

AGAINST SEARCH ENGINE VOLITION

ABSTRACT

The concept of volition in copyright arose as a way to insulate Internet Access Providers and Bulletin Board Systems from liability for direct infringement prior to the 1998 “safe harbor” provisions of the Digital Millennium Copyright Act. The volition requirement purported to be a technical test, finding that volition was lacking where the technology, like a copy machine, responded “automatically” to a “third party request.” However, the metaphor of the copy machine and the application of volition in recent cases suggest that the requirement is not wholly technical. Instead, courts use the technical prongs of volition to approximate when the copying is sufficiently close to the infringing conduct that the machine owner himself can be said to trespass on the exclusive domain of the copyright owner. The problem is this: despite the clear distance between search engines and uploaders of infringing copies, the “third party request” element of the volition test would presently prevent courts from finding that search engines lack volition. This Comment argues that the test should be subtly retooled to accommodate the search engine model.

206**ALB. L.J. SCI. & TECH.****[Vol. 18**

TABLE OF CONTENTS

ABSTRACT	205
I. INTRODUCTION.....	207
II. THE VOLITION REQUIREMENT.....	210
A. What is Volition?	211
B. Why Does it Matter?	213
III. VOLITION AND THE ORIGINAL ISPs: INTERNET ACCESS PROVIDERS AND BULLETIN BOARD SYSTEMS.....	215
IV. SEARCH ENGINES AND VOLITION	218
A. Are Search Engines more like Web sites or ISPs?	219
B. The Implications of Parker v. Google	221
V. ADAPTING VOLITION TO ACCOMMODATE SEARCH ENGINES ...	223
A. Re-envisioning Third Party Requests	223
B. A Challenge and a Response.....	225
VI. CONCLUSION.....	226

I. INTRODUCTION

The concept of volition in copyright arose as a way to insulate two types of Internet Service Providers (“ISPs”)—Internet Access Providers (“IAPs”) and Bulletin Board Systems (“BBSs”)—from liability for direct infringement prior to the 1998 “safe harbor” provisions of the Digital Millennium Copyright Act (“DMCA”). It developed as a response to the conflict between direct infringement and the copy-based nature of cyberspace. Because direct copyright infringement required a plaintiff to show only two elements—ownership of a valid copyright in the infringed material, and “copying” by the defendant¹—ISPs were inherently violating copyright whenever a user transmitted or uploaded an infringing copy of a work.²

Faced with cases alleging infringement by ISPs, bulletin boards, and other computer operators, courts began to read a volition requirement into the concept of direct infringement in order to prevent liability and the attendant upheaval of our modern communications network.³ Courts found that the Copyright Act’s treatment of infringement as the violation of the exclusive rights of the copyright holder implied that an element of direct action or participation was necessary for direct infringement.⁴ Thus, the requisite volition was held not to exist where a technology responded automatically to a third party

¹ Sid & Marty Krofft Television Prods., Inc. v. McDonald’s Corp., 562 F.2d 1157, 1162 (9th Cir. 1977); Religious Tech. Ctr. v. Netcom On-Line Commc’n Servs., Inc., 907 F. Supp. 1361, 1366–67 (N.D. Cal. 1995) (citation omitted).

² See Niva Elkin-Koren, *Copyright Law and Social Dialogue on the Information Superhighway: The Case against Copyright Liability of Bulletin Board Operators*, 13 CARDOZO ARTS & ENT. L.J. 345, 365 (1995); see also Perfect 10, Inc. v. Cybernet Ventures, Inc., 213 F. Supp. 2d 1146, 1167 (C.D. Cal. 2002). Computer technology, and in particular the Internet, has created a challenge to copyright’s strict liability scheme. Because of the architecture of the web and the workings of computer technology, almost any business that utilizes computer hardware to create access to the Internet or to store content may find its hardware creating or displaying infringing material as a result of decisions by third-parties (the system’s users) without the business doing any truly volitional actions.

Perfect 10, 213 F. Supp. 2d at 1167.

³ See, e.g., *Netcom*, 907 F. Supp. at 1370.

⁴ *Id.* at 1370–71 (comparing *Netcom* to a situation where the defendant is not found liable for a direct infringement in the absence of direct participation); see also 17 U.S.C. § 501(a) (Supp. 2004); Jennifer L. Kostyu, Comment, *Copyright Infringement on the Internet: Determining the Liability of Internet Service Providers*, 48 CATH. U. L. REV. 1237, 1253 (1999).

request.⁵

Courts made the following analogy: an ISP is the same as a copy machine in that it functions automatically at the request of a user, without any intervention or oversight by the machine's owner.⁶ Like a copy machine owner, the ISP owner does not engage in activities reserved to copyright owners; she merely supplies a tool that third parties might use to infringe copyright.⁷ Therefore, because copy machine owners are not liable for direct infringement when users make illegal copies, the courts said that an ISP should not be liable as a direct infringer when its system is similarly used by a subscriber to transmit an infringing copy.⁸

However, the metaphor of the copy machine and the application of volition in recent cases suggest that the requirement is not wholly technical. Instead, courts use the technical prongs of volition to approximate when the copying is sufficiently close to the infringing conduct that the machine owner himself can be said to trespass on the exclusive domain of the copyright owner.⁹ The result is that we see courts finding

⁵ *Netcom*, 907 F. Supp. at 1370; see also Jennifer E. Markiewicz, Comment, *Seeking Shelter From the MP3 Storm: How Far Does the Digital Millennium Copyright Act Online Service Provider Liability Limitation Reach?*, 7 COMMLAW CONSPPECTUS 423, 430–31 (1999).

⁶ See, e.g., *Netcom*, 907 F. Supp. at 1369.

⁷ *Id.* at 1368, 1369–70.

⁸ *Id.* at 1369–70; see also *Parker v. Google, Inc.*, 422 F. Supp. 2d 492, 497 (E.D. Pa. 2006).

The United States Court of Appeals for the Fourth Circuit has concluded that while the Copyright Act does not require that the person intentionally infringe a copyright, “it nonetheless requires *conduct* by a person who causes in some meaningful way an infringement.” . . . Courts have drawn an analogy between an internet service provider (“ISP”), which maintains “a system that automatically transmits users' material but is itself indifferent to the material's content,” and the owner of a copy machine who makes its use available to the public.

