

Emergent Works*

Bruce E. Boyden**

INTRODUCTION

Nineteen sixty-four was a very important year for the Copyright Office. The copyright revision effort that would eventually become the Copyright Act of 1976 was in full swing; draft bills that largely resembled the final Act were introduced in July.¹ The goal of this effort was to update the old 1909 Act to fully account for the incredible proliferation of mass media. Just as that effort was shifting to Congress, however, over at the Copyright Office, the first harbinger of a new set of problems was sounding: the Office issued its first registrations for computer programs.²

Almost immediately, there were difficulties. Programs were submitted in various formats, some in printed source code, some on punch cards; one program was submitted only on magnetic tape, its length so great that if it had been printed out the paper would have stacked twelve inches high.³ Amidst these early difficulties, Copyright Office personnel noted another oddity as well: some applicants were submitting, not programs themselves, but works *created* by a computer program. The works submitted for registration included “a musical composition created by computer, . . . an abstract drawing, and . . . compilations of various kinds, which were at least partly the ‘work’ of computers.”⁴ In his annual report to Congress, Register Abraham Kaminstein observed that these submissions raised “difficult questions of authorship,” namely whether “the ‘work’ is basically one of human authorship, with the computer merely being an assisting instrument, or whether the traditional elements of authorship in the work . . . were actually conceived and executed not by man but by a machine.”⁵

The question has puzzled copyright scholars for decades. A number of law

* This Essay is based on a talk that was given on October 2, 2015, at the Kernochan Center Annual Symposium at Columbia Law School.

** Associate Professor, Marquette University Law School.

1. *Compare* Copyright Revision Bill, H.R. 11947, 88th Cong. (1964), *and* Copyright Revision Bill, S. 3008, 88th Cong. (1964), *with* Copyright Act of 1976, Pub. L. No. 94-553, 90 Stat. 2600. *See also* Abraham L. Kaminstein, SIXTY-EIGHTH ANNUAL REPORT OF THE REGISTER OF COPYRIGHTS 1 (1966), <http://www.copyright.gov/reports/annual/archive/ar-1965.pdf> [<https://perma.cc/G6YE-SP8U>].

2. Abraham L. Kaminstein, SIXTY-SEVENTH ANNUAL REPORT OF THE REGISTER OF COPYRIGHTS 4 (1965), <http://www.copyright.gov/reports/annual/archive/ar-1964.pdf> [<https://perma.cc/WN7Y-Y2DA>].

3. Kaminstein, *supra* note 1, at 3.

4. Kaminstein, *supra* note 2, at 5.

5. *Id.*

review articles have been written on the issue, beginning only a few years after Kaminstein's reports. Both the Commission on New Technological Uses of Copyrighted Works (CONTU) and the Office of Technological Assessment have addressed the matter. The lure of the question has proven almost irresistible, like a sort of copyright koan, through which a student of copyright law might achieve enlightenment: can a work be authored by something that is not an author? But while serving as an intellectual exercise, the problem of computer-generated works has never seemed very pressing. As the CONTU Report concluded in 1978, given the state of technology at the time, the issue was "too speculative to consider."⁶

Until now, that is. We are on the cusp of a very significant change. Programs can now generate music that is commercially viable as background music, or write poetry that is difficult to distinguish from that written by humans.⁷ They can write simple news stories, such as breaking news about earthquakes or sports.⁸ They can generate personalized reports from databases, which is essentially what a Google search result is. Or they can create automated videos, as Facebook has done for its users occasionally.⁹ And they are doing so for financial gain.

What is perplexing about these situations is that they seem to be instances in which we have something that looks like a "work," but there may be no person whose actions resemble those of a traditional "author." That is, computer-generated works pose a problem for what might be called the "standard model" of copyright law, under which a person, the author, produces a work that is then conveyed to the audience through some sort of medium such as a book. With computer-generated works, the production and conveyance steps become intertwined, with one work—the computer program—producing another work each time it is used. Those subsequent works may contain creative elements that were not present in the first, and thus lack an easily identifiable human origin. When it comes to what might be "emergent works"—works that consist largely of creative elements that have emerged unbidden from the operation of the program—who should be considered the author?

This is a question that pushes the edge of the envelope of established doctrine.

6. NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS, FINAL REPORT 44 (1978), <http://people.ischool.berkeley.edu/~bearver/mediawiki/images/8/89/CONTU.pdf> [<https://perma.cc/C7DQ-XJQ7>] [hereinafter CONTU Report]. Arthur Miller endorsed the same conclusion in 1993. See Arthur R. Miller, *Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU*, 106 HARV. L. REV. 977, 1049 (1993).

7. See JUKEDECK, <https://www.jukedeck.com/> (last visited Feb. 6, 2016); "Bot or Not?", BOTPOET, <http://botpoet.com/> (last visited Oct. 2, 2015).

8. See Steve Lohr, *In Case You Wondered, a Real Human Wrote This Column*, N.Y. TIMES (Sept. 11, 2011), <http://www.nytimes.com/2011/09/11/business/computer-generated-articles-are-gaining-traction.html> [<https://perma.cc/HL9B-5D6C>]; Gregory Ferenstein, *An Algorithm Wrote the LA Times Story About the City's Earthquake Aftershock Today*, TECHCRUNCH (Mar. 17, 2014), <http://techcrunch.com/2014/03/17/an-algorithm-wrote-the-la-times-story-about-the-citys-earthquake-aftershock-today/> [<https://perma.cc/S68K-SFH9>]; Matt Carlson, *The Robotic Reporter*, 3 DIGITAL JOURNALISM 416 (2015).

9. See Angela Moscaritolo, *Facebook "Year in Review" Will Omit Bad Memories*, PC MAGAZINE (Dec. 17, 2015), <http://www.pcmag.com/article2/0,2817,2496767,00.asp> [<https://perma.cc/2MWD-AZSG>].

Copyright law has rarely been called upon to determine the authorship rights underlying material that is only apparently the product of human creativity. Judge Jerome Frank long ago wrote that even involuntary creations may be protected by copyright, suggesting that a painter's inadvertent markings, caused by "bad eyesight or defective musculature, or a shock caused by a clap of thunder," may produce material that "the 'author' may adopt . . . as his and copyright it."¹⁰ But if such a situation has ever arisen, it has been so infrequent that there has been little felt need to carefully delineate the boundaries of authorship. That is likely to change, however, as computer programs become not only more adept at simulating human creativity, but more importantly, commercially successful at doing so.¹¹

In this Essay I draw three conclusions. First, although there have long been difficulties in copyright law in identifying authors and authorship, computer-generated works represent a novel problem, one that, in the fashion of most Internet law conundrums, fails to fit well in existing doctrinal categories. That problem is also growing in practical significance as computers are increasingly able to create works unassisted by humans.

