to note 311 *infra* and accompanying text (discussing how something might be just a part of a “system” while not being considered a system in and of itself).

311. See, e.g., *American Dental Ass’n v. Delta Dental Plans Ass’n*, 126 F.3d 977, 980–81 (7th Cir. 1997):

A dictionary cannot be called a “system” just because new novels are written using words, all of which appear in the dictionary. Nor is word-processing software a “system” just because it has a command structure for producing paragraphs. The Code [a taxonomy of dental procedures and the subject matter of the case] is a taxonomy, which may be put to many uses. These uses may be or include systems; the Code is not.

Id.

312. See, e.g., *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240, 1251 (3d Cir. 1983) (discussing Franklin’s defense that Apple’s operating systems are not copyrightable because the operating systems are “purely utilitarian works” and “. . . Apple is seeking to block the use of the art embodied in its operating systems”); *Music Licensing and Related Copyright Issues Before the Subcommittee on Courts and Intellectual Property of the House Committee on the Judiciary*, 105th Cong. (1997), 1997 WL 11234980, at *15 (prepared statement of Patrick Collins on behalf of SESAC, Inc.) (discussing the pre-1978 practice of not giving copyright protection to the underlying musical scores on phonograph records); *Lotus Dev. Corp. v. Borland Int’l Inc.*, 49 F.3d 807, 813–14 (1st Cir. 1995) (discussing Borland’s argument that the facts from *Baker v. Selden* are identical and, therefore, the “user interface” of Lotus 1-2-3 is not copyrightable), *aff’d by an equally divided Court*, 516 U.S. 233 (1996); *White-Smith Music Publ’g Co. v. Apollo Co.*, 209 U.S. 1, 11–12 (1908) (stating Apollo’s contention that the perforated rolls of music used in connection with the mechanical production of music such as by player pianos are not subject to copyright protection).

313. *Baker v. Selden*, 101 U.S. 99, 106 (1879) (concluding that blank account books are not copyrightable, because they are described and illustrated in a different book that is copyrightable).

314. I have long since concluded that an unwritten rule requires *Baker v. Selden* to be cited whenever a copyright case involves new technology.

but those choices are much more narrowly circumscribed because the schema must reflect the database structure.

C. Standards and Compatibility Approach

We know that some courts favor promoting the use of *de facto* standards.³¹⁵ If our hypothetical schema is widely used, it may well become such a standard. Indeed, an argument about standards in an XML schema case would likely have more force than the same argument in *Lotus*-type cases. Recall that in the *Lotus* case, the plaintiff's product was produced for individual sales. As far as we can tell from the court's opinion, Lotus 1-2-3 was not expressly promoted or touted as being an industry "standard."³¹⁶ An author of the XML schema might well be expressly promoting or touting the schema for just such standardization purposes. Industry adoption as a standard is, in large part, what gives a schema its value because third parties can more easily exchange or access data that is formatted in accordance with a commonly-used standard schema.³¹⁷

A useful example for thinking about the standards issue appears in the schemas that are currently under development for use in the legal system.³¹⁸ Electronic filing of court documents is one area in which considerable work has been underway for some time.³¹⁹ A schema for such documents would be of greatest use if it were adopted as an official standard by a given jurisdiction, and would be of even greater use if adopted by, say, all state courts or all federal courts or both.³²⁰ Legal schemas, thus

315. Refer to Part II.A.2 *supra* (discussing the use by several courts of *de facto* standards or standards of practice as a way of finding certain subject matter not copyrightable).

316. *Lotus v. Borland*, 49 F.3d 807, 809–810 (1st Cir. 1995) (stating that Lotus 1-2-3, like many other computer programs, allows users to write "macros," and that by using Lotus 1-2-3, users can shorten the time needed to operate certain programs), *aff'd by an equally divided Court*, 516 U.S. 233 (1996).

317. See Rapoza, *supra* note 282, at *1 (stating that XML schemas, a new standard adopted by the World Wide Web Consortium, make it easier for business partners to exchange and integrate information).

318. William Skeels IV, *What Can Lawyers Expect as the Internet Matures?*, LEGAL TECHNOLOGY PRODUCT NEWS, December 2000, available through Westlaw at 1200 LTPN 128, at *3 (discussing a group called LegalXML's work on creating an XML-based standard that describes legal documents and other information to help facilitate electronic filing in courts).