Parker, 422 F. Supp. 2d at 497 (citing *CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 549, 551 (4th Cir. 2004)); see also *Playboy Enters., Inc. v. Russ Hardenburgh, Inc.*, 982 F. Supp. 503, 512 (N.D. Ohio 1997) (citing *Netcom*, 907 F. Supp. at 1369); *Marobie-Fl, Inc. v. Nat'l Ass'n Fire Equip. Distribs.*, 983 F. Supp. 1167, 1178 (N.D. Ill. 1997) (“Like a copying machine owner, [the Web hosting service] did not actually engage in any of these activities itself. Accordingly, [the Web hosting service] may not be held liable for direct infringement.”); Nicole Bashor, Comment, *The Cache Cow: Can Caching and Copyright Co-exist?*, 6 J. MARSHALL REV. INTELL. PROP. L. 101, 119 (2006).

⁹ *LoopNet*, 373 F.3d at 550 (4th Cir. 2004).

There must be actual infringing conduct with a nexus sufficiently close and causal to the illegal copying that one could conclude that the machine owner himself trespassed on the exclusive domain of the copyright owner. The *Netcom* court described this nexus as requiring some aspect of volition or causation.

volition where a BBS operator actively solicits infringing material,¹⁰ and—contrary to a strict reading of the “automatic response” element—not finding volition where a BBS operator actively reviews and manually approves such uploads.¹¹

In 2006, this volition requirement was extended to the context of Internet search engines.¹² Like IAPs and BBSs, search engines are integral to the working of the Internet. They catalogue the World Wide Web and thus make it more accessible to Internet users. Search engines do not generally aim to violate the copyright of the content they collect and organize, but merely aggregate online information in a way that is economically feasible—automatically, without waiting for third party requests.¹³ Google, for example, trolls the Web through automatic technology (its bots), organizes that information into a distributable form (its index), and then makes that information available to the public via its Web site.¹⁴

The problem is that despite the clear distance between search engines and uploaders, such that infringing copies may be automatically added to the search engine’s cache, the absence of the “third party request” element would necessarily deem a search engine as having volition. The information is not being sent via a “third party request”; instead, search engines access information through the absence of “robots.txt” instructions embedded in HTML pages—thereby running afoul of a literal reading of the “third party request” language.¹⁵ To put it simply, search engines, while in some sense responsive, are not passive in the original volitional sense because they take the affirmative

Id. (citing *Netcom*, 907 F. Supp. at 1370).

¹⁰ *Playboy*, 982 F. Supp. at 513.

¹¹ *LoopNet*, 373 F.3d at 556 (noting that LoopNet’s anti-infringement procedures allow it to be “compared to an owner of a copy machine who has stationed a guard by the door to turn away customers who are attempting to duplicate clearly copyrighted works.”).

¹² See *Parker*, 422 F. Supp. 2d at 497 (“It is clear that Google’s automatic archiving of USENET postings and excerpting of websites in its results to users’ search queries do not include the necessary volitional element to constitute direct copyright infringement.”).

¹³ See Sergey Brin & Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, <http://infolab.stanford.edu/~backrub/google.html> (last visited March 13, 2008) (“In order to scale to hundreds of millions of web pages, Google has a fast distributed crawling system.”).

¹⁴ *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1110 (D. Nev. 2006).

¹⁵ *Id.* at 1113, 1115.

act of crawling the Web.¹⁶

This Comment argues that the test should be adapted to recognize search engine non-volition. Part II outlines the volition requirement and explains why the requirement still matters after the adoption of DMCA “safe harbor” provisions protecting ISPs. Part III explains how volition evolved to protect those technological intermediaries who are at a satisfactory distance from the infringing conduct.¹⁷ It then shows how the proxies for volition—an “automatic response” to a “third party request”—are able to reflect that distance between the infringing conduct and the technological copying in the IAP and BBS context.

Part IV explains why the same proxies fail to vindicate non-volition in the search engine context. It argues that this disparity both explains and justifies the fragmentary application of the *Netcom* analysis in *Parker v. Google*.¹⁸ Parts V and VI contend that rather than ignoring half of the volition analysis as the court did in *Parker*, courts should adapt the “third party request” element to recognize search engine non-volition via its role as a conduit between content providers and Internet users and its responsiveness to “robots.txt” files.

II. THE VOLITION REQUIREMENT

The volition requirement is included in the Copyright Act only by implication. The Copyright Act states that direct infringement occurs when a defendant violates one of the exclusive rights of the copyright holder, such as reproducing the copyrighted work, preparing derivative works, distributing copies

¹⁶ See *id.* at 1115 (providing an example of when the search engine Google was not passive but instead “when its ‘Googlebot,’ like an ordinary Internet user, made the initial copies of the Web pages containing [the plaintiff’s] copyrighted works and store[d] those copies in the Google cache.” The court explained that Google was in some sense responsive when “Google’s computer respond[ed] automatically to the user’s request.”). See also Brin, *supra* note 13 (explaining that by web crawling Google is affirmatively gathering various URLs and storing them).

¹⁷ *CoStar Group, Inc. v. Loopnet, Inc.*, 373 F.3d 544, 551 (4th Cir. 2004) (requiring a person to engage in “volitional conduct – specifically, the act constituting infringement – to become a direct infringer” is “especially important when it is applied to cyberspace” and defendants “are conduits from or to would-be copiers and have no interest in the copy itself”).

¹⁸ See *Parker v. Google, Inc.*, 422 F. Supp. 2d 492, 497 (E.D. Pa. 2006) (discussing that the search engine’s automatic activity caused the necessary volitional element to be missing).

to the public, or publicly displaying the work.¹⁹ Courts have read this statutory focus on activities reserved to copyright owners to require “some element of direct action or participation” for direct infringement.²⁰ That requirement, serving to demarcate the boundary between direct infringement and contributory infringement, is volition. This Part explains the modern origin of the volition requirement as it was used to insulate ISPs from direct infringement liability and explains why the requirement still matters even after its codification in the DMCA “safe harbor” provisions.²¹

A. *What is Volition?*

Volition arose as a solution to the problem of direct copyright infringement in cyberspace. Direct copyright infringement requires that a plaintiff show two elements: ownership of a valid copyright in the infringed work, and copying by the defendant.²² With the advent of network communications such as the Internet, the traditional definition of “copying” created interpretive problems for the direct infringement analysis.²³ Specifically, the notion of what constituted “copying” was called into question by the transitory and automatic nature of the copies created in the process of digital communication.²⁴

One strand of this dilemma emerged in copyright claims against ISPs. By virtue of the copy-based nature of their technology, ISPs were inherently violating copyright whenever a valid copyright was transmitted over their networks.²⁵ As such,

¹⁹ 17 U.S.C. § 501(a) (Supp. IV 2000); 17 U.S.C. § 106(1)–(3), (5) (2000 & Supp. IV 2000).