Second, the problem of computer-generated works is not a single problem, but rather a set of related problems, some of which are easier than others to resolve. The most difficult involve what might be called "emergent works"—works of apparently creative expression that arise from the operation of a program but cannot be traced directly to a human source.

Third, consideration of the class of computer-generated works indicates a potential criterion for whether a person should be considered the author of a given work: whether that person could predict the work's content with reasonable specificity before it is rendered or received by the user. That indicates a possible test that might be used to determine the authorship of computer-generated works, one akin to elements of a negligence claim in tort law: whether the output of the program was reasonably foreseeable to the putative author. Such an inquiry essentially asks of the would-be author that he or she establish what copyright law has demanded of authors for more than a century, that the work was the product of his or her imagination and conception of it.

I. MAPPING THE PROBLEM OF COMPUTER-GENERATED WORKS

Authorship lies at the heart of modern copyright law. The Constitution gives Congress the power to establish a particular sort of copyright system: one that grants "Authors" the exclusive right to their "Writings."¹² The Copyright Act of 1976, like the copyright acts that preceded it, implements the constitutional grant by providing exclusive rights only to works created by authors.¹³ In order for a

10. *Alfred Bell & Co. v. Catalda Fine Arts, Inc.*, 191 F.2d 99, 105 (2d Cir. 1951).

11. Jukedeck's service providing video background music, for example, charges \$22 per song to license use by large businesses, and \$199 to own any copyright outright. See "Pricing", JUKEDECK, <https://www.jukedeck.com/pricing> [<https://perma.cc/6BAU-XZ2S>] (last visited Feb. 6, 2016).

12. U.S. CONST. art. I, sec. 8, cl. 8.

13. Section 102(a) provides that copyright inheres only in "works of authorship," and § 201(a)

work to receive copyright protection, there must be a person or entity that qualifies as an “author.”

Despite its centrality, however, authorship has never been formally defined. It has instead been assumed that anyone who creates a sufficiently original work, one that meets the minimal threshold for creativity under the Copyright Act, is necessarily an “author” under the statute. For more than a century, the threshold for creativity has purposefully been kept at a minimal level, following Justice Holmes’s concern that federal judges were competent at assessing neither popular culture nor avant-garde art.¹⁴ Although courts must still determine which parts of a work may contain creative expression, for purposes of determining copyrightability, or assessing infringement, or deciding between competing claims to have authored a work, that inquiry has typically been carried out not by assessing the creativity of the material itself, but rather by considering the actions of the putative author—did he or she make nonobvious choices, and exercise judgment, in constructing the work?¹⁵ If so, the court merely assumes that those choices and judgment calls—the author’s process—led to something creative in the work. As Justice Holmes concluded as far back as 1903, “[p]ersonality always contains something unique. It expresses its singularity even in handwriting, and a very modest grade of art has in it something irreducible, which is one man’s alone.”¹⁶ Whatever creative expression might be, if the author made the right type of choices, it is assumed to be present in the work somewhere.

II. ALGORITHMS AND COPYRIGHT’S STANDARD MODEL

Computer-generated works destabilize copyright law’s approach to authorship by obscuring the connection between the creative process and the work. Once that happens, it will no longer be possible to simply assume that all minimally creative elements stemmed from the mind of one or more human authors. Consider a newspaper-story-writing algorithm. The story as a whole may appear to be the product of human ingenuity. But how much of the story was actually written by a person can depend on how the algorithm operates.

The earliest story-generating algorithms simply combined a template with a continuously monitored updated information source. So, for example, an algorithm like the *Los Angeles Times*’s “Quakebot” might be programmed to begin, “A [#.#]-magnitude earthquake struck [REGION] at [TIME] today, causing [severe/widespread/little/no reported] damage.”¹⁷ Most of the story has been written in advance by human authors, with only isolated blanks to be filled in,

provides that such copyrights are owned initially by “the author or authors of the work.”

14. See *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239, 251–52 (1903).

15. See, e.g., *CCC Info. Servs. v. MacLean Hunter Mkt. Reports*, 44 F.3d 61 (2d Cir. 1994); *Kregos v. Associated Press*, 937 F.2d 700 (2d Cir. 1991); *Childress v. Taylor*, 945 F.2d 500 (2d Cir. 1991).

16. *Bleistein*, 188 U.S. at 250. Ed Lee has referred to this as copyright’s “happy doctrinal détente.” Edward Lee, *Digital Originality*, 14 VAND. J. ENT. & TECH. L. 919, 920 (2012).

17. See *Robot Writes LA Times Earthquake Breaking News Article*, BBC NEWS (Mar. 18, 2014), <http://www.bbc.com/news/technology-26614051> [<https://perma.cc/3375-LD9G>].

“mad-lib” style, by the algorithm when it detects a report from its information source that fits its criteria.¹⁸ But story-generating algorithms have become far more sophisticated in just a few years. The news stories produced by firms such as Narrative Science and Automated Insights still use canned phrases and structural templates, but the templates are more complex than simply filling in the blanks, resulting in news stories that are difficult to tell from those written, albeit under deadline, by humans.¹⁹ Eventually news algorithms will construct their stories essentially the same way humans do: by selecting from a large number of phrases, idioms, paragraph structures, and story formats based on the underlying facts.²⁰

Such programs pose a problem for copyright’s standard model of authorship. As typically envisioned, copyright protection can be divided into two phases, a creation phase and a communication phase. During the creation phase the author, often conceived of as a single person such as a novelist,²¹ produces a finished work that is fully captured in some sort of medium—books, sheet music, scripts, optical disks, electronic files, or the like. During the communication phase, the finished product is then conveyed to an audience. The distinction between the two phases is important, because copyright protection arises with the first stage, and applies to the second: once an author has created a fixed copyrightable work, copyright gives exclusive rights over the means of conveying that work—physical media,²² performances to an audience,²³ or transmissions of its images.²⁴ The idea is that the copyright owner gets compensated for having produced the work in the first stage out of the license fees for communicating the work to a public audience in the second.