319. See Jody Rolnick, *Eugene, Ore., Company Seeks to Shift Courts, Lawyers into Online Legal Filing*, KNIGHT-RIDDER TRIBUNE BUSINESS NEWS: THE REGISTER GUARD, June 18, 2001, 2001 WL 23527262, at *3 (describing endeavors by numerous companies to enter into contracts with different court systems to provide e-filing systems and e-storage service for court documents).

320. Refer to note 317 *supra* and accompanying text (stating that an industry's adoption of a certain schema gives it value because the schema then may become a

officially adopted, would then resemble the electrical codes or building codes already discussed.³²¹

A court inclined to give weight to the standards issue would be strongly inclined to find that a legal schema, adopted into law (perhaps as an administrative rule of a court), would not be copyrightable.³²² Whatever arguments are persuasive in connection with the *desire* of users to be able to use “standard” spreadsheet menu commands, to pick one example, would certainly apply *a fortiori* to a court requirement that attorneys *must* use a particular standard for the filing of documents.

D. *This Article’s Approach*

This Article recommends that new subject matter cases be analyzed with regard to: (1) whether incentives would make a significant difference in the quantity of works of that subject matter that are produced and disseminated; (2) whether a range of different quality levels is possible for that type of work; and (3) whether potential customers of the work would be likely to care about those differences. Applied to XML schemas, this approach asks: (1) whether the incentive of copyright would likely be significant for the creation and dissemination of such schemas; (2) whether a range of different quality levels exist for schemas (beyond what is good enough to bring an associated product like an XML database or document to market); and (3) whether consumers would be likely to care about those differences.

The strongest argument for copyright’s incentive to make a difference in the quantity of XML schemas produced is that such schemas are not essential to the use of an XML database.³²³ As noted, schemas offer a new level of function and the potential for much wider access,³²⁴ but they are not essential to the use of a database, XML or otherwise.³²⁵ Because XML schemas are new

commonly-used standard; once a standard, the schema then facilitates communication between its users because people in that particular field become familiar with using that schema).

321. Refer to notes 61–62 *supra* and accompanying text (noting that courts using the standards approach will not find a new work copyrightable if it has been adopted as a “standard” because it would be unfair to the competition and consumers).

322. Refer to note 62 *supra* and accompanying text (explaining that one reason why courts use the *de facto* standards approach is to prevent consumers from having to retain or relearn how to use new products).

323. See Simon, *supra* note 279, at *4 (noting that an XML schema is not necessary to use an XML document; a DTD may be an acceptable alternative to a schema).

324. Refer to notes 282–83 *supra* and accompanying text (discussing the advantages to using XML schemas in an XML database).

325. See Simon, *supra* note 279, at *4 (noting that users can also use DTDs to use an XML document instead of an XML schema).

in the past few years,³²⁶ databases created and used before that time—and there must be hundreds of thousands, if not millions³²⁷—would not have an XML schema associated with them, and yet are, by hypothesis, in productive use already. From that perspective, perhaps a strong incentive would be needed to bring forth very many schemas at all.

But that possibility seems unlikely. If XML schemas were actually useful, actually provided a greater level of functionality than a database without such a schema, one would imagine that that functionality alone would bring forth a great many schemas, whether copyrightable or not. In that case, perhaps no extra copyright incentive would be needed.³²⁸ One additional fact would be important here, however, and that is the likelihood of public disclosure and dissemination. If disclosure or distribution has value, then we should care not only about schema creation, but about schema disclosure and distribution. It might turn out that private organizations have the strong incentive of added functionality to develop XML schemas for internal use; but that lacking copyright protection, they might choose not to disclose such schemas to the public generally, relying instead on trade secret law and actual secrecy to prevent circulation or disclosure.³²⁹ It would be quite possible to use a schema internally and not make it known publicly—in that way, XML schemas are different from, say, the Lotus 1-2-3 user interface, which must be disclosed to the public in order to sell the Lotus 1-2-3 program.³³⁰ For that matter, every database necessarily has

326. *Id.* at *1 (noting that XML Schema only recently attained final recommendation status from the W3C and are now beginning to gain acceptance from companies).

327. *There are Free Databases on the Web—But Where?*, THE INFORMATION ADVISOR, June 1, 2001, available at 2001 WL 13572678, at *1 (noting that the Gale Directory of Databases provides information on 130,000 public databases but provides information on only “a few hundred Internet databases,” and is by no means a comprehensive compilation of all the databases available on the Internet today).

328. Refer to note 164 *supra* and accompanying text (noting that if a certain product would have been created with or without copyright protection, the copyright protection does not provide an incentive for creation of that particular product; instead, copyright protection does nothing more than prevent competition and keep prices high for consumers).