²⁰ *See, e.g.*, *Playboy Enters. v. Russ Hardenburgh, Inc.*, 982 F. Supp. 503, 512, 513 (N.D. Ohio 1997).

²¹ 17 U.S.C. § 512(b)(1).

²² *Religious Tech. Ctr. v. Netcom On-Line Commc’n Servs., Inc.*, 907 F. Supp. 1361, 1366–67 (N.D. Cal. 1995); *Sid & Marty Krofft Television Prods., Inc. v. McDonald’s Corp.*, 562 F.2d 1157, 1162 (9th Cir. 1977).

²³ *Elkin-Koren*, *supra* note 2, at 381–82; *see also Perfect 10, Inc. v. Cybernet Ventures, Inc.*, 213 F. Supp. 2d 1146, 1167 (C.D. Cal. 2002) (explaining that the Internet causes copyright infringement issues).

²⁴ *Elkin-Koren*, *supra* note 2, at 382; *see also Netcom*, 907 F. Supp. at 1368–69. The *Netcom* court states that in the digitalized world when a system makes temporary copies, the system is not always deemed to have caused the copying. *Id.* at 1368, 70. In addition, the court recognizes *Elkin-Koren* who comments on the difficulty of applying copyright concepts to a digital environment. *Id.* at 1369 n.11.

²⁵ *See CoStar Group, Inc. v. Loopnet, Inc.*, 373 F.3d 544, 551 (4th Cir. 2004) (explaining that an ISP automatically receives infringing materials at the

ISPs were at risk of being held liable for direct infringement claims for any infringing copies uploaded by their subscribers.²⁶ Rather than stringently apply the law and potentially hinder a technology making the Internet possible, courts began to read an element of direct action or participation into the law of direct infringement.²⁷ The courts recognized that ISPs were not the type of infringer that copyright was intended to protect against, and that ISPs had been implicated only because of the copy-based nature of their communications regime; courts therefore held that such service providers lacked the volition necessary to directly infringe the rights of copyright owners.²⁸

In perhaps the earliest and most famous of the cyberspace volition cases, *Religious Technology Center v. Netcom On-Line Communication Services, Inc.* (“Netcom”), the United States District Court for the Northern District of California explicitly addressed the copyright and cyberspace dilemma with respect to ISPs.²⁹ The court noted that to impose liability based on the traditional analysis alone would “result in liability for every single . . . server in the worldwide link of computers transmitting [a user’s] message to every other computer.”³⁰ For this reason, the court held that Netcom did not directly infringe when it installed and maintained software that temporarily stored copies of subscriber messages on Netcom servers while automatically forwarding those messages onto Usenet.³¹ It held, in short, that Netcom lacked the requisite volition.³²

The volition requirement that emerged from *Netcom* thus precluded liability where Internet copying was “automatically

instigation of the subscriber); see also *Netcom*, 907 F. Supp. at 1367–68 (providing an example of the basic copy-based process of ISP technology).

²⁶ See *Sega Enters. v. Maphia*, 857 F. Supp. 679, 686 (N.D. Cal. 1994) (citing a general rule that when BBS users uploaded unauthorized video games to a BBS, “[t]he unauthorized copying of copyrighted computer programs [was] *prima facie* an infringement of the copyright”); see also Elkin-Koren, *supra* note 2, at 363–64.

²⁷ See, e.g., *Netcom*, 907 F. Supp. at 1375. (“Where a defendant has knowledge of the primary infringer’s infringing activities, it will be liable if it ‘induces, causes or materially contributes to the infringing conduct of the primary infringer.’”) (citation omitted).

²⁸ *Id.* at 1369–70.

²⁹ *Id.* at 1365.

³⁰ *Id.* at 1369.

³¹ See *id.* at 1368–69.

³² See *id.* at 1368–70 (“Although copyright is a strict liability statute, there should still be some element of volition or causation which is lacking where a defendant’s system is merely used to create a copy by a third party.”).

made on their computers using their software as part of a process initiated by a third party.”³³ The court explained that Netcom was not liable for direct infringement because its “act of [creating] a system that automatically and uniformly creates temporary copies of all data sent through it [by subscribers] is not unlike that of the owner of a copying machine who lets the public make copies with it.”³⁴ Where a third party used a defendant’s system to create a copy, then, the requisite element of volition was missing, making “contributory infringement, not direct infringement,” the appropriate test.³⁵

B. *Why Does it Matter?*

Since the 1998 addition of the Digital Millennium Copyright Act (“DMCA”) “safe harbor” provisions, the volition requirement is no longer the only tool used to insulate ISPs from copyright infringement liability. The safe harbor provisions statutorily immunize ISPs from copyright infringement for copies made in the transmission of information, provided:

(1) the transmission of the material was initiated by or at the direction of [another party]; (2) the transmission . . . is carried out through an automatic technical process . . . ;(3) the service provider does not select the recipients of the material except as an automatic response to the request of another person; (4) the [transmitted] material . . . is [not otherwise made] accessible [or permanently stored]; and (5) the material is transmitted through the system or network without modification of its content.³⁶

In other words, the safe harbor provisions essentially codify the *Netcom* principles in Title II of the DMCA. According to the House Report of Title II of the DMCA:

As to direct infringement, liability is ruled out for passive, automatic acts engaged in through a technological process initiated by another. Thus the bill essentially codifies the result in the leading and most thoughtful judicial decision to date:

³³ *Religious Tech. Ctr. v. Netcom On-line Commc’n. Servs. Inc.*, 907 F. Supp. 1361, 1368 (N.D. Cal. 1995).

³⁴ *Id.* at 1369.