It is of course true, as James Grimmelman argues elsewhere in this issue, that

18. This is how the story-generation algorithms appear to work that are described in Erick Schonfeld, *Automated News Comes To Sports Coverage Via StatSheet*, TECHCRUNCH (Nov. 12, 2010), <http://techcrunch.com/2010/11/12/automated-news-sports-statsheet/> [https://perma.cc/SZZ3-MAB2]; Kevin Roose, *I Just Wrote 7 Blog Posts in Less Than 3 Seconds. Here’s the Secret*, FUSION (Oct. 21, 2015), <http://fusion.net/story/217513/bot-writer/> [https://perma.cc/A9J8-4TGF].

19. Samantha Goldberg, *Robot Writers and the Digital Age*, AM. JOURNALISM REV. (Nov. 25, 2013), <http://ajr.org/2013/11/25/computer-might-replace-robot-journalism-digital-age/> [https://perma.cc/MJ3E-V9T9]; Celeste Lecompte, *Automation in the Newsroom*, NIEMAN REPORTS (Sept. 1, 2015), <http://niemanreports.org/articles/automation-in-the-newsroom/> [https://perma.cc/8PRQ-7KE8]; Steven Levy, *Can an Algorithm Write a Better News Story Than a Human Reporter?*, WIRED (Apr. 24, 2012), <http://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/> [https://perma.cc/7SKY-2B25]; Shelley Podolny, *If an Algorithm Wrote This, How Would You Even Know?*, N.Y. TIMES (Mar. 8, 2015), http://www.nytimes.com/2015/03/08/opinion/sunday/if-an-algorithm-wrote-this-how-would-you-even-know.html?_r=0 [https://perma.cc/PX7Q-NHK3].

20. While news algorithms may improve in their writing ability, they will still be reliant on human reporting for everything but the most basic stories. See Ian Crouch, *The Sportswriting Machine*, NEW YORKER (Mar. 26, 2015), <http://www.newyorker.com/news/sporting-scene/the-sportswriting-machine> [https://perma.cc/D6XC-3E5E].

21. Obviously this is a highly stylized portrait of even most novelists.

22. The reproduction right, derivative works right, and distribution right prohibit certain means of conveying the work through the creation or distribution of physical media, including electronic storage. See 17 U.S.C. § 106(1)–(3).

23. See 17 U.S.C. § 106(4), (6).

24. The right of public display, 17 U.S.C. § 106(5), once the “first sale” right in § 109(c) is removed, essentially protects against remote viewing of a fixed image.

the standard model has always had complications, both in terms of how works are produced and how they are conveyed.²⁵ Some of those complications have been quite challenging, but none disrupt the fundamental distinction between production and conveyance. For example, in contrast to the archetype of a single author drafting a work *ex nihilo*, it is often the case that authors work in groups or even large industrial organizations dedicated to the output of cultural or informational products.²⁶ That has led to difficulties in identifying who among many collaborators may be an author.²⁷ Similarly, authors are heavily influenced by their cultural surroundings, and as a result an author's creative impact does not end once the finished product is disseminated. Works beget other works, and it can be difficult to determine where influence ends and appropriation begins.²⁸ Neither of these problems, however, challenges the boundary between production and communication. One set of conflicts concerns whose contributions count; the other concerns whether downstream uses of the finished work constitute infringement.

With respect to conveyance, there are other complications. For example, it has long been the case that the delivery of some works from authors to audience occurs through communication to some sort of intermediary, either a device or human performers. Indeed, works intended for performance are not usually even intended to be seen by the ultimate audience. Rather, they consist of a set of instructions to the performers, such as sheet music or the script of a play; the ultimate audience only indirectly perceives the work through the performance. This complicates the distinction between work and conveyance, because a human performer in rendering a work will inevitably add his or her own interpretation to it; a work intended for human performance is never transparently conveyed.²⁹ Although the authorship inherent in performance leads to the occasional conundrum, by and large conflicts over performance are infrequent enough that the system has been able to ignore them.³⁰

25. James Grimmelmann, *There's No Such Thing as a Computer-Authored Work—And It's a Good Thing, Too*, 39 COLUM. J.L. & ARTS 403 (2016).

26. The "standard model" is based in part on what has been called the "Romantic notion" of authorship, that is, the image of a single creative genius toiling away in isolation to produce some work. As critics have observed, the portrayal of authorship as part of an industrial process is far less attractive as lobbying fodder, but such processes produce much of what is actually protected by copyright. See, e.g., Peter Jaszi, *Toward a Theory of Copyright: The Metamorphoses of "Authorship,"* 1991 DUKE L.J. 455 (1991).

27. See *Childress v. Taylor*, 945 F.2d 500 (2d Cir. 1991); *Thomson v. Larson*, 147 F.3d 195 (2d Cir. 1998); *Aalmuhammed v. Lee*, 202 F.3d 1227 (9th Cir. 2000).

28. See *Castle Rock Enmt., Inc. v. Carol Publ'g Group, Inc.*, 150 F.3d 132 (2d Cir. 1998); *Three Boys Music Corp. v. Bolton*, 212 F.3d 477 (9th Cir. 2000); *Cariou v. Prince*, 714 F.3d 694 (2d Cir. 2013).

29. See *Garcia v. Google, Inc.*, 743 F.3d 1258 (9th Cir. 2015); *Erickson v. Trinity Theatre, Inc.*, 13 F.3d 1061 (7th Cir. 1994). Indeed, in some works, such as jazz compositions, the work is written with the understanding that some amount of it will be filled in by the performers later. Such compositions are a special case of sequential authorship resembling those on the right side of Figure 1 below.

30. This is not to deny that there are serious tensions between assigning a low threshold for creativity and denying the authorship of performers, even if that conflict has been papered over. See generally Rebecca Tushnet, *Performance Anxiety: Copyright Embodied and Disembodied*, 60 J.

III. SPECTRUMS OF AUTHORSHIP

Computer-generated works—works produced with the aid of a computer—are potentially different. With a computer-generated work, the finished work—the one displayed on the user’s screen or heard on his or her speakers—is not even presented until there is some amount of input from the user. That input can be greater or lesser, but the intervention of the algorithm before the work as displayed is even finalized is a novel situation. It is particularly perplexing for copyright purposes if seemingly creative portions of the work emerge for the first time through the operation of the program. Who should be considered to be the author of those portions?