329. *Cf.* GREGORY M. WASSON, MISAPPROPRIATION OF TRADE SECRET UNDER UNIFORM TRADE SECRETS ACT § 1(A)(2) (1991), 12 AM. JUR. PROOF OF FACTS 3d 711 (discussing the different views on the purpose of trade secret law, and noting that while some believe the purpose of trade secret law is to encourage invention, others believe that trade secret law is an attempt to “enforce some degree of commercial morality,” and still others feel that the purpose of trade secret law is to protect people from another’s breach of faith and to keep others from taking one’s secrets by immoral means).

330. See Simon, *supra* 279, at *4 (noting that an XML schema is not necessary to use an XML document); Lotus Dev. Corp. v. Borland Int’l Inc., 49 F.3d 807, 809 (1st Cir. 1995) (stating that a user needs the menu command hierarchy in Lotus 1-2-3 in order to use the

some sort of organization and arrangement, which amounts to an informal “schema.” How many of us in the public are aware of these informal “schemas”? Are we aware of the organization and structure of the databases that we face daily, from Web site logs to credit card accounts to mortgage payment schedules? Converting these internal, informal schemas to formal XML schemas would not magically bring them to the public’s attention. Providing them with copyright protection would at least permit and even encourage organizations to disclose them to the public.³³¹ The application of copyright protection would also likely pre-empt trade secret protection and in that way make public disclosure more likely as well.³³²

Other forces are at work that would provide an incentive to the disclosure of many XML schemas besides copyright, however. Although such schemas can certainly be used internally by a single organization, one of the touted virtues of schemas is that they make it much easier for companies to deal electronically with each other.³³³ One likely source of schema development is therefore trade associations. An association-developed schema could be shared among all member companies precisely for the purpose of those companies exchanging data (exchanging current inventory data with suppliers to facilitate automatic just-in-time delivery by those suppliers, for example). Perhaps trade association-developed schemas, shared among many companies, would in that way come to the attention of the public.

Even in this instance, though, it is not obvious that the members of a trade association have any incentive to disclose such schemas outside the association. Why should a potentially valuable XML schema, the development of which was funded by dues-paying members, be given away to others who do not similarly pay dues? Again, if these schemas have value to the public, then extending copyright protection to them might be useful to provide an incentive for groups like trade associations to let the schemas come to the public’s attention.³³⁴

data and information in the program), *aff’d by an equally divided Court*, 516 U.S. 233 (1996).

331. See 17 U.S.C. § 102(a) (1994) (stating that copyright protection prevents the reproduction of original works).

332. Margaret H. Kavalaris, *INTELLECTUAL PROPERTY PROTECTION BASICS* § 3(C)(1)(a) (1995), 403 PLI/PAT 143, 157 (stating that federal copyright laws still preempt state trade secret laws).

333. See Rapoza, *supra* note 282, at *1 (stating that XML schemas make it easier for businesses to exchange and integrate information).

334. Refer to notes 163–65 *supra* and accompanying text (postulating that copyright law should provide incentives to encourage public disclosure of the creative work and discourage companies and businesses from keeping the product to themselves).

Finally, in the future, we may see businesses formed on the basis of offering a widely accessible database for a fee. Such businesses could rely on a publicly-accessible XML schema to facilitate the public's access to their databases. Businesses like this, unlike those that use schemas only internally or only within the members of a trade association or similar group, would therefore have an incentive both to create a schema in the first place, and also to make it publicly known and available. In this latter case, the schema would be much like the menu command hierarchy in *Lotus*—necessarily created, and “good enough,” in order to be packaged and sold along with the access to the underlying database.³³⁵ Yet, as might be true with a menu command hierarchy, a range of different quality levels might still be possible for such a schema, beyond whatever quality level proved “good enough” to justify bringing the underlying database to market at all. If that were true, then the application of copyright would make a difference to that quality level: the more easily a developer can capture the benefits of greater investment in a schema's quality by protecting it with copyright, the more likely the developer is to make that greater investment and to produce a schema of that higher quality.

Would consumers care about the different quality levels? One can imagine any result here. Like many readers, I have had at least some direct, hands-on experience with menu command hierarchies, buying application programs to run on given hardware and software platforms, and using game programs and consoles. I have not had comparable experience with the use of publicly available XML schemas and consequently lack that same experiential intuition, so my speculations are even more speculative.