³⁵ *Id.* at 1369 (quoting Elkin-Koren, *supra* note 2, at 363) (“[C]ontributory infringement is more appropriate for dealing with BBS liability, first, because it focuses attention on the BBS-users relationship and the way imposing liability on BBS operators may shape this relationship, and second because it better addresses the complexity of the relationship between BBS operators and subscribers.”).

³⁶ 17 U.S.C. § 512(a)(1)–(5) (2000).

Religious Technology Center v. Netcom On-Line Communications Services, Inc. . . . In doing so, it overrules these aspects of *Playboy Enterprises, Inc. v. Frena* . . . insofar as that case suggests that such acts by service providers could constitute direct infringement, and provides certainty that *Netcom* and its progeny, so far only a few district court cases, will be the law of the land.³⁷

Given that codification, there would seem little need to continue the common law application of volition. Yet, as the Fourth Circuit explained in *LoopNet*, that would doubly be a mistake. First, the final version of the legislation ultimately failed to unconditionally exempt ISPs from direct infringement liability as the above-quoted House Report suggests.³⁸ Instead, the safe harbor guarantees DMCA protection only if ISPs fulfill certain obligations, such as taking down infringing material at the request of the copyright owner.³⁹ While such obligations may appear slight, this means that the protection guaranteed by the safe harbor provisions and the volition requirement are not wholly overlapping, but complementary fields of protection.

Second, in rejecting CoStar's argument that Congress's codification of *Netcom* in the DMCA essentially overrides and replaces the *Netcom* decision, thereby obviating the volition analysis, the Fourth Circuit explained that CoStar had it exactly backward: "When Congress codifies a common-law principle, the common law remains not only good law, but a valuable touchstone for interpreting the statute, unless Congress explicitly states that it *intends* to supplant the common law."⁴⁰ The court also pointed to the legislative history of the DMCA, citing the final conference report which states:

As provided in subsection (l), Section 512 is not intended to imply that a service provider is or is not liable as an infringer either for conduct that qualifies for a limitation of liability or for conduct that fails to so qualify. Rather, the limitations of

³⁷ *ALS Scan, Inc. v. RemarQ Cmty., Inc.*, 239 F.3d 619, 622 (4th Cir. 2001) (discussing H.R.REP. NO. 105-551(1) (1998)) (citations omitted).

³⁸ *See CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 554 (4th Cir. 2004).

³⁹ *Id.* ("[U]nlike the earlier version, which would have exempted ISPs unconditionally for direct liability for automatic processes, the enacted law requires ISPs to fulfill certain '[c]onditions for eligibility' for the safe harbors.") (citation omitted); *see also* 17 U.S.C. § 512(c)(1)(C) (2000) ("[U]pon notification of claimed infringement [an ISP is not liable if it] responds expeditiously to remove, or disable access to, the material that is claimed to be infringing or to be the subject of infringing activity.")

⁴⁰ *Id.* at 553 (emphasis supplied) (citations omitted).

2008] AGAINST SEARCH ENGINE VOLITION 215

liability apply if the provider is found to be liable under existing principles of law.⁴¹

Thus, as the Fourth Circuit concludes, the volition requirement is not superseded by the DMCA. Rather, the volition analysis precedes the “safe harbor” analysis because it is part of deciding the scope and nature of prima facie copyright liability.⁴² Volition determines whether direct infringement is possible; safe harbor provisions determine whether there is liability notwithstanding a finding of direct infringement. As the court put it, the DMCA’s safe harbor for ISPs operates as “a floor, not a ceiling, of protection.”⁴³

III. VOLITION AND THE ORIGINAL ISPS: INTERNET ACCESS PROVIDERS AND BULLETIN BOARD SYSTEMS

The *Netcom* volition analysis proceeds through the metaphor of the copy machine owner that makes his machine available to the public—finding no volition where the machine automatically responds to a third party request. *LoopNet* has, however, more recently emphasized that the technical elements of the test do not constitute volition but merely serve as proxies for it.⁴⁴ The elements help courts determine whether the copying is sufficiently close to the infringing conduct that the machine owner himself can be said to trespass on the exclusive domain of the copyright owner. This Part examines how the *Netcom* proxies successfully vindicate the *LoopNet* distance in the IAP and BBS contexts.

The concept of volition was first applied to IAPs and BBSs.⁴⁵ An IAP is a business or organization that provides an access point to the Internet.⁴⁶ Through dial-up, DSL, or other

⁴¹ *Id.* (citing H.R.REP. NO. 105–796, at 73 (1998) (Conf. Rep.)) (emphasis supplied).

⁴² *Id.*

Thus the DMCA was intended not to change the “evolving” doctrines on ISP liability for copyright infringement, which included *Netcom* and *Frena*, but to offer a certain safe harbor for ISPs. Courts were left free to continue to construe the Copyright Act in deciding the scope and nature of prima facie liability.

Id. (emphasis supplied).

⁴³ *CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 555 (4th Cir. 2004).

⁴⁴ *Religious Tech. Ctr. v. Netcom On-Line Commc’n Servs., Inc.*, 907 F. Supp. 1361, 1369, 1370 (N.D. Cal. 1995).

⁴⁵ *See, e.g., Netcom*, 907 F. Supp. at 1381–82; *Loopnet*, 373 F.3d at 555.

⁴⁶ *Playboy Enters., Inc. v. Webbworld, Inc.*, 991 F. Supp. 543, 552 (N.D. Tex. 1997) (“Unlike the defendant service provider in [an IAP], Webbworld did not

connective technology, IAPs are a conduit for Internet information, though they may also provide services like domain name registration and hosting, web hosting, and collocation.⁴⁷ Some well-known IAPs include America Online, CompuServe, NetZero, SBC/Yahoo, Verizon, and Earthlink. Like IAPs, BBSs allow users to dial into an electronic system in order to be connected with other users.⁴⁸ BBSs were an early, localized type of Internet service provider that facilitated access to an online space where users could talk, publish, and download games and software.⁴⁹

In *Netcom*, the court dealt with both of these types of ISPs. There, the infringing network user, Erlich, originally uploaded his messages to the BBS. The messages were then automatically copied to Netcom's servers and thereby made available to other users of the Internet.⁵⁰ Because the infringing user gained access to the Internet through the BBS "support.com," and the "BBS [was] not directly linked to the Internet, but gain[ed] its [access] through the facilities of . . . Netcom On-Line Communications, Inc.," both companies were alleged to have directly infringed a copyright by copying the uploaded works.⁵¹ In recognizing that the two service providers were not closely involved with the infringing conduct of Erlich, the court found that neither the IAP (Netcom) nor the BBS (support.com) had the requisite volition for direct infringement because their role, like the "owner of a copying machine who lets the public make copies with [his machine]," was an automatic response to a third party request.⁵²

In a later case, *Playboy v. Russ Hardenburgh*, the court addressed a situation where there was a close nexus between the copying by the BBS and the infringing conduct by the network

function as a mere provider of access.").