Commentators have considered several possibilities. The software programmer who wrote the program that produced the display might be the author;³¹ the user of the computer might be the author; the two might be joint authors,³² or neither might be the author.³³ Most fancifully, the program itself might be considered the author.³⁴ Prior scholarship examining this issue has tended to propose a single solution to the problem. For example, the CONTU Commission declared the “obvious answer” to be that the “one who employs the computer”—i.e., the user—authors the works that result.³⁵ Annemarie Bridy largely agrees, viewing a computer program as a sort of agent of the user, whose works might be thought to be produced by the user as a work made for hire.³⁶ Other scholars, however, have argued that the author of the software should be held to be the source of any content it generates, or that the Constitution’s requirement of authorship rules out protection of non-human creativity.³⁷

The circumstances under which computer-generated works are prepared seem too varied to permit a single solution, however.³⁸ For example, assigning initial

COPYRIGHT SOC’Y U.S.A. 209 (2013).

31. See Evan H. Farr, *Copyrightability of Computer-Created Works*, 15 RUTGERS COMP. & TECH. L.J. 63, 79–80 (1989).

32. See W. Joss Nichols, *Painting Through Pixels: The Case for a Copyright in Videogame Play*, 30 COLUM. J.L. & ARTS 101, 122 (2007).

33. See Ralph D. Clifford, *Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand Up?*, 71 TUL. L. REV. 1675 (1997).

34. See Karl F. Milde, Jr., *Can a Computer Be an “Author” or an “Inventor”?*, 51 J. PAT. OFF. SOC’Y 378, 393 (1969) (suggesting that program could at least be considered a co-author).

35. CONTU Report, *supra* note 6, at 45. The Report continues: “The simplicity of this response may obscure some problems, though essentially they are the same sort of problems encountered in connection with works produced in other ways.” *Id.* See also Grimmelmann, *supra* note 25 (reaching the same conclusion).

36. See Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 1 (2012); see also Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185, 1226–28 (1986) (arguing that rights should accrue to the user of the program as the best practical solution).

37. See Clifford, *supra* note 33; Farr, *supra* note 31, at 30; see also EUGENE VOLOKH & DONALD M. FALK, FIRST AMENDMENT PROTECTION FOR SEARCH ENGINE SEARCH RESULTS (2012), <http://www.volokh.com/wp-content/uploads/2012/05/SearchEngineFirstAmendment.pdf> [https://perma.cc/Y8RY-5RU2] (arguing that Google is the author of Google search results for First Amendment purposes).

38. See Miller, *supra* note 6, at 1059.

ownership of a work to the operator of a computer program threatens to hinge ownership rights on the press of a button. Consider two proprietors of software that automatically generates musical works on request. Developer 1 runs the program on a private server, instructing it to generate songs whenever a customer submits an order, and then sending only the song file to the customer. Developer 2 sells exactly the same program to end users for download, after which the users can generate as many songs as they like.³⁹ Under a theory that assigns ownership based on operation of the program, Developer 1 owns the copyrights to the songs its program generates, but Developer 2 does not, even though it is the same program. Authorship rights should not depend on something both arbitrary and trivial.⁴⁰

Instead, it would appear that determining authorship of a computer-generated work will require an assessment of where the creative elements in the output of the program originate. That assessment will depend on the answers to two questions: (1) between the user and the programmer, who contributed more to the final product? and (2) what was the quality and nature of the material each contributed? As a result, there is not just one problem of computer-generated works, there is a class of problems, with the analysis depending on the sort of mixture of user and programmer expression represented in the output of a particular program.

There is one important distinction that should be drawn at the outset, namely that, in the case of computer-generated works, there are at least two works at issue, one of which produces the other. First, there is the software code that implements the intermediary work, the computer program. That program is obviously a literary work authored by the programmer. Second, there is the downstream work generated by the program—a movie, videogame, song, news article, poem, or other work.⁴¹ The programmer's control over the software work does not necessarily mean control over the generated work. If the standard model of copyright described above is correct—that what copyright provides is exclusive control over the communication of the author's prior fixed message—then what must be determined is whether the generated work contains any message from an author that needs protecting. And the answer to that depends on how much of the developer's meaning passes through the program down to the ultimate work. That is, it depends on how *transparent* the program is to any meaning intended by the programmer to be contained in the generated work.

The novelty of computer-generated works is that they can combine inputs from any of three different sources: material the programmer provides, material the user provides, or material collected and produced by the operation of the program itself. In order to map the terrain of computer-generated works, it would be most helpful

39. Assume, like the *Ode on a Grecian Urn* in Learned Hand's famous example, the program was independently created. See *Sheldon v. Metro-Goldwyn Pictures Corp.*, 81 F.2d 49, 54 (2d Cir. 1936).

40. There is an overlap here between the notion of making authorship of a computer-generated work hinge on who presses a button, and making performance of a work provided by an automated service hinge on such a fact, which a majority of the Supreme Court recently rejected. See *Am. Broad. Cos., Inc. v. Aereo, Inc.*, 134 S. Ct. 2498 (2014); Bruce E. Boyden, *Aereo and the Problem of Machine Volition*, 2015 MICH. ST. L. REV. 485.

41. There could even be programs that write other programs. See, e.g., JOHN H. HOLLAND, *ADAPTATION IN NATURAL AND ARTIFICIAL SYSTEMS* (1992).

to consider first the easiest examples: situations where almost all of the output is coming from material provided by one of the two human sources. For example, a program could do something as simple as rendering a pre-existing work on the screen or speakers, as occurs with a software music player, a software DVD player, or some software for playing electronic audiobooks. In such a case, the software program is essentially completely transparent in conveying the meaning of the author of the rendered work from author to audience, and the user conversely has very limited control over how the work unfolds.⁴²

Now consider a program, such as a word processor, that provides a nearly blank canvas to the user with no guidance for how to proceed. Programs such as word processors are mere tools for composing works, performing the same function as literal blank canvasses, and from which almost all expressive content in the output originates with the user.⁴³ The software in such cases will be transparent to user meaning but not programmer meaning. The programmer cannot predict what works are going to emerge from its program, whereas the user can. It seems clear in that case that the output works are authored by the user.

These simple examples appear to demonstrate two things. First, determining the authorship of a work will depend on whose meaning or message is embodied in that work. In the easy cases, the computer program is relatively transparent to the meaning input by either programmer or user; but there may be other cases, in which the source of any apparent meaning in a work output by the computer is more obscure.