One can envision a world in which XML schema designers debate passionately the virtues of using one type of approach to schema design over another, or extol the elegance of one syntactical formulation over another, and that no one else would care a wit about these debates or instances of elegance. If that were true, then different quality levels would exist for schemas, but they would be of little interest or use to consumers of those schemas. In that case, copyright protection would be superfluous.³³⁶

335. *Lotus*, 49 F.3d at 809 (stating that a user would not be able to use Lotus 1-2-3 without the menu command hierarchy that accompanies the program), *aff'd by an equally divided Court*, 516 U.S. 233 (1996).

336. Refer to note 169 *supra* and accompanying text (noting that if consumers did not care about the differences in the quality of the various products, then copyright protection is irrelevant).

One can also envision a market for schemas in which the consumers are businesses, universities, or other organizations. It might turn out that these consumers are keenly aware that some schemas are faster for computers to process than others, some are easier to understand by humans as well as computers than others, some are easier to combine with different types of databases or software than other schemas, some are easier to transport to different platforms than others, or in general, that XML schemas—even those that describe the same underlying database—might vary from each other in qualitatively important ways even if individual users were unaware of those ways. If that proved true, then the argument for treating XML schemas as copyrightable works of authorship would be strong indeed.³³⁷ We would want schema developers to be able to capture the full value of their efforts at greater quality so that they would invest at the proper level to achieve that quality. Copyright protection is a useful and appropriate way to accomplish just that result.³³⁸

Finally, one can also envision a world in which all the above scenarios are true. Schema designers might debate commercially unimportant differences, even while they pay great attention to other differences that do matter to schema consumers. If any of the various possibilities I have sketched out seems intuitively likely to me, it is this latter one—that all possibilities are likely!

In that case, the argument for copyright protection is very strong.³³⁹ Indeed, if the factual findings supported the conclusion that a range of different quality levels is possible for XML schemas, and that the relevant consumers of those schemas would care about those differences, then the analysis of this Article shows that such schemas should very much be copyrightable.

V. CONCLUSION

In 1976, Congress clarified that copyright's subject matter did not comprise tangible objects, but rather the abstract category of "works of authorship," without regard to any particularly medium of fixation. But Congress did not clarify

337. Refer to notes 166–68 *supra* and accompanying text (stating that the existence of significant differences between the quality of products is a strong factor in determining the strength of an incentive for copyright protection for the superior product).

338. Refer to notes 166–68 *supra* and accompanying text (noting that offering copyright protection to products beyond the "good enough" level is a way of encouraging businesses to produce a product of "higher" quality).

339. Refer to note 169 *supra* and accompanying text (discussing three factors which, if established, provide a strong argument for copyright protection).

what exactly is a “work of authorship” in the first place. Consequently (and perhaps inevitably), issues continue to arise from time to time about what new forms of expression should or should not be considered copyrightable “works of authorship.” In recent decades this question has arisen regarding computer software menu command hierarchies, data input formats on “punch cards,” computer operating system software, case reporter pagination, and in other contexts.

Courts seem to use very informal analyses in these cases, at times asking about the definition of a statutory term without regard to any underlying policy; at times asking whether a merger of idea and expression has taken place; at other times suggesting that essentially antitrust concerns such as the avoidance of monopoly or the need to foster competition play an important role in making decisions about new subject matter.

What is surprising in many of these cases is the lack of consideration of one of the most fundamental policies of copyright law, the establishment of incentives to creation. This Article has reviewed several illustrative “new subject matter” cases identify the types of analyses used. This Article then proposed a different approach to such cases, one more firmly grounded in the notion of copyright law as an incentive. Close attention to incentives leads to the proposal of a three-factor analysis. This analysis asks whether the creation of a copyright right in some new subject matter would substantially increase the quantity of the works produced; whether a range of varying quality levels for such works is possible; and whether consumers of the work would care about such qualitative differences. The answers to these questions turn largely on factual findings. In regard to any new subject matter at issue, if the findings are positive—that is, if the application of copyright would increase the quantity of works produced; if a range of different quality levels is possible for the type of work; and if consumers would likely care about those quality differences—then courts should determine that the subject matter at issue is copyrightable.

To anchor this proposed new analysis, this Article has reassessed the illustrative cases already discussed in light of the new analysis. Many of these previous cases were in fact decided “correctly” by the terms of the analysis, but one prominent case, the *Lotus v. Borland*, was not. This Article has tried to understand the way that future courts might deal with a particular form of new subject matter that has not, at this point and to the author’s knowledge, actually been litigated but seems likely to be litigated in the not-too-distant future: the subject matter known as “XML schemas.” It now concludes with an

acknowledgment that the issue turns on factual findings, but that such findings will very likely show that XML schemas should be considered an appropriate subject matter for copyright's protection.