⁴⁷ BusinessDictionary.com, <http://www.businessdictionary.com/definition/internet-access-provider-IAP.html> (last visited Mar. 20, 2008).

⁴⁸ Reference for Business, <http://www.referenceforbusiness.com/encyclopedia/Bre-Cap/Bulletin-Board-Systems-BBS.html> (last visited Mar. 20, 2008).

⁴⁹ *See id.*

⁵⁰ *See Netcom*, 907 F. Supp. at 1366.

⁵¹ *Id.*

⁵² *Id.* at 1369, 1372, 1382 ("The court believes that Netcom's act of designing or implementing a system that automatically and uniformly creates temporary copies of all data sent through it is not unlike that of the owner of a copying machine who lets the public make copies with it.").

user sufficient to find the former had volition.⁵³ There, the court found that volition existed where the operator of a BBS not only knew the infringing activity was occurring, but actively solicited the uploading of that content by its users.⁵⁴ In such a scenario, the technical volition elements vindicated the close nexus between the infringing conduct and the copying because the BBS was not automatically responding to third party requests, but rather actively requesting infringing third party uploads.⁵⁵

More recently, in *LoopNet*, the Fourth Circuit has shown how the copy machine metaphor can be used in the BBS context to accommodate situations where the nexus is distant and the *Netcom* “automatic response” element fails to vindicate that distance. In that case, the court decided whether a BBS that personally reviewed third party uploads of pictures for real estate listings had the requisite volition for direct infringement where a user of the service uploaded an infringing picture.⁵⁶ Despite the fact that LoopNet personnel approved the upload and publication of the infringing copy, and therefore failed the “automatic response” half of the *Netcom* analysis, the court employed the copy machine metaphor in a way that did not prioritize the technical elements of volition, but instead reflected the distance between the infringing conduct and LoopNet’s copying.⁵⁷

While a fair application of the metaphor would have LoopNet with its finger on the copy machine button, pressing it only after a cursory look at what the user placed on the machine,⁵⁸ the court found that LoopNet served as a guard at the door of the

⁵³ See generally *Playboy Enters., Inc. v. Russ Hardenburgh, Inc.*, 982 F. Supp. 503, 511–14 (N.D. Ohio 1997).

⁵⁴ *Id.* at 513. But see *Sega Enters. v. MAPHIA*, 948 F. Supp. 923, 932 (N.D. Cal. 1996) (extending volition immunity to a situation where the operator of a BBS not only knew the infringing activity was occurring, but also solicited others to upload games).

⁵⁵ See *Playboy Enters.*, 982 F. Supp. at 513.

⁵⁶ *CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 547, 556 (4th Cir. 2004).

⁵⁷ *Id.* at 547, 550.

A LoopNet employee then cursorily reviews the photograph (1) to determine whether the photograph in fact depicts commercial real estate, and (2) to identify any obvious evidence, such as a text message or copyright notice, that the photograph may have been copyrighted by another. If the photograph fails either one of these criteria, the employee deletes the photograph and notifies the subscriber. Otherwise, the employee clicks an “accept” button that prompts LoopNet’s system to associate the photograph with the web page for the property listing, making the photograph available for viewing.

Id. at 547.

⁵⁸ *Id.* at 547.

copy machine room, excluding certain unwanted visitors.⁵⁹ The court thus recognized that LoopNet's review of uploaded images, far from being close to the infringing conduct, was an effort to protect (at least in part) against just such infringement; LoopNet wanted to prevent the posting of unauthorized images, but simply had no way to provide effective oversight where the image lacked an embedded copyright notice.⁶⁰

Thus, insofar as the IAP and BBS contexts are concerned, volition is a flexible and cohesive test because the proxies generally succeed. They detect a lack of volition in *Netcom* where the ISPs have no connection to the infringing conduct other than indiscriminately and inadvertently transmitting a copy of the infringing work,⁶¹ and they find volition in *Playboy* where the ISP's copying was related to the infringing conduct because the infringing conduct was encouraged.⁶² Moreover, even where the proxies would seem to fail—such as in *LoopNet* where the “automatic response” element would militate for volition⁶³—the copy machine metaphor is able to stretch far enough to accommodate the gap between the machine owner's copying and the infringing conduct.⁶⁴

IV. SEARCH ENGINES AND VOLITION

The question today is how volition applies to ISPs of a different sort: search engines. This Part explains why the *Netcom* proxies that were developed in the context of IAPs and BBSs fail to approximate the *LoopNet* distance between the copying and infringement in the search engine context. It then argues that this divergence both explains and justifies the fragmentary application of the *Netcom* analysis in *Parker v. Google*.⁶⁵

⁵⁹ *Id.* at 556 (“To invoke again the analogy of the shop with the copy machine, LoopNet can be compared to an owner of a copy machine who has stationed a guard by the door to turn away customers who are attempting to duplicate clearly copyrighted works.”).

⁶⁰ *Id.*

⁶¹ *Religious Tech. Ctr v. Netcom On-line Commc'n Servs., Inc.*, 907 F. Supp. 1361, 1381–82 (N.D. Cal. 1995).

⁶² *Playboy Enters. Inc. v. Russ Hardenburgh, Inc.*, 982 F. Supp. 503, 514 (N.D. Ohio 1997).

⁶³ *CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 547 (4th Cir. 2004) (discussing the manual review, conducted by Loopnet employees, of photographs uploaded by Loopnet subscribers).

⁶⁴ *Id.* at 556 (applying the copy machine metaphor).

⁶⁵ 422 F. Supp. 2d 492 (E.D. Pa. 2006).