Second, whether a computer-generated work transmits a person's meaning or message correlates with whether they would be able to predict the output of the computer program in operation. In the case of a CD being played back by a software player, a user listening to the work for the first time would not be able to predict the output of the player; the composer of the song would. Conversely, the programmer of a word processing program is unable to foresee any particular document that might be written by its users. The ability of a person to foresee the work as rendered by a device or process can serve as a proxy for whether that person's meaning or message is embodied in that work.

Additional examples help to fill in missing pieces of the puzzle. Figure 1 below illustrates the relationship between the programmer contributions and user contributions to a work generated by a computer. The x-axis represents the "fixed input," in other words, the content of the program written by the programmer—or more generally, the content added initially to any sort of iterative process that results in another, different work at the end.⁴⁴ The y-axis represents the "progressive input," that is, the content input by a downstream user of the program, which can be a single action, such as hitting "play" on a software CD player, or can

42. The user can start, stop, fast forward, and rewind, but otherwise has little influence over how the work is rendered.

43. The programmer may control incidental features of the literary works produced with a word processor, such as the range of typeface options presented, but otherwise has no influence.

44. The examples need not be limited to computer-generated works, as the inclusion of *Wildflower Works* demonstrates.

be a series of interactions with the program, such as the user who types a document using a word processor.

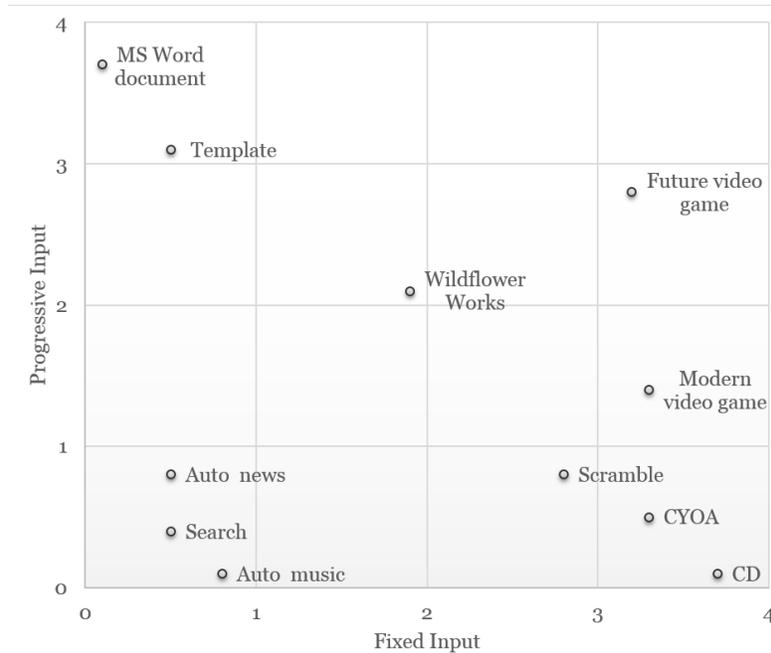


Fig. 1: Examples of fixed vs. progressive input to a work

A word-processing document is an extreme example of a computer-generated work in which there is a high amount of user-added content in the output of the program, and a low amount of programmer-added content. It would therefore fall in the upper left corner of the graph.⁴⁵ Slightly less extreme would be a general template provided by the programmer, such as the outline of a business letter, where most of the detail is still supplied by the user. In the opposite corner, at the lower right, are examples of high programmer contribution and low user contribution, such as a CD played by a software player, or a “Choose Your Own Adventure” story that allows the reader to select a path through the work from a finite number of options. In both cases the works that unspool are almost entirely determined by the fixed contributions of the writers.

As the amount of user control over the work increases, the programmer contributions decrease and the examples move closer to the center of the graph.⁴⁶

45. The x and y values assigned to the data points in Figure 1 are completely arbitrary and for illustrative purposes only.

46. Grady Hendrix, *Choose Your Own Adventure: How the Cave of Time Taught Us to Love Interactive Entertainment*, SLATE (Feb. 18, 2011), <http://www.slate.com/id/2282786/> [<https://perma.cc/GB5N-7LQP>].

In the early 1980s, federal courts were presented with claims that the audiovisual displays of early arcade games such as *Scramble*⁴⁷ and *Defender*⁴⁸ were not authored by the game developers because the player had the ability to control some elements on the screen, such as their ship. While true, the courts held that the amount of variation introduced by the player was insignificant compared to the fixed elements that had been supplied by the game developer.⁴⁹ Modern video games are probably somewhat closer to the center; players can control the camera angle and the sequence of events, and exercise rudimentary control over dialogue, but players still generally have little control over plot, characters, settings, graphics, or sounds. In all of these cases the programmer's meaning is more or less transparent through the program, and the user has very little contribution that alters that meaning.

The examples plotted thus far fall roughly on a line with slope of -1, representing an inverse relationship between programmers and users. These are computer-generated works in which essentially all of the content in the program output is coming from some human source, and the only question is who it is.⁵⁰ As the programmer contribution increases, the user's decreases and vice versa. It seems clear thus far that at least at the extremes what determines authorship of the output of a computer is whether a person's intended meaning reliably or predictably forms part of that output.

The more difficult cases involve a different relationship between content sources, in which the humans are contributing roughly equivalent amounts of material, such that neither the programmer's meaning nor the user's meaning dominates the output. For example, in the top right corner are works that result from both a high level of contribution from programmers and from users as well. It is not clear if any current systems fall into that category, but in the future, it might be possible to make video games in which the players do more than simply react to the story presented onscreen, but participate in creating that story as it unfolds, somewhat like the holodeck on the starship *Enterprise*.⁵¹ In such a case, both player and programmer would be contributing character ideas, dialogue, and plot elements, to be mixed together by the program in ways that perhaps neither party could fully predict. The end result would resemble a joint work, in that it would be the product of contributions made by two or more persons intended to be combined into a single audiovisual work, but that audiovisual work would be one that was assembled without human guidance.⁵² The joint work would emerge from the operation of the program, instead of a close collaboration between authors.

Near the middle of the graph is an example that, while not involving computers,

47. See *Stern Elecs., Inc. v. Kaufman*, 669 F.2d 852 (2d Cir. 1982).

48. See *Williams Elecs., Inc. v. Artic Int'l, Inc.*, 685 F.2d 870 (3d Cir. 1982).

49. See *Stern*, 669 F.2d at 856 (“[T]he repeated appearance of the same sequence of numerous sights and sounds in each play of the game defeats [the defendants’] argument.”).

50. See Fig. 2 *infra*.

51. See generally JANET H. MURRAY, *HAMLET ON THE HOLODECK: THE FUTURE OF NARRATIVE IN CYBERSPACE* (1998).

52. Tyler T. Ochoa, *Who Owns an Avatar? Copyright, Creativity, and Virtual Worlds*, 14 *VANDERBILT J. ENT. & TECH. L.* 959 (2012).

apparently involves a roughly equivalent mixture between prior planning and post-fixation adjustments, mediated by non-human forces: Chapman Kelley's *Wildflower Works*, the subject of *Kelley v. Chicago Park District*.⁵³ *Wildflower Works* was an installation consisting of three beds of wildflowers, which Kelley initially seeded according to his plan and which Kelley and volunteers then tended to each year by weeding and replanting. As in *Stern* and *Williams*, the question was whether *Wildflower Works* was either sufficiently fixed, or sufficiently the product of the plaintiff's authorship, to be copyrightable.⁵⁴ But unlike in the two arcade game cases, the Seventh Circuit held the work lacked both fixation and authorship. The *Kelley* court held that the near-perpetual change caused by natural growth and random events precluded any particular state of the gardens from being recognized as a fixation of the work, a conclusion that seems somewhat in tension with recognizing copies in the volatile memory of a computer as fixations.⁵⁵ But it also held that the amount of post-seeding change that occurs to the work, out of Kelley's control, precluded his claim of authorship of the gardens. Flowers bloomed at varying times, according to natural forces, or randomly failed to sprout at all; seeds dispersed to the winds, and weeds appeared. The court found that this amount of variation overwhelmed whatever meaning or message Kelley may have intended to send with either his initial arrangement or even with his post-planting adjustments: "Most of what we see and experience in a garden—the colors, shapes, textures, and scents of the plants—originates in nature, not in the mind of the gardener."⁵⁶

In the bottom left corner are perhaps the most difficult cases, in which the meaning of the output has little to do with any message contributed by either programmer or user. Contemporary examples include automatic background music generators,⁵⁷ or news article generators,⁵⁸ or poetry bots,⁵⁹ or search engine results.⁶⁰ It is admittedly difficult to determine from news reports just how much of the output of, say, current music generation programs consist of mixtures of human-authored phrases entered into each program. To the extent a program constructs music, news stories, or the like out of large blocks of preexisting content randomly arranged, then such programs are more like the templates considered earlier. Conversely, to the extent such programs merely scramble or randomly generate their inputs, then there may not be much meaning in the output, human or otherwise.⁶¹ But clearly the trend is toward development of programs that generate

53. 635 F.3d 290 (7th Cir. 2011).

54. *Id.* at 303.

55. See *NLFC, Inc. v. Devcom Mid-Am., Inc.*, 45 F.3d 231, 235 (7th Cir. 1995) (citing *MAI Systems Corp. v. Peak Computer, Inc.*, 991 F.2d 511, 519 (9th Cir. 1993)).

56. *Kelley*, 635 F.3d at 304.

57. See *JUKEDECK*, *supra* note 11.

58. See Podolny, *supra* note 19.

59. See "Bot or Not?", *supra* note 7.

60. The meaning of a search engine result is thin, but is something like, "These web pages are relevant to your request." See James Grimmelmann, *Speech Engines*, 94 MINN. L. REV. 868, 933 (2014); Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445, 1469 (2013).

61. Many of the poems produced by "Bot or Not?" appear to have this character, although some

outputs that are both novel and apparently meaningful. If the algorithm is complex enough, even the programmer may not be able to say what works might come out of it. In other words, the programmer may not be able to predict what content might emerge. The user of the program likewise may add little more to it than simply hitting a button.

If the line $y = 4 - x$ in Fig. 1 represents a balance between two human authors, the programmer and the user, then the line $y = x$ represents authorship emerging at least in part from somewhere else, namely from the operation of the program itself.⁶² In the upper right, where both programmer and user have contributed meaning that is conveyed in the work that is output, the program plays the role of editor, or perhaps co-author, joining the contributions together in ways perhaps neither human could have predicted. But in the lower left, where human inputs contribute very little meaning that is conveyed in the output, then whatever aesthetic or informational value the output contains emerges from the program itself. It is in such situations that the claim is sometimes made that the computer program itself “authors” the work—in other words, that the output is unpredictable and not transparent, even to the authors or users of the program.

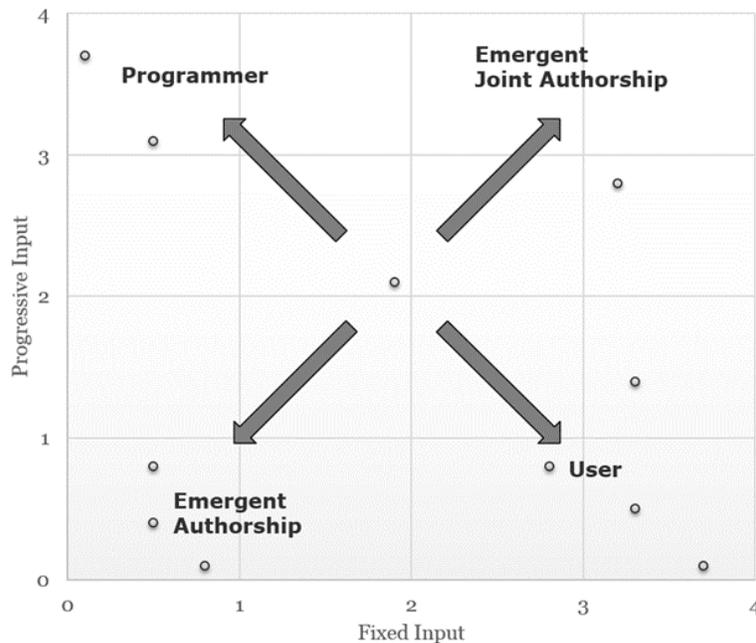


Fig. 2: Types of authorship resulting from computer-generated works

of those are purportedly written by humans. See also Andrew R. Brown et al., *Techniques for Generative Melodies Inspired by Music Cognition*, 39 *COMPUTER MUSIC J.* 11, 14 (2015) (melody generation by random-walk process “lacks larger-scale organization and the ‘meandering’ is, and sounds, directionless”).