2008] AGAINST SEARCH ENGINE VOLITION 219*A. Are Search Engines more like Web sites or ISPs?*

Anyone that has traversed the Internet is immediately familiar with search engines. A user may go to Google.com or Yahoo.com, for example, type a search query into a text box, hit Enter, and the search engine will compare the text string to its index and provide a list of matching results.⁶⁶ It may display, among other things, the title of the Web page, which often serves as a link to the Web page, a snippet of the content of the page, the full URL for the page, and a link to the search engine's cached copy of the page.⁶⁷

What is perhaps a little less familiar is the process by which search engines operate. Because the amount of data on the Internet is massive, search engines cannot index it manually, but rely on software agents known as "web crawlers" or "spiders" to automatically scan Internet content.⁶⁸ That software organizes or indexes the information it comes across, and stores a copy of each Web page it finds in a repository known as a cache.⁶⁹ If a Web site owner does not wish to have his page indexed, or if he wishes not to have it added to the search engine's cache, he can signal the search engine to that effect.⁷⁰

⁶⁶ See, e.g., *Perfect 10 v. Google, Inc.*, 416 F. Supp. 2d 828, 832 (C.D. Cal. 2006).

Google operates different search engines for various types of web content. All search queries are text-based, *i.e.*, users input text search strings representing their query, but results can be in the form of text, images, or even video. Thus, for example, Google's basic web search, called Google Web Search, located at <http://www.google.com/>, receives a text search string and returns a list of *textual* results relevant to that query.

Id. (citation omitted).

⁶⁷ *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1111 (D. Nev. 2006).

When Google displays Web pages in its search results, the first item appearing in each result is the title of a Web page which, if clicked by the user, will take the user to the online location of that page. The title is followed by a short "snippet" from the Web page in smaller font. Following the snippet, Google typically provides the full URL for the page.

Id.

⁶⁸ See generally Brin, *supra* note 13 (explaining how search engines function).

⁶⁹ *Id.*

⁷⁰ See *Field*, 412 F. Supp. 2d at 1112.

A principal way for Web site owners to communicate with Google's robot is by placing specific instructions in "meta-tags" within the computer code (called HTML) that comprises a given page. When the Googlebot visits a page, it reads through this code. If it encounters meta-tags, it follows the instructions provided. Thus, for example, a site owner can place the following meta-tag within a page to tell Google's robot not to analyze the page or include it in Google's Web index and search results: "<META NAME='ROBOTS' CONTENT= 'NOINDEX, NOFOLLOW'" . . . Using meta-tags, a Web site

The impact these characteristics should have on volition is not immediately clear. On the one hand, search engines would appear analogous to the Netpics Web site in *Playboy Enterprises v. Webbworld, Inc.*⁷¹ In that case, Webbworld obtained and then sold images it collected from newsgroups, including images copyrighted by Playboy.⁷² Through what it called “ScanNews” software, the Web site would troll selected adult-oriented newsgroups, download feeds, and then separate out text from images.⁷³ It would then create “thumbnail” copies of the images, making those available to Webbworld subscribers for a subscription fee of \$11.95 per month.⁷⁴ On these facts, the United States District Court for the Northern District of Texas explained that Webbworld had the requisite volition for direct infringement because it took “affirmative steps to cause the copies to be made.”⁷⁵

Search engines, in much the same way, are not passive conduits in the IAP or BBS sense. Information is not sent through them by third parties for third parties; instead, search engines actively go out and seek information on the Internet.⁷⁶ Like Webbworld’s Netpics, Google trolls the web, copies information, and then makes that information available to the users of its Web site.⁷⁷ Indeed, it is arguable that Google, like Webbworld, even profits from its ability to aggregate and redistribute the information it collects through its online advertising revenue.⁷⁸

The *Netcom* analysis therefore initially suggests that search engines should have volition. As Nicole Bashor explains, in the

owner can also tell Google’s robot that it can include a given page in Google’s index, but that it should not provide a “Cached” link to that page in Google’s search results. To do so, the Web site owner uses a “no-archive” meta-tag “<META NAME=‘ROBOTS’ CONTENT=‘NOARCHIVE’>”.

Id. at 1112–13 (citations omitted).

⁷¹ *Playboy Enters., Inc. v. Webbworld, Inc.*, 991 F. Supp. 543, 550, 553 (N.D. Tex. 1997).

⁷² *Id.* at 552.

⁷³ *Id.* at 550.

⁷⁴ *Id.*

⁷⁵ *Id.* at 552 (quoting *Religious Tech. Ctr. v. Netcom On-line Comm’n Servs., Inc.*, 907 F. Supp. 1361, 1381 (N.D. Cal. 1995)).

⁷⁶ See Brin, *supra* note 13.

⁷⁷ *Id.*

⁷⁸ See, e.g., Google AdSense, <https://www.google.com/adsense/login3?gsessionid=7RNW-14xmMU> (last visited Mar. 13, 2008); BBC News, *Advertising Revenue Boosts Google*, Feb. 1, 2005, <http://news.bbc.co.uk/2/hi/business/4227967.stm>.

2008] AGAINST SEARCH ENGINE VOLITION 221

quintessential search engine index and cache context, Google is not like the copy machine owner who allows the public to access the machine but rather more like the copy machine owner who uses his machine to make infringing copies.⁷⁹

This comparison, however, overlooks a crucial difference. Despite a search engine's automatic collection of infringing material, the search engine's conduct, like that of an IAP or BBS, is not very close and causal to the illegal copying. One would be hard pressed to say that a search engine trespassed on the exclusive domain of the copyright holder by merely crawling and compiling sources already posted on the Web by a third parties; after all, it is exactly this role of the third party in the IAP and BBS context that provides the requisite distance for non-volition.

Furthermore, a search engine's preference for actively collecting information rather than passively receiving it does not rise to the level of solicitation implicated in *Playboy*.⁸⁰ Search engines, like their IAP and BBS counterparts, do not knowingly infringe the copyright of the information they copy, nor do they seek to profit at the expense of the copyright owner.⁸¹ Their advertising revenue, unlike that of the Netpics Web site just discussed, is not intimately tied to particular infringing conduct. It is instead a function of the number of users searching for a particular topic. Google illustrates this point by applying its ad words service only to general Web searches and not to image searches or (most like Webworld's Netpics) image caches.

B. The Implications of Parker v. Google

The question is how courts should apply the volition test to recognize the non-volition of search engines. The United States District Court for the Eastern District of Pennsylvania suggests one approach in *Parker v. Google*. In that case, the court found that the automatic copying process alone was sufficient to show a lack of volition, thus insulating Google from a direct copyright infringement claim.⁸² The court reached its conclusion by

⁷⁹ Bashor, *supra* note 8, at 119 (“[T]he search engine is most analogous to the person using the copy machine to copy a book as opposed to an innocent service provider.”).

⁸⁰ See *supra* note 56 and accompanying text.

⁸¹ These are, of course, generalizations that are by nature not universal to every specific circumstance.

⁸² See *Parker*, 422 F. Supp. 2d at 497 (“It is clear that Google’s automatic archiving of USENET postings and excerpting of websites in its results to users’ search queries do not include the necessary volitional element to constitute

drawing an analogy between Google's search engine and an ISP. It succinctly explained that "[w]hen an ISP automatically and temporarily stores data without human intervention so that the system can operate and transmit data to its users, the necessary element of volition is missing."⁸³ The other half of the volition test—the third party request or affirmative steps element—was not even invoked. If the court had applied that portion of the volition test, it could hardly have failed to recognize the affirmative steps involved in Google's directing its bots to troll and cache the Web.

Commentators have attempted to explain the result in *Parker* by suggesting that, to the extent the court failed to find volition, it got the decision wrong. As Bashor argues in her article, *The Cache Cow*, the court's application of only the "automatic" portion of volition makes little sense.⁸⁴ In a world where almost any manual process can be handled automatically, particularly the copying of electronic information, it is inherently problematic to apply only the "automatic response" element when analyzing volition—otherwise, every would-be-infringer could simply avoid liability by becoming a software developer.⁸⁵

Yet recognizing that the automatic element alone is insufficient for the purpose of analyzing volition does not compel the conclusion that the "third party request" element must be strictly applied in every situation. Instead, courts should recognize, as I believe the Eastern District of Pennsylvania did in *Parker*, that the technical elements of the test sometimes fail to vindicate volition. The upshot is that where the "third party request" element would find that search engines have volition, and the search engine is at a sufficient distance from the infringing conduct, the "automatic" element should dominate the

direct copyright infringement.”).

⁸³ *Id.*

⁸⁴ See Bashor, *supra* note 8, at 119.

Finally, the court could apply the volitional conduct requirement and find it is not satisfied. The *Parker* and *Field* courts adopted this view and, accordingly, the search engines did not satisfy this requirement and were not liable for direct copyright infringement. The best view, however, is to apply the volitional conduct requirement and find it is satisfied by virtue of the back end programming by the search engines.

Id.

⁸⁵ *Cf. id.* (“When someone controls a software program, or in this case, a robot, liability should not be lost. As a result, search engines should be held responsible for the actions of their robots because they directed the robots to perform the web site copying.”).

court's analysis.⁸⁶

V. ADAPTING VOLITION TO ACCOMMODATE SEARCH ENGINES

Ignoring half of the *Netcom* volition analysis as the court did in *Parker v. Google* is, however, a less than ideal solution. A better approach might be to simply adapt the “third party request” proxy to recognize search engine non-volition. This Part takes up that task by focusing on the responsive elements of search engines: their role as a conduit between content providers and Internet users, and their reliance on “robots.txt” files. It then raises and engages a challenge to this characterization of search engines as responsive to third party requests—their exercise of editorial control.

A. *Re-envisioning Third Party Requests*

Search engine non-volition can be recognized within the *Netcom* framework if courts apply a less rigid view of the “third party request” element. Such a view would recognize and prioritize two responsive aspects of the search engine model: their role as conduits between content providers and Internet users, and their reliance on “robots.txt” files.

The United States District Court for the District of Nevada recently emphasized the former in *Field v. Google*.⁸⁷ In that case, the court applied § 512(b)(1)(B) of the DMCA to Google before granting the search engine's motion for summary judgment of non-infringement. In doing so, it characterized search engines in a way that supports a broad reading of the “third party request” element of volition: the search engine's effective function is to serve as a conduit between third parties. It noted, for example, that “one of Google's principal purposes in including Web pages in its cache is to enable subsequent users to access those pages”⁸⁸ Additionally, the court explained that “Google's cache is a repository of material that operates between the individual posting the information, and the end-user requesting it.”⁸⁹

Yet what is most important for the present argument is the

⁸⁶ *Parker v. Google, Inc.*, 422 F. Supp. 2d 492, 497 (E.D. Pa. 2006).

⁸⁷ *Field v. Google, Inc.*, 412 F. Supp. 2d 1106, 1123–24 (D. Nev. 2006). *See generally* 17 U.S.C. § 512(b) (2000) (mandating a safe harbor provision that under certain circumstances on-line material is immune from liability).

⁸⁸ *Field*, 412 F. Supp. 2d at 1124.

⁸⁹ *Id.*

court's application of §512(b)(1)(B). That section of the DMCA provides that the owner of a cache is immunized from liability where the material in question is transmitted from the person who makes it available online to a person other than himself at the direction of the other person.⁹⁰ In finding that Google satisfied this requirement the court employed a loose reading that parallels my recommendation for the "third party request" element of volition. The court explained that "Field transmitted the material in question, the pages of his Web site, to Google's Googlebot at Google's request. Google is a person other than Field. Thus, Google's cache meets the requirement of Section 512(b)(1)(B)."⁹¹ In other words, the court concluded that despite the fact that Google took the initiative in obtaining the material, the requirement that the material was sent by the Web site owner was nonetheless satisfied because the search engine was in some sense responding to Field's request.

This suggests the second argument for the responsive nature of search engines: how they can be understood as responding to third party requests through their reliance on robots.txt files. While search engine bots are initially sent out to catalogue the Web, their actions are not wholly subject to the discretion of their creator. If a Web site owner does not wish to have his page indexed by a search engine, the owner can instruct a search engine's bots to that effect by including the proper "robots.txt" designation: <meta name = "robots" content = "noindex, nofollow">.⁹² Likewise, if a Web site owner wishes to prevent his content from being added to a search engine's cache, he can effect that desire by including a "noarchive" tag.⁹³ Because third party

⁹⁰ 17 U.S.C. § 512(b)(1)(B).

⁹¹ *Field*, 412 F. Supp. 2d at 1124.

⁹² *Id.* at 1112.

A principal way for Web site owners to communicate with Google's robot is by placing specific instructions in "meta-tags" within the computer code (called HTML) that comprises a given page. When the Googlebot visits a page, it reads through this code. If it encounters meta-tags, it follows the instructions provided. Thus, for example, a site owner can place the following meta-tag within a page to tell Google's robot not to analyze the page or include it in Google's Web index and search results: "<META NAME='ROBOTS' CONTENT='NOINDEX, NOFOLLOW'>".

Id. (citations omitted). See also Bashor, *supra* note 8, at 108; Google Webmaster Help Center, <http://www.google.com/remove.html> (last visited Mar. 20, 2008).

⁹³ *Id.* at 1112-13.

Using meta-tags, a Web site owner can also tell Google's robot that it can include a given page in Google's index, but that it should not provide a "Cached"

uploaders of content have these means to limit a search engine's acquisition of information, the absence of such instruction suggests that search engines are responding to implicit third party requests to index and cache content.

B. A Challenge and a Response

There is, however, a challenge to this characterization of search engines as responding to third party requests: search engines typically retain a measure of editorial control. As Professor Goldman has observed, search engines often ignore metatags or choose to incorporate only portions of very large files.⁹⁴ These practices suggest that search engines are less of a conduit between third parties and more of a publisher of third-party content. After all, a search engine can hardly be described as responding to a third party's request, however broad the definition may be, if all it does is ignore that request.

But this editorial control, for two reasons, is ultimately inapposite to volition. First, a search engine's decision to ignore keyword and description metatags or failure to include full copies of large files in its index⁹⁵ are not editorial decisions about the content or whether to reproduce the information at all (by, for example, ignoring a "no index" tag); rather they are decisions affecting the speed and integrity of the search engine's results. They are thus editorial in the same way that a copy machine's paper size, resolution, and color settings are editorial—that is to say, they may well affect the output, but they are arguably more administrative or structural in nature than anything else. Second, to the extent that search engines engage in editorial decisions that go beyond such administrative concerns, it is not clear that those decisions place the search engine any closer to the infringing conduct. Exactly the opposite may in fact be true: search engines could well exercise extended editorial control in a good faith effort to exclude infringing content from their catalogue. Such an effort, *LoopNet* teaches, is not and should not

link to that page in Google's search results. To do so, the Web site owner uses a "no-archive" meta-tag "`<META NAME='ROBOTS' CONTENT='NOARCHIVE'>`".

Id.

⁹⁴ Eric Goldman, *Search Engine Bias and the Demise of Search Engine Utopianism*, 9 YALE J.L. & TECH. 111, 113 n.7 (2006), available at <http://ssrn.com/abstract=893892>.

⁹⁵ Eric Goldman, *Deregulating Relevancy in Internet Trademark Law*, 54 EMORY L.J. 507, 567 (2005).

be the source of a machine owner's volition.

VI. CONCLUSION

Since its inception the notion of volition in copyright has served to insulate unwitting ISPs from claims of direct infringement arising from the conduct of their users.⁹⁶ While Congress has stepped in to shield such ISPs through the "safe harbor" provisions in the DMCA, there is no doctrinal reason the volition test should not also insulate other ISPs, including Internet search engines, from direct infringement.⁹⁷ After all, the volition analysis precedes rather than merely predates the "safe harbor" provisions. It determines whether direct infringement is possible, while the safe harbor provisions determine whether there is liability notwithstanding a *prima facie* finding of direct infringement.⁹⁸

Given the continuing importance of volition, courts should carefully consider the evolving nature of the analysis. They should recognize that *Netcom's* "automatic response" to a "third party request" test is not the sum and substance of volition, but rather a proxy for the Fourth Circuit's articulation of volition as "actual infringing conduct with a nexus sufficiently close and causal to the illegal copying that one could conclude that the machine owner himself trespassed on the exclusive domain of the copyright owner."⁹⁹ Thus, where the *Netcom* elements fail to approximate this underlying nexus, as in the case of a BBS owner that personally and manually accepted an infringing upload,¹⁰⁰ those elements should be circumvented via the copy machine metaphor if possible.

In the case of Internet search engines, there is typically a similar distance between the search engine's copying and the third party uploader's infringing conduct—a search engine has neither knowledge nor control of the infringing content. But as the *Parker* court's one-pronged volition analysis suggests,¹⁰¹ there is at present no similar mechanism capable of circumventing the "third party request" element's finding of volition in the search engine context. Because search engines use an opt-out means of

⁹⁶ *See supra* text accompanying Part II.A.

⁹⁷ *See supra* text accompanying Part II.B.

⁹⁸ *Id.*

⁹⁹ *CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 550 (4th Cir. 2004).

¹⁰⁰ *Id.* at 559.

¹⁰¹ *Parker v. Google, Inc.*, 422 F. Supp. 2d 492, 497 (E.D. Pa. 2006).

2008] AGAINST SEARCH ENGINE VOLITION 227

collecting online information rather than an opt-in model, they actively copy rather than passively respond to a third party requests.

While courts could simply follow the *Parker* court's lead and apply only the elements of the *Netcom* test that vindicate the underlying volition nexus, this Comment argues that a better approach might be to adapt the technical elements of the test in a manner similar to *LoopNet*'s adaptation of the copy machine metaphor.¹⁰² Specifically, search engine non-volition could be recognized within the existing *Netcom* framework if courts employ a broader reading of the "third party request" element that acknowledges both a search engine's position as a conduit between third parties and its responsiveness to "robots.txt" files.¹⁰³

*Matthew D. Lawless***

¹⁰² *LoopNet*, 373 F.3d at 550.

¹⁰³ See *supra* text accompanying Part V.A.

* J.D. Candidate, Indiana University School of Law—Bloomington, 2008; M.A., Texas Tech University, 2005; B.A., Michigan State University, 2001. An early draft of this Comment received first prize in the American Society of Composers, Authors and Publishers's Nathan Burkan Memorial Competition for best student paper on Copyright Law at the Indiana University School of Law—Bloomington. The author wishes to thank Professors Kevin E. Collins, Eric Goldman, and Miquel Peguera for their trenchant advice. Their suggestions resulted in many improvements, although perhaps not as many as they would have wished.