62. See Fig. 2.

The prospect of emergent works challenges contemporary copyright law in two ways. First, it poses a challenge for copyright's nondiscrimination principle: the doctrine that copyright protection should not depend on judicial determinations of artistic merit. As a result of that doctrine, all but the most rudimentary works are protected.⁶³ It is a principle that made sense at a time when both modern art and popular culture were beginning their ascendance as cultural forces.⁶⁴ But the effect was to eliminate any need to demonstrate a connection between acts of authorship and the creative elements present in a work. Instead, courts ask, as a proxy for creativity, only whether the putative author engaged in expressive activity, making choices, or judgments, or using skill; if so, then the court simply presumes, as Justice Holmes suggested in *Bleistein*, that there is "something" in the work, somewhere, that is copyrightable.⁶⁵ Even when infringement is asserted, there is no attempt to determine whether the defendant took what is creative in the work, beyond the most obvious limits; closer questions are assigned to the jury with instructions that border on the comical.⁶⁶

Severing the causal connection between author and expression matters little in an environment in which all works are produced by persons and there are only occasional disputes as to whom.⁶⁷ But it leaves existing copyright doctrine unable to address the situation where the output of a computer appears to have minimally creative expression, but that expression is not directly traceable to the content input by any human. One tempting response might be denial: to deny that computers really can generate creative expression without that expression having been built in, or to deny the significance of such a fact. But it is certainly possible for expression that strikes audiences as creative to emerge unbidden from a computer program. Not only is it apparently happening, but there is no reason to suppose that it is not feasible. Advances in the science of complexity within the last thirty years have made it clear that complex, seemingly meaningful patterns can emerge, counterintuitively, from repeated iterations of even just a few simple rules.⁶⁸ The

63. See *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239, 250 (1903).

64. See *id.* at 251–52.

65. See, e.g., *CDN Inc. v. Kapes*, 197 F.3d 1256 (9th Cir. 1999); *Am. Dental Ass'n v. Delta Dental Plans Ass'n*, 126 F.3d 977 (7th Cir. 1997); *CCC Info. Servs. v. MacLean Hunter Mkt. Reports*, 44 F.3d 61 (2d Cir. 1994).

66. Under the test for substantial similarity, or intrinsic similarity as it is called in the Ninth Circuit, the factfinder is asked, among other things, to determine if the works share the same "total concept and feel"—taken literally, whether they share unprotectable ideas and a sensory impression that is irrelevant for most works. See *Cavalier v. Random House, Inc.*, 297 F.3d 815, 822 (9th Cir. 2002); *Boisson v. Banian, Ltd.*, 273 F.3d 262, 272–73 (2d Cir. 2001).

67. Those disputes, when they do arise, can be difficult to resolve, however. See, e.g., *Garcia v. Google Inc.*, 766 F.3d 929 (9th Cir. 2014), *rev'd en banc*, 786 F.3d 733 (9th Cir. 2015).

68. See JOHN H. HOLLAND, *EMERGENCE: FROM ORDER TO CHAOS* 225 (1997) (emergence defined as "regularities in system behavior that are not revealed by direct inspection of the laws satisfied by the components"); MELANIE MITCHELL, *COMPLEXITY: A GUIDED TOUR* 12 (2009) (complex systems consisting of "large networks of components with no central control and simple rules of operation" can "give rise to complex collective behavior, sophisticated information processing, and adaptation via learning or evolution").

sort of regularly occurring, yet novel and dynamic, patterns that define meaningful expression are an unpredictable mix of randomness and order that, while fantastically complex, are not beyond the realm of computation.⁶⁹

It is also significant, both practically and theoretically. As a practical matter, a version of copyright law that does not recognize the significance of emergent works will assign authorship arbitrarily based on whose possession the computer happened to be in when it was used. And as a theoretical matter, there is no good reason to assign initial ownership rights over such works to anyone. No one needs to be incentivized to produce the output of the program; it is a simple matter to generate works once the program is set up. Nor is there any additional need to incentivize the creation of the program, since that incentive is provided by the copyright in the program as a literary work, and access to the program can be controlled by license and by property rights over the server it runs on. Contractual restrictions attaching to the use of the program itself are likely all the programmer would need to prevent competitors from using the program to generate songs for resale in a competing service.⁷⁰ Emergent works likewise need no protection under a moral rights theory as the expression of a human being's personhood, nor under a natural rights theory as intellectual labor, because again there is neither human creativity nor labor involved in their production.⁷¹

A more difficult question is whether the user of a program that produces emergent works may claim copyright protection over them. There is case law suggesting that not only are choices and selections in producing a work acts of authorship sufficient to give rise to copyrightability, but that such acts can be performed after a work is created accidentally or by natural processes. Thus, for example, in Judge Frank's famous hypothetical in *Alfred Bell*, if a shock of thunder produces an involuntary streak across the painting that the painter decides to keep, the creativity comes from the post-hoc decision to *keep* the streak rather than delete it. In other words, an author can add creativity to elements of a work by "ratifying" their presence post hoc, even if they were not consciously planned. But even if that suggestion is accurate, no court has found a single such ratification decision by itself to give rise to a copyright over the thing ratified, and such a result would be inconsistent that the doctrine that compilation copyrights are thin.⁷² Selecting materials to be part of a compilation may justify a copyright over the entire set, but not to the individual selected pieces.

IV. FORESEEABLE AUTHORSHIP

There is thus no simple answer to the question of who authors computer-

69. See Nihat Ay et al., *Effective Complexity and Its Relation to Logical Depth*, 56 IEEE TRANSACTIONS ON INFO. THEORY 4593 (2010) (complexity of strings experiences discontinuous jump at a point in between repeating patterns and randomness). Tracing this argument any further is beyond the scope of this Essay. For the moment, it suffices to note that computer-generated creativity can exist.

70. See *ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447 (7th Cir. 1996).

71. For more on the labor theory or personality-based theory of copyright, see Justin Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L.J. 287 (1988).

72. See *Feist Publ'ns., Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 349 (1991).

generated works. Users author the outputs of programs that consist merely of composition tools. Programmers author the outputs of programs that merely render works input beforehand.⁷³ Both may author works that combine contributions together in unforeseen ways.⁷⁴ And it is possible that no one authors works produced by a program—one that is itself authored—where the authors of the program cannot predictably communicate any particular message or expression to its users. This is new territory for copyright law. Authorship of the output of a computer program would no longer be determined by a straightforward causal analysis, based on whether the putative author produced the work in the right sort of way, but rather would become in at least some instances a probabilistic inquiry. A person would be the author of the output of a program only if he or she could predict, or rather foresee, that output, more or less.

That shift is a fairly confounding one for present doctrine. It is reminiscent of the shift that occurred in tort law in the nineteenth century, in which actors began to be held responsible for actions that did not directly and intentionally injure anyone, but rather created an intangible risk of unintended injury. This was the tort of negligence, formed to deal with the last great technological disruption, the Industrial Revolution in the United States.⁷⁵ Suddenly it was possible to be held liable in tort for actions which did not always, or even frequently, cause injury to a third party, but rather raised the risk of injury to third parties. For a long time after this development, courts and scholars resisted thinking of negligence liability in terms of probabilities, however, seeking instead certainties. That is, well into the twentieth century courts and legal scholars engaged in a search for specific, rigid rules for determining “fault” as a basis for negligence.⁷⁶ It was only in the 1920s that that search was abandoned, and the law began to embrace the idea that negligence liability would lie where the defendant increased someone’s risk beyond a level where he or she should have foreseen an unreasonably high probability of harm.⁷⁷

Something similar may happen for computer-generated speech. It may take some time for legal decisionmakers to become accustomed to the idea that speech may arise in some cases only indirectly from human actions. That will require an adjustment in the doctrines used to determine authorship, from those that focus on

73. That is the case, at least, if the programmer creates a work embedded in the software, such as a video game. Alternatively, if the program renders a work in a separate data file, such as a software mp3 player or DVD player, then the creators of the file are the authors of the work that is rendered; the point here is that the authors are on the “production” side of the standard model.

74. This is a difficult question, which I will not explore further here. For authorship of a joint work, it must be true that the authors “prepared the work with the intention that their contributions be merged into inseparable or interdependent parts of a unitary whole,” and that each have the intent to be co-authors. See *Childress v. Taylor*, 945 F.2d 500, 507 (2d Cir. 1991). Typically the authors work together in arranging the contributions, so that at least one human is in charge of the merging process. But is human control necessary? If contributions are intended to be merged together into one work, does that constitute the necessary “prepar[ation]” of the work?

75. See LAWRENCE M. FRIEDMAN, *A HISTORY OF AMERICAN LAW* 223 (3d ed. 2005).

76. G. EDWARD WHITE, *TORT LAW IN AMERICA: AN INTELLECTUAL HISTORY* 40–41, 57–58 (expanded ed. 2003).

77. *Id.* at 94–95.

the connection between actions and speech to those that focus instead on states of mind. That has already started occurring when it comes to the issue of direct infringement—that is, the inclusion of copies of someone else’s works in what one writes or publishes. Twenty years ago, as the first infringement suits arose alleging automated copying, courts began requiring plaintiffs to demonstrate a conscious act to establish direct liability, namely that the infringement was the product of human “volition.”⁷⁸

The elimination of an assumption that authorship caused a work to be created will have destabilizing effects on copyright law. Detaching speech from speaker pulls apart interests that are almost always conjoined in the law: the speaker’s interest in expressing his or her thoughts, the listener’s interest in being enlightened, and the societal interest in the dissemination of information. For copyright, it has been assumed that the economic value of a work—that is, the audience demand for it—was attributable to the author’s creativity, thus justifying the reward of exclusive rights over some portion of that value for the author. But now there are works in which the aspects the audience is most interested in are generated by computer. To the extent those aspects dominate, there is no reason to provide exclusive rights in each such work to the creator of the program that generates them, who already will enjoy exclusive rights to the literary work that embodies the program.

While that may be correct in theory, it will be challenging to put into practice. It may be that the distinction between computer-generated works and human-authored works is very difficult to enforce, by either the Copyright Office or courts or parties outside of litigation, such as ISPs. That is, it may be difficult or impossible to tell in the future whether a song or news article or a short film is computer-generated or human-generated just by looking at it. If such works become prevalent at a reasonably high quality, that may require a reassessment of whether the exclusive rights of copyright are even necessary. But as long as they are, some method of sorting out authors from non-authors will be required.

The most reasonable solution would be to alter the nature of proof of authorship we require from copyright claimants. Tort law shifted from negligence to strict liability in cases where proof of negligence became too onerous.⁷⁹ Copyright could similarly shift to requiring a copyright claimant to prove human authorship of certain forms of computer-generated works, not simply by making choices about the program that generated them, but by establishing that the output foreseeably includes a meaning or message that the author wishes to convey to his or her audience. Only the author would have knowledge of this, making it natural to put the affirmative burden on him or her to establish it.

When it comes to torts, the necessary foreseeability for negligence liability is whether a reasonable person, knowing what the defendant knew, would have

78. See *Religious Tech. Ctr. v. Netcom On-Line Comm’n Servs., Inc.*, 907 F. Supp. 1361 (N.D. Cal. 1995).

79. See David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 132 (2014).

perceived the risk.⁸⁰ That perception may be clouded; a defendant foresees a risk not if they can predict a particular result with certainty, but if they can predict results of that sort with a minimum probability that depends on the magnitude of the harm.⁸¹ In other words, liability attaches for negligence if the defendants' action foreseeably conveys a risk to the plaintiff.

Authorship of emergent works may involve a similar inquiry. If a reasonable person, knowing what the putative author knew, would have tried to convey a given meaning to an audience through the computer program at issue, then the communication of that meaning qualifies as an authored work. The test for authorship should be whether the putative author foreseeably communicated that meaning to the audience. Although it involves probabilities and communicative intent, that is the sort of question a jury can feasibly evaluate as a matter of fact, and it may need to be added to the showing a plaintiff must make to prevail in cases involving computer-generated works in which some of the creativity is likely to be emergent.

CONCLUSION

We are at the dawn of the age of emergent works. Legal scholars and practitioners in 1850, before they had even conceived of a separate area of the law known as "torts," would have found it difficult to imagine the rise of the tort of negligence, let alone products liability a century later. It is difficult now to imagine the legal rules in a world in which the drudgework of creation may be assigned to robots, but the joy of creation remains human. It is difficult to foresee what continuing need the incentives and rewards of copyright will still serve in such a future. But the onset of the problem is clearly upon us.

80. DAN B. DOBBS, *THE LAW OF TORTS* 335 (2000); 3 FOWLER V. HARPER ET AL., HARPER, JAMES AND GRAY ON TORTS § 16.9 at 519–20 (3d ed. 2007).

81. See HARPER ET AL., *supra* note 80, § 16.9 at 523–